TECHNIKI MULTIMEDIALNE

Artur Przelaskowski

Warszawa, grudzień 2011

Wprowadzenie

Multimedia stały się ważnym składnikiem życia codziennego współczesnego człowieka. Często są narzędziem jego pracy intelektualnej i zawodowej, przyczyniają się w istotny sposób do zniesienia bariery czasu i przestrzeni pomiędzy ludźmi. Ubogacają komunikację, często są niezbędnym elementem rozrywki, nauki, poznania, wyrazu artystycznego, nowych doświadczeń. Odgrywają coraz większą rolę we wspomaganiu diagnosty medycznej, leczeniu, profilaktyce. Dla osób niepełnosprawnych są szansą skutecznej rehabilitacji, ale też aktywnego udziału w życiu świata.

W rozwoju współczesnych multimediów ważny okazał się rozwój technologii doskonalących formy przekazu (telekomunikacja, teleinformatyka), umożliwiających hierarchiczne osadzanie treści, odwołania do ogólnie dostępnych, dynamicznych źródeł wiedzy, prostą interakcję na odległość itp. Istotne stało się pojawienie pojęcia hipertekstu, czyli tekstu, w którym wszystkie ważniejsze pojęcia mają "niewidoczne" odnośniki (hiperłącza) encyklopedyczne i słownikowe. Odnośniki te za jednym kliknięciem przywołują teksty informacyjne odwołujące się to innych terminów i pojęć kluczowych, o których można by otrzymać informacje objaśniające itd. Miało to olbrzymie znaczenie dla rozwoju Internetu, ale też nowoczesnych nośników typu CD-ROM, a więc także dla multimediów.

Współczesna definicja multimediów nie jest prosta. Można o nich mówić w różny sposób, można je rozumieć odmiennie, można wreszcie przypisywać im różną rolę praktyczną, a więc odnosić je do różnego typu zastosowań. W **kręgu artystów** chodziło o identyfikację praktyk twórczych o charakterze interdyscyplinarnym, łączących rozmaite elementy wizualne i środki ekspresji, takie jak: filmy, muzyka, słowo mówione, projekcje świetlne, taniec, rysunek, malarstwo, grafika, fotografia, itp. Przy czym często sztukę multimedialną definiowano w kategoriach tzw. kultury komputerowej, wspomaganej efektami grafiki komputerowej, wirtualną rzeczywistością, przekazem internetowym.

Zasadnicze znaczenie słowa "multimedia" występuje w kontekście realnego, czyli aktualnego i zamierzonego przekazu informacji. **Informacja**, w pierwszym przybliżeniu, to dane z przypisaną treścią, która jest użyteczna dla odbiorcy

(prawdziwie lub domniemanie). Informacja przekazywana w danej chwili ma formę strumienia danych (tj. sekwencji danych potencjalnie nieograniczonej, z indeksem czasowym) o określonej reprezentacji. Środkiem przekazu staje się wszystko (sprzęt w danej technologii, protokoły, warstwa fizyczna, oprogramowanie itp.), co umożliwia przekaz informacji.

W kontekście przekazu komputerowego, w środowisku sieci lokalnych czy globalnych (np. pomiędzy serwerem a klientem), dużego znaczenie nabiera pojęcie **strumieniowania**, czyli przesyłania danych w formie strumienia, któremu towarzyszy wykorzystywanie kolejno napływających danych bezpośrednio po ich otrzymaniu. Typową realizacją jest bezpośrednie wyświetlanie filmu odbieranego jako otwarty strumień, bez wstępnego ustalenia całkowitego rozmiaru zamierzonego przekazu. Otrzymywany strumień może być jednocześnie rejestrowany w lokalnych archiwum lub też nie. Typowym przykładem jest monitoring kamerą z interfejsem sieciowym, kiedy to rejestrowane są jedynie fragmenty czy też usługi interaktywnej telewizji cyfrowej wykorzystującej protokół IP (IPTV): wideo na życzenie VoD (*V*ideo on Demand) czy też funkcja sieciowego magnetowidu NPVR (ang. Network Personal Video Recorder), dająca możliwość nagrywania (a potem korzystania z nich) aktualnie nadawanych programów telewizyjnych na dysku sieciowym operatora.

Multimedia to różne środki przekazu informacji, przy czym ta różnorodność dotyczy w pierwszej kolejności informacji (rodzaj, semantyka, treść), w drugiej - form przekazu (reprezentacja, jakość), a dopiero na końcu chodzi o zróżnicowanie środków (technologia, skala, zasady). Istotnym elementem jest synchronizacja przekazywanych strumieni informacji, wzajemna zależność treści komplementarnych względem siebie. Łączenie poszczególnych strumieni w jeden przekaz informacji powinno dać **efekt synergii**. Atrakcyjność multimediów wynika przede wszystkim z szybkiego dostępu do bogatych źródeł informacji, z coraz doskonalszych technologii odtwarzania i prezentacji tej informacji, możliwości jej gromadzenia, wymiany, obróbki, czy też upowszechniania własnych źródeł informacji. Bariery wynikające z fizycznych ograniczeń człowieka giną, a otwierają się nowe szanse rozwoju, współpracy, aktywnego udziału w życiu świata.

W kontekście multimediów występuje istotne pojęcie **mediów cyfrowych**, czyli określonej formy użytkowania treści multimedialnych, takiej jak Internet, telewizja cyfrowa, telefonia komórkowa, poczta elektroniczna, dystrybucje DVD, itp. Odwołuje się ono do starszego pojęcia mediów, czyli środków audiowizualnych przekazu informacji i rozrywki, takich jak radio i telewizja, filmy, nagrania dźwiękowe, magazyny, gazety, książki, płyty, taśmy magnetofonów, plakaty itp. Coraz trudniej dziś rozróżnić pojęcia mediów od mass mediów, czyli środków masowego przekazu. Dzięki technologicznym możliwościom rejestracji i komunikacji danych, sekretny zapis dokonany w domowym zaciszu może w jednej chwili stać się informacją przekazaną na cały świat.

Podstawową cechą mediów cyfrowych jest ich zdalna dostępność przez sieci telekomunikacyjne, z której wynika konieczność unormowania sposobów reprezentacji danych i ich opisów (metadanych) jako warunek skutecznej wymiany informacji [27]. Warto także zwrócić uwagę na proces integracji poszczególnych mediów wokół przekazu multimedialnego, a więc rejestratorów dźwięku i obrazu, komputerów, edytorów tekstu, narzędzi zapewniających interaktywny przekaz, koderów, nośników, monitorów, głośników, drukarek, skanerów, kart muzycznych, telewizyjnych, graficznych, sterowników, oprogramowania, itd. Celem integracji źródeł informacji, jej treści i formy oraz sprzętu i oprogramowania systemu multimediów jest efekt synergii, spodziewany u odbiorcy.

Multimedia są technologicznym, niedoskonałym odpowiednikiem naturalnego, ludzkiego przekazu informacji, obejmującego wielość zmysłów i różnorodny charakter komplementarnych sposobów wyrażania treści (modulacja głosu, gestykulacja, wyraz i mimika twarzy, zapach, sposób dotykania, ubiór, itp.). Pozyskując, dostarczając i prezentując informacje, kształtujące w dużym stopniu wiedzę o świecie współczesnego człowieka, multimedia stają się wyspecjalizowanym narzędziem poznawania świata, łamiącym istniejące dotąd bariery ograniczeń czasowych, przestrzennych, związanych z dostępem i komunikacją globalną. Są przy tym jednocześnie źródłem wielu zagrożeń, z których warto wymienić przede wszystkim pokusę zastępowania świata realnego wytworem globalnego przekazu informacji, wykrzywiania obrazu świata realnego natłokiem zbędnych treści i selektywnie wybieranych zestawień budujących fałszywą wiedzę. Multimedia mogą stać się ucieczką, zamiast szansą. Wydaje się, że pomocnicza funkcja wspomagania z zachowaniem dominującej roli ludzkiej osoby oraz przyłożenie większej wagi do treści oraz wiarygodności przekazu nie ograniczy roli multimediów, przyczyni się natomiast do bardziej harmonijnego i spójnego ich rozwoju.

Wśród zagadnień stanowiących podstawę opracowań współczesnych aplikacji multimedialnych znajdują się:

- **pozyskanie źródłowych danych multimedialnych** poprzez akwizycję, rejestrację, zapis danych z wykorzystaniem takich urządzeń jak kamery, aparaty fotograficzne, zestawy mikrofonowe, skanery, satelity, czujniki, systemy obrazowania medycznego, itp.;

- **składowanie i transmisję danych multimedialnych**, z istotną rolą efektywnej kompresji, transkodowania, kodowania archiwizacyjnego i nadmiarowego, kontroli i korekcji błędów transmisji, protokołów transmisji zapewniających interakcję, łatwy i szybki dostęp, struktur i mechanizmów bazodanowych, itd.;

- **opis treści multimedialnych** poprzez ekstrakcję efektywnych deskryptorów (hipertekst, metadane, deskryptory numeryczne, semantyczne), indek-

sowanie (szybkość i zasoby), wyszukiwanie (selektywność i formy zapytań), itd.;

- **ulepszanie multimediów**, w tym poprawa jakości (redukcja szumów, artefaktów i zniekształceń geometrycznych, zwiększanie rozdzielczości, poprawa kontrastu), poprawa percepcji treści, ekstrakcja treści ukrytej, uzupełnianie treści - wypełnianie brakujących fragmentów obrazu czy zapisu dźwiękowego, itd., syntezę treści metodami grafiki komputerowej (problem realizmu scen, wizualizacja danych), ale też metody oceny użyteczności danych multimedialnych (oceny ich jakości, wiarygodności, przydatności w konkretnych zastosowaniach);

- **analiza treści multimedialnych**, w tym segmentacja obiektów, ekstrakcja i selekcja cech, rozpoznawanie w inteligentnych systemach multimedialnych (rozpoznawanie twarzy, mowy, identyfikacja osób, detekcja obiektów, zmian patologicznych w obrazach medycznych, śledzenie obiektów);

- **prezentacja multimediów**, za pomocą wielokanałowych systemów odtwarzania dźwięku, nowoczesnych monitorów, telewizorów, systemów wizualizacji 3W, form wielkorozdzielczej reklamy dynamicznej, itp.;

- **ochrona danych**, w tym kryptologia, techniki szyfrowania, uwierzytelnienia, szyfrowania, ukrywanie treści metodami steganografii i znaków wodnych, itd.;

- **integracja** treści przekazu wielostrumieniowego, zarówno na poziomie transmisji czy składowania (różne formy synchronizacji i wzajemnego referowania danych), jak też metod odbioru przez użytkownika (zintegrowane modele odbiorcy, interakcja i sterowanie jakością);

- **standaryzacja**, w tym zwiększanie kompatybilności, *przezroczystości* technologii, tworzenie standardów technologii przyszłości, poszukiwanie nowych technik dostosowanych do wyzwań współczesności.

- inne.

Będziemy je nazywali **technikami multimedialnymi**. Znajomość niezbędnych podstaw teoretycznych, algorytmicznych oraz uwarunkowań realnych implementacji ww. technik warunkuje efektywne wykorzystanie oraz twórcze kształtowanie rozwoju świata multimediów.

Podręcznik ten stanowi zwarte, chociaż ze względu na szeroki zakres zagadnienia – pojemnościowo obszerne kompendium wiedzy oraz podstawowych umiejętności z zakresu technik multimedialnych. Jest efektem doświadczeń Autora, tak w pracy dydaktyczne, jak naukowej i aplikacyjnej w zakresie multimediów.

Znalazło w niej także wyraz szereg opinii i doświadczeń osób, które odkrywały przede mną tajniki świata multimediów, od których uczyłem się fascynacji obrazem, dźwiękiem, nowymi technikami i zastosowaniami. Mam tutaj na myśli w pierwszej kolejności Pana Doktora Mariana Kazubka oraz Panów Profesorów: Władysława Skarbka oraz Zbigniewa Kulkę z Politechniki Warszawskiej, Marka Domańskiego z Politechniki Poznańskiej, Ryszarda Tadeusiewicza z Akademii Górniczo-Hutniczej.

Książka składa się z pięciu zasadniczych części, pokrywających się ze strukturą rozdziałów. Pierwsza część zawiera ogólną charakterystykę multimediów, koncentrując się wokół zagadnienia zintegrowanego przekazu międzyludzkiego. Zwrócono uwagę na takie zagadnienia jak sama koncepcja przekazu informacji, rola oraz zadania nadawcy i odbiorcy, rejestracja i prezentacja informacji multimedialnej, ocena użyteczności narzędzi multimedialnych czy też powody tak znaczącej dziś roli multimediów.

Druga część dotyczy problemu skutecznego reprezentowania informacji, zwracając uwagę tak na aspekty składni, ogólniej syntaktyki, jak i semantyki. Wybór przestrzeni opisu sygnałów, usuwanie nadmiarowości, docieranie do form przekształceń porządkujących i dobrze opisujących przekaz ogrywają kluczowe znaczenie w efektywnym kodowaniu i indeksowaniu danych. Dobór sposobu reprezentacji danych możliwie zwartej, a jednocześnie odnoszącej się do przejrzystych znaczeń wyrażanych przez nie treści wpływa na użyteczność każdej aplikacji multimedialnej.

W trzecim rozdziale zawarto przegląd wybranych metod komputerowego przetwarzania informacji. Dobra reprezentacja ułatwia i usprawnia proces przetwarzania danych, przetwarzanie zaś pozwala ustalić reprezentację informacji użyteczną w danym zastosowaniu. Komputerowe przetwarzanie to zbiór rozwiązań zapewniających skuteczność przekazu multimedialnego, dostosowanie do potrzeb odbiorcy, urealnienie celów nadawcy, ale też zabezpieczenie przed ograniczeniami całego procesu przekazywania informacji. Ukazano głównie metody podstawowe, które mogą stać się inspiracją do ciągle poszukiwanych usprawnień, które coraz bardziej koncentrują się wokół takich pojęć jak technologie semantyczne, wykorzystanie efektu synergii wielostrumieniowego przekazu, integracja człowiek–komputer–człowiek.

Czwarta część stanowi rozwinięcie omawianych wcześniej teorii i metod w kierunku opisu narzędzi, konkretnych realizacji służących użytkowaniu informacji. Podane przykłady różnego typu aplikacji, rezultaty badań i eksperymentów, krótka charakterystyka obszaru zastosowań itd. zwracają uwagę na tak bardzo ważny pragmatyzm opracowań multimedialnych. Celem jest zachęcenie do własnego, ale bardziej przemyślanego użytkowania przekazu informacji. Ciekawym aspektem jest też weryfikowanie istniejących metod i narzędzi w kontekście nieograniczonych, coraz nowych wyzwań współczesności.

W piątej, ostatniej części zamieszczono krótką charakterystykę wybranych, stale doskonalonych standardów, które stanowią fundament rozwoju współczesnych multimediów, pozwalając realizować szeroko dostępne aplikacje multimedialne w dowolnej skali. Jest on zwieńczeniem rozważań dotyczących praktycznego aspektu rozwoju multimediów.

Zasadnicze treści tej książki można podzielić na 12 wykładów, należących do czterech grup tematycznych:

1. Charakterystyka multimediów:
   a) multimedialny przekaz informacji,
   b) metody rejestracji informacji,
   c) wizualizacja, percepcja i ocena użyteczności multimediów;
2. Informacja przekazu multimedialnego:
   a) reprezentowanie danych,
   b) podstawy kodowania informacji,
   c) podstawy indeksowania informacji;
3. Komputerowe przetwarzanie multimediów:
   a) metody ulepszania danych,
   b) metody analizy i syntezy danych,
   c) metody użytkowania informacji;
4. Standardy multimedialne:
   a) kompresja multimediów według standardów,
   b) opis multimediów według standardów,
   c) użytkowanie multimediów według standardów.

Zaproponowano także sześć ćwiczeń pogłębiających praktyczne aspekty prezentowanych zagadnień:

1. Rejestracja obrazu i dźwięku;
2. Postrzeganie przekazu informacji;
3. Kompresja multimediów;
4. Indeksowanie danych multimedialnych;
5. Komputerowe przetwarzanie danych multimedialnych;
6. Użytkowanie multimediów.

Zagadnienia dotyczące multimediów wymagają nieustannej aktualizacji, która wobec gwałtownego rozwoju dyscypliny będzie śledzić najnowsze trendy przy jednoczesnej selekcji osiągnięć trwałych, istotnych. Dlatego też zamiarem autora jest systematyczna aktualizacja i dostosowywane treści cyfrowej wersji tego opracowania do potrzeb współczesności. Wyrazem troski o wiarygodność przekazu, będą także dokonywane uzupełnienia formy według reguł sztuki multimediów.

Warszawa, grudzień 2011

Spis treści

Rozdział 1

Multimedia jako zintegrowany przekaz międzyludzki

Multimedia to odpowiedni kontekst dokonującego się realnie, czyli aktualnego i zamierzonego przekazu informacji, w którym istotne są:

- forma przekazu jako strumień danych, tj. sekwencja danych potencjalnie nieograniczona (bezpośrednio po otrzymaniu napływają kolejne dane), z indeksem czasu rzeczywistego, o określonej reprezentacji;
- integracja rożnych rodzajów informacji rozdzielonej na komplementarne, synchronizowane strumienie, zróżnicowane w zakresie:
  - natury przekazu – obraz, dźwięk, tekst, metadane itd.,
  - charakteru treści wynikającego zazwyczaj z natury danych,
  - formy reprezentowania informacji, zarówno w zakresie syntaktycznym, jak też semantycznym,

  z oczekiwanym efektem synergii;
- urządzenia i systemy, w tym m.in.
  - rejestratory (cyfrowe detektory obrazów, kamery, aparaty fotograficzne, zestawy mikrofonowe, studia nagrań itd.),
  - procesory przetwarzania danych (sygnałowe, graficzne, procesory efektowe, korektory graficzne, stoły mikserskie itp.),
  - technologie komunikacji (przewodowej, bezprzewodowej, sieci lokalne i globalne, protokoły, urządzenia dostępowe itd.),

  - systemy komputerowe ze specjalizowanym oprogramowaniem (nagrywarki, odtwarzacze multimedialne, przeglądarki itp.),
  - sprzęt do prezentacji (monitory, wyświetlacze, systemy wizualizacji 3W, systemy odsłuchowe, zestawy głośnikowe itd.);

- specyficzne cechy przy projektowaniu, konstruowaniu i doskonaleniu użytecznych rozwiązań, takie jak:
  - naśladowanie naturalnego, ludzkiego, możliwie kompletnego sposobu przekazywania informacji, podmiotowość użytkownika – odbiorcy przekazu;
  - przełamywanie barier oraz istniejących w kontaktach międzyludzkich (schemat nadawca – odbiorca) ograniczeń fizycznych,
  - szybki i selektywny dostęp do bogatych, wiarygodnych źródeł informacji,
  - stały rozwój, w tym doskonalenie technologii rejestracji, komunikacji i prezentacji, zwiększanie możliwości w zakresie gromadzenia i wymiany danych, opracowywanie skuteczniejszych metod wyszukiwania i obróbki informacji wobec rosnącej skali rozpowszechniania i poprawy łatwości obsługi, przy jednoczesnej indywidualizacji i szerszym wykorzystaniu mechanizmów oceny użyteczności itp.

Zaś **media cyfrowe** to

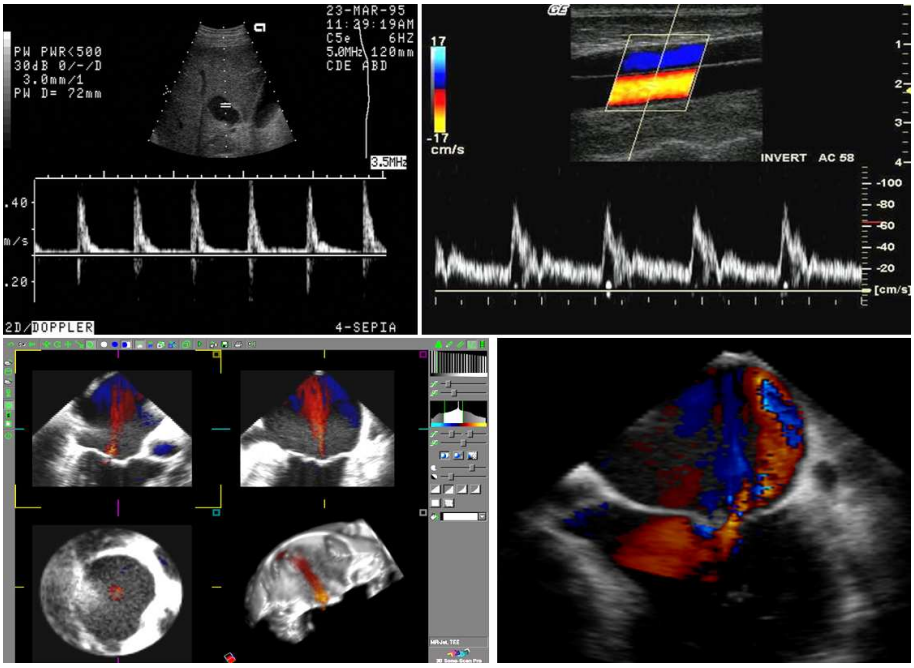
- określone formy użytkowania treści multimedialnych, takie jak Internet, telewizja cyfrowa, telefonia komórkowa, poczta elektroniczna, dystrybucje DVD itp.;
- zdalna dostępność przez sieci telekomunikacyjne, z której wynika konieczność unormowania sposobów reprezentacji danych i ich opisów (metadanych) jako warunek skutecznej wymiany informacji.

Możliwa jest też definicja multimediów jako mediów cyfrowych z

- przekazem kilku strumieni informacji w czasie, synchronizacją czasową i semantyczną przekazu;
- integracją treści i formy przekazu w celu uzyskania efektu synergii;
- treścią będącą odwzorowaniem/odbiciem naturalnych sposobów komunikacji według schematu człowiek–zmysły–poznanie.

Przykłady powszechnego wykorzystania multimediów można mnożyć – od drobnych aplikacji realizujących przekaz wielostrumieniowy na potrzeby lokalne,

domowe czy sąsiedzkie (np. urządzenie do identyfikacji gości z kamerą i mikrofonem, czy odtwarzacz DVD z monitorem), aż do kluczowych dziś mediów cyfrowych, takich jak cała infrastruktura internetu czy telewizji cyfrowej. Warto także zwrócić uwagę na kilka mniej typowych przykładów, jak chociażby telediagnostyczne badanie ultrasonograficzne (USG) serca (echo serca). Mamy tutaj do czynienia z zapisem sygnału EKG (elektrokardiograficzny zapis czynności elektrycznej serca), dźwiękowym monitoringiem funkcji zastawek (dopplerowski pomiar prędkości przepływu z przesunięciem rejestrowanych zmian częstotliwości sygnału w zakres akustyczny), ruchomym obrazem morfologii struktur oraz barwnych, dwuwymiarowych rozkładów przepływu krwi w komorach i przedsionkach - zobacz rys. 1.1. Badaniu towarzyszy także szereg informacji podawanych w postaci tekstowej oraz graficznej, a całość przekazu informacji ma charakter głęboko interaktywny.



Rysunek 1.1: Kompozycja treści multimedialnej w badaniu medycznej diagnostyki obrazowej (echokardiografii) jako przykład specyficznego wykorzystania multimediów.

Innym przykładem jest multimedialna scenografia z okazji koncertów, przedstawień, obchodów. Obok telebimów pojawiają się inne urządzenia, np. kurtyny diodowe, które pozwalają na dosłownie malowanie scen obrazami i kolorami, zwiększając siłę przekazu i urealniając go (rys. 1.2). Ciekawą możliwością jest także multimedialna prezentacja obiektu ze zdjęciami sferycznymi, wizualizacją przestrzenną, obrotem w dowolnym kierunku, zbliżeniem, przesunięciem, jednoczesnym podglądem lokalizacji na mapie miasta, informacją dźwiękową o istot-

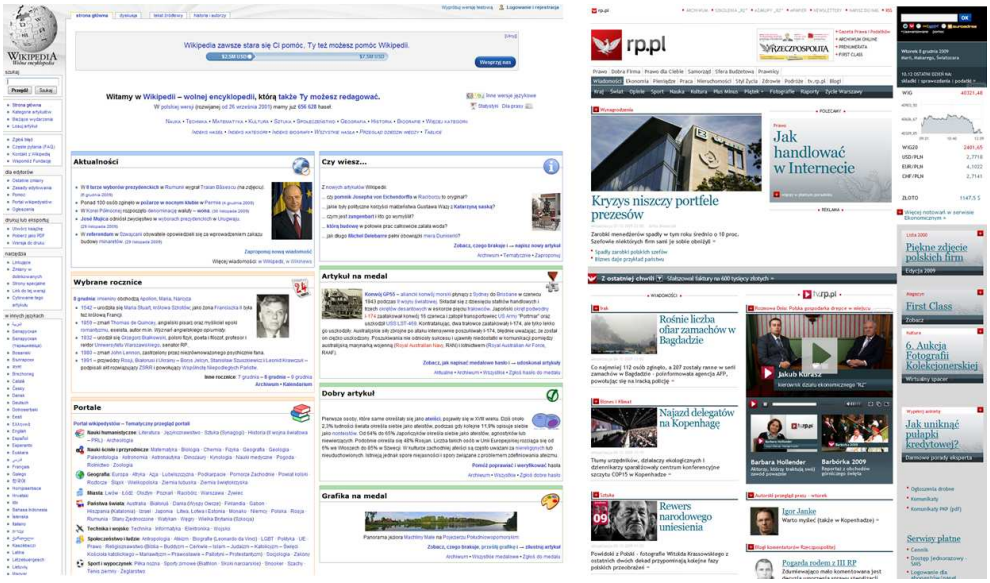
nych walorach turystycznych, muzyką charakterystyczną dla określonego czasu i miejsca, monitoringiem ze strumieniem wideo, pozwalającym podejrzeć aktualną pogodę, sytuację w obiekcie, w bezpośrednim otoczeniu, itp. (rys. 1.3).

Rysunek 1.2: Multimedialna scenografia koncertowa; źródło: zaczerpnięte ze strony http://marketingowiec.pl/artykul/multimedialne-sceny-na-dni-miast.

Rysunek 1.3: Multimedialna prezentacja obiektu ze zdjęciami sferycznymi, z aktualną lokalizacją na mapie miasta; źródło: zaczerpnięte ze strony http://www.meetingspoland.pl/searcher/presentations/Bohema/prezentacja_bohema.html.

Każdy niemal dzień przynosi nowe pomysły, realizacje, zamierzenia, doskonalące i rozszerzające świat multimediów – przykłady można mnożyć. Wystarczy wymienić portale edukacyjne, encyklopedie, internetowe wersje popularnych dzienników (rys. 1.4), portale społecznościowe, sterowanie robotami "widzącymi" i "słyszącymi", wirtualne światy (np. Second Life – rys. 1.5)), najnowsze rozwiązania telewizji przestrzennej (3W) czy technologie wykorzystywane w tworzeniu takich filmów jak Avatar. Warto śledzić najnowsze pomysły w obszarze techno-

logii multimedialnych oraz ich zastosowań, zwracając przy tym baczną uwagę na kluczowe trendy, które warunkują realny wzrost ich użyteczności. Lista zastoso-



Rysunek 1.4: Multimedialny charakter stron współczesnych portali internetowej encyklopedii (Wikipedii) oraz dziennika *Rzeczpospolita*.

Rysunek 1.5: Multimedialne drugie życie, czyli przykład wglądu w wirtualny świat SecondLife; jest to darmowa platforma, udostępniona publicznie w 2003 roku przez firmę Linden Lab, bazująca na rozbudowanej sieci gridowej. Dzięki niej można mieć wirtualną namiastkę coraz większego zakresu różnorodnych zadań życiowych, od nauki w szkole, po zmyślne formy realizacji zawodowej, leczenie czy prowadzenie biznesu. Jest to przykładowa forma realizacji postulatów Stanisława Lema: *Należy tchnąć ducha w maszynę. Źle mówię: należy sporządzić nowy świat, nadrzędny, pojęciowy, więc zbudowany z informacji i dać go człowiekowi. Nie tracąc ziemi spod nóg, człowiek zamieszka w tym świecie.* – fragment *Godziny przyjęć profesora Tarantogi* (słuchowisko radiowe, zamieszczone w zbiorze opowiadań Powtórka, Wydawnictwo Literackie, 1979).

wań technik multimedialnych jest więc szeroka. Podsumowując, warto wspomnieć przede wszystkim

- przeglądarki, wyszukiwarki internetowe, aplikacje zarządzające rozproszonymi bazami danych, semantyczne technologie internetowe;
- interfejsy człowiek–komputer, zintegrowane formy komunikacji wielozmysłowej;
- nowoczesną telewizję, wideo czy kino na żądanie, zdalny magnetowid, multimedialne radio, telefon;
- gry komputerowe z animacją domyślną, behawioralną, uczące się zachowań i preferencji użytkownika, bez ograniczeń poziomu abstrakcji spodziewanych efektów;
- internetowe transmisje "na żywo"– mecze, koncerty, teleturnieje, prezentacje reklamowe, telekonferencje, telekonsultacje, portale społecznościowe;
- nauczanie na odległość, e-edukację, cyberprzestrzenie, realistyczne światy wirtualne, portale wiedzy;
- specjalizowane portale gazet, sklepów, telewizji, radia, innych instytucji, e-handel;
- telemedycynę, opiekę domową, teleoperacje, telerejestrację itp..

W rozdziale tym szczególną uwagę zwrócono na podstawe zagadnienia dotyczące multimediów, uniwersalne fundamenty, które są ważne dziś i stanowią punkt wyjścia w debacie na temat technologii jutra. Chodzi o zasady formowania przekazu informacji oraz wykorzystanie ogólnych koncepcji "rozumowania multimedialnego" od wczesnego etapu rejestracji danych i kształtowania strumienia danych, aż po prezentację realnej informacji, z ogólnym zarysem procedur obróbki danych pozwalających uzyskać efekt dostosowany do wymagań odbiorcy.

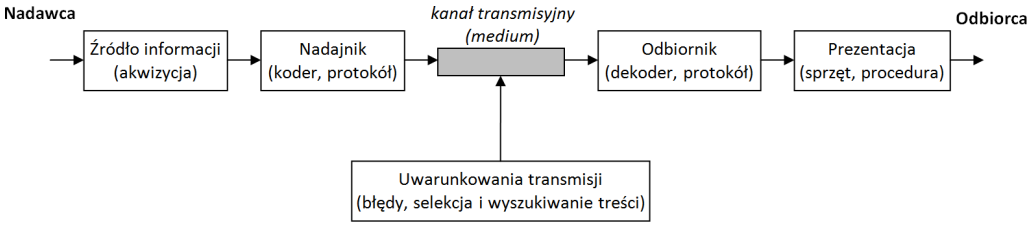
1.1 Przekaz multimedialny

Przekaz multimedialny może obejmować wiele strumieni informacji o odmiennej naturze nośnika – sygnału oraz zawartej treści, m.in.

- obrazy pojedyncze, wideo (zapis ciągły, przerywany, w zmieniających się warunkach akwizycji, filmy, animacje);
- dźwięk – audio, w tym mowa, śpiew, muzyka, odgłosy (może stanowić centralny strumień przekazu, uzupełniający lub też dodatkowy, np. alternatywna wersja językowa filmu);
- obrazy grafiki komputerowej (rastrowo, wektorowo, z modelami obiektów, kodowaniem), stanowiące zasadniczy przekaz (np. filmy animowane), uzupełniający (wizualizacja dodatkowych informacji, reklamy itp.) lub stanowiące interaktywny interfejs oddziaływania na treść przekazu;
- teksty (objaśniające, uzupełniające, definiujące źródło informacji lub autorów opracowania, itd.);
- dane mieszane, w tym hybrydowe archiwa, katalogi, dyski, dokumenty cyfrowe – np. książka łącząca tradycyjną formę cyfrową z odwołaniami do filmów, zapisów dźwięku, referencji do określanej bazy danych czy adresu URL;
- metadane, stanowiące opis zasadniczych strumieni informacyjnych (np. modele, deskryptory, struktury informacji, itp.);
- dane pomiarowe o różnym charakterze, np. zapisy czujników śledzących procesy fizyczne, dotyczące np. uwarunkowań zapisu głównych strumieni informacyjnych, pomiaru wartości temperatury, ciśnienia czy wilgotności;
- instrukcje sterujące, dotyczące np. uwarunkowań transmisji (protokołów, parametrów typu skala, rodzaj progresji treści), możliwych opcji wyboru jakości lub alternatywnej treści przekazu, doboru profilu użytkownika, itp.;
- warstwa synchronizacji przekazu służąca integracji treści wielostrumieniowej, w tym znakowanie zakresów, referencje, alternatywy odwołań, znaczniki czasu, mechanizm zapewniający ciągłość w czasie przekazu, itp.;
- inne.

Wielostrumieniowy przekaz informacji odbywa się w kontekście określonych uwarunkowań transmisji, czyli procesu przesyłania strumienia danych multimedialnych. Występuje więc nadawca wyposażony w źródło informacji, wybrany kanał transmisyjny (medium służące przekazaniu sygnału źródłowego będącego

nośnikiem informacji) wyposażony w nadajnik (dostosowujący sygnał przenoszący informację do celów transmisji) i odbiornik (w możliwie wierny sposób rekonstruujący postać sygnału źródłowego), oraz odbiorca (interpretator, użytkownik informacji) - rys. 1.6.



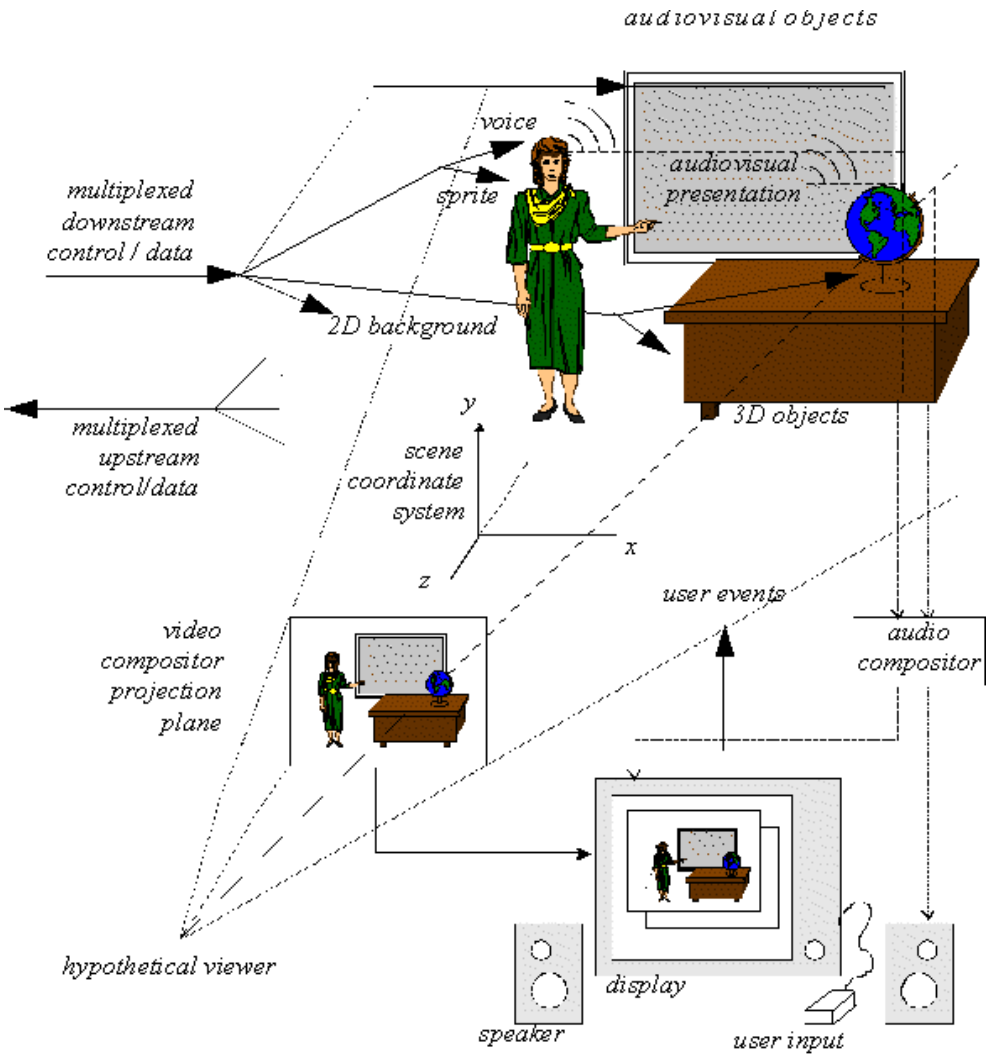
Rysunek 1.6: Schemat przekazu strumieni informacji multimedialnej od nadawcy do odbiorcy.

Przekazowi informacji towarzyszy szereg problemów realizacyjnych, związanych z jakością danych źródłowych, ograniczeniami kanału transmisji, zewnętrznymi źródłami zakłóceń czy różnego typu przekłamań, wreszcie z klarownością odczytu przekazywanej treści. Zależnie od zastosowań, występują niekiedy trudności z dopasowaniem intencji nadawcy do oczekiwań czy potrzeb odbiorcy, wynikające między innymi z różnego typu nieporozumień, odmiennego rozumienia charakteru danych, różnej interpretacji danych, inaczej odczytywanego znaczenia przesyłanych danych (odmienna funkcja semantyczna przypisywana danym przez nadawcę i odbiorcę), błędnego odczytania preferencji odbiorcy czy wręcz zasadniczej różnicy celów, wymagań i możliwości.

Rozumienie charakteru danych, właściwe odczytanie pełnej treści przekazu oraz poprawna interpretacja treści skutkujące pełnym zrozumieniem przekazu informacji (inaczej poprawnym odbiorem przekazywanej informacji) jest **fundamentalną zasadą** warunkującą sukces przekazu. Aby zrealizować aplikację działającą zgodnie z tą zasadą trzeba w pierwszej kolejności dostosować ją do charakteru danych multimedialnych. Warto pamiętać, że dane multimedialne są nośnikiem informacji multimedialnej, która jest podzielona na kilka zależnych od siebie strumieni o odmiennym, specyficznym i komplementarnym względem siebie charakterze. Przynajmniej jeden ze strumieni informacji powinien mieć znaczniki czasowe, czyli wskaźniki odnoszące się do upływającego czasu rzeczywistego, co urealnia przekaz multimedialny.

Schematyczną kompozycję sceny multimedialnej, analizowanej według reguł multimedialnego standardu MPEG-4, definiującego szereg aplikacji multimedialnych przedstawiono na rys. 1.7.

Podstawowym elementem sceny jest obraz – zarówno statyczny (inaczej pojedynczy) jak i dynamiczny (tzw. ruchomy, składający się z sekwencji pojedynczych obrazów – wideo). Charakter obrazu może być bardzo zróżnicowany: naturalny



Rysunek 1.7: Kompozycja sceny multimedialnej według standardu MPEG-4 (zachowano oryginalny język opisu normy http://mpeg.chiariglione.org/standards/mpeg-4/mpeg-4.htm).

(zdjęcie rzeczywistości obserwowanej ludzkim okiem) i graficzny (sztuczny, uzyskany za pomocą modeli i algorytmów komputerowych), oczywisty w treści oraz trudny w interpretacji, niejednoznaczny, z treścią bardzo subtelną, wręcz ukrytą.

Dalej scenę definiują uzupełniający dźwięk i tekst, czasem dominująca muzyka, kiedy indziej głos lektora – wyjaśniający treść sceny, zwracający uwagę, zachęcający, itp. Mogą być także ukazywane obiekty przestrzenne, symbole, syntetyczne formy wyrazu uzupełniające zdjęcia natury, itd. Wreszcie podmioty i przedmioty sceny, obiekty i ich wzajemne relacje, tło, ruch kamery nadający zrozumienie sytuacji zapisanej sceną oraz tak istotna dla użytkownika możliwość

interakcji, ingerencji w określone parametry sceny. Ponadto scena musi mieć przypisany zestaw urządzeń, rejestratorów i odtwarzaczy, czasem sprzętu kształtującego na bieżąco strumień informacji, przetwarzającego, opisującego. Wszystko to razem wymaga integracji, nawet jeśli nie ma zależności treści, to musi być synchronizacja przekazu, wspólne odniesienie do czasu i przestrzeni. Integracja ta służy jednemu wspólnemu celowi multimedialnego przekazu – informowaniu odbiorcy.

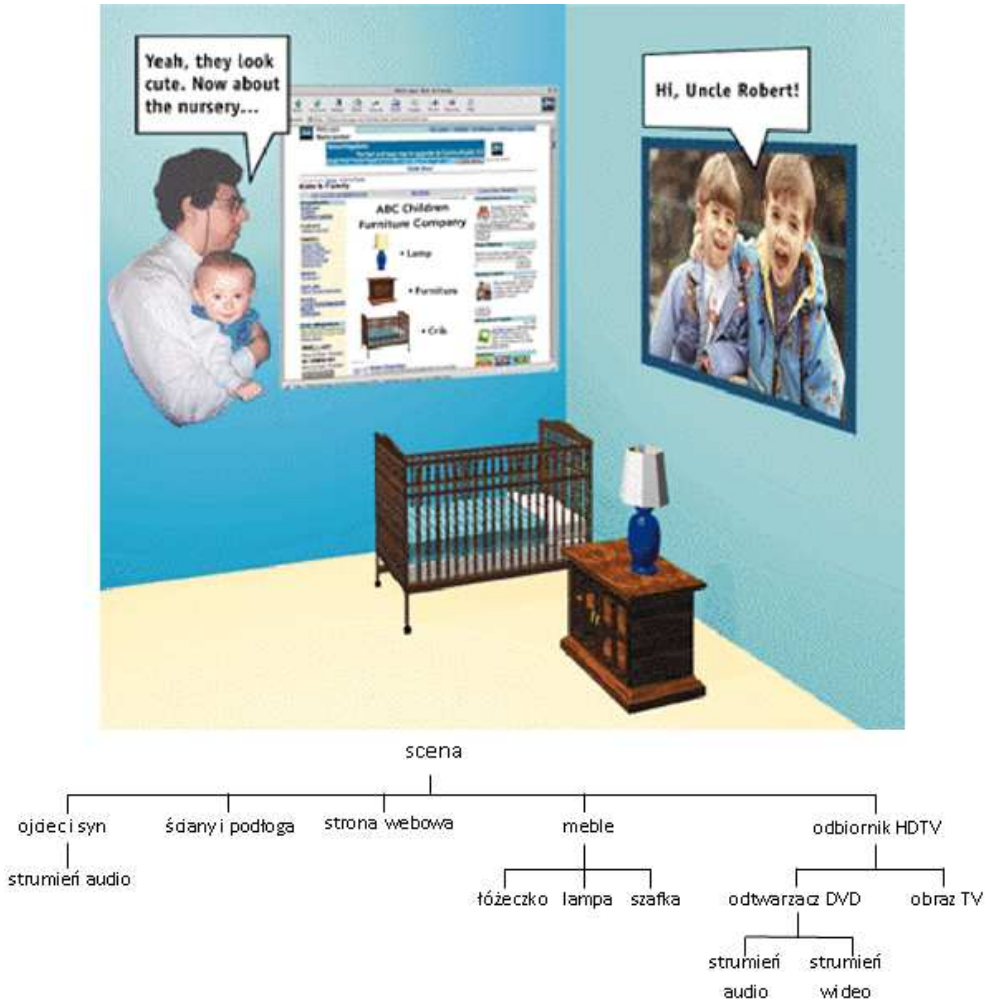
Strumienie danych multimedialnych, które opisują taką scenę, to przede wszystkim długie sekwencje obrazów (sekwencje wizyjne) z przewagą scen o ograniczonej informacji. Wideo jest zazwyczaj zintegrowane z dźwiękiem (sygnał audio), nieraz tekstem, grafiką, uzupełnione danymi kontrolnymi, systemowymi, umożliwiającymi interakcję, znacznikami synchronizacji itd. Specyficzne okoliczności ich wykorzystywania nakładają dodatkowe wymagania na sposób rejestracji, przysyłania i prezentacji danych: realność odbieranej treści i wiarygodność odczytywanej informacji, możliwość pracy w czasie rzeczywistym przy zmieniającej się przepływności sieci, skuteczne zabezpieczenie przed błędami transmisji, interakcję, efektywne indeksowanie informacji w celu sprawnej jej selekcji w bazie multimedialnej, itp. Celem jest wielobodźcowe oddziaływanie na zmysły człowieka w możliwie wiarygodnej formie, wiernej realiom natury w dopasowaniu możliwych form do istoty przekazywanej treści.

Przykład takiej sceny, łączącej różne rodzaje multimediów według koncepcji MPEG-4, ukazujący bogactwo przekazu informacji przedstawiono na rys. 1.8. Scena zawiera wiele obiektów niezależnych bądź złożonych. Mężczyzna z dzieckiem komunikuje się on-line ze swoją żoną (strumień wideo wklejony z kształtem dobranym do obiektów zainteresowania i towarzyszący mu zapis dźwięku). Wspólnie analizują ukazaną jak obraz na ścianie webową stronę sklepu z meblami. Wybierane meble są przestrzennie wizualizowane, przestawiane, obracane, wpasowywane w powstającą kompozycję pokoju dziecka, stanowiąc centralne obiekty sceny. Tło stanowią ściany i podłoga centralnego pomieszczenia, a dodatkowy strumień informacji wizualizowany na sąsiedniej ścianie to film odtwarzany z rodzinnych zasobów DVD. Wypadkowa konstrukcja tak złożonej sceny może być opisana za pomocą struktury drzewa jak na rys. 1.8.

1.1.1 Obraz, sekwencja wizyjna

Kategoria obrazów obejmuje obok klasycznie rozumianych obrazów statycznych przekazujących treść o typowym charakterze płaszczyznowym, także sekwencje wideo oraz zobrazowania przestrzenne (3W), obrazy grafiki komputerowej, animacje, rysunki, wykresy, a nawet fragmenty tekstu.

Postrzeganie jest naturalnym sposobem poznawania świata, przy czym bodźce wzrokowe wydają się zajmować tutaj miejsce szczególne. O ile porozumiewanie się z innymi ludźmi może być łatwiejsze za pomocą głosu – dzieci głuchonieme



Rysunek 1.8: Przykład sceny multimedialnej z wieloma obiektami obsługiwanymi za pomocą różnorodnych strumieni zintegrowanego przekazu multimedialnego. Pod obrazem sceny umieszczono drzewiastą strukturę opisu wieloobiektowej treści sceny; źródło: na podstawie opracowania M. Alberink, S. Iacob, The MPEG-4 standard. Overview of the MPEG-4 standard, fora and tools. Telematica Instituut, document of project GigaCE/D1.11, 2001.

wolniej rozwijają się od dzieci niewidomych – to już percepcja informacji jest pełniejsza, szybsza i bardziej jednoznaczna za pomocą obrazu. Tekst czy dźwięk bardziej pobudzają wyobraźnię, siłą obrazu jest większa zdolność poznawania świata.

Obraz lepiej nadaje się do tworzenia prostej symboliki, zwartego wyrażenia nieraz bardzo bogatej treści. Dlatego obrazy są tak ważnym nośnikiem wykorzystywanym we współczesnych systemach informacyjnych – według niektórych szacunków poprzez formę obrazu dociera do nas blisko 80% wszystkich informa-

cji. Taki przekaz jest bardziej pożądany przez odbiorców, pozawala lepiej zrozumieć naturę zdarzeń, jakby pełniej w nich uczestniczyć. Przekaz obrazu staje się kluczowy we wszystkich istotnych dziś mediach cyfrowych. Przykładowo, telefon komórkowy, w ślad za komputerowymi interfejsami użytkownika, w coraz większym stopniu orientowany jest wokół wysokiej klasy wyświetlacza obrazów, a wideorozmowy, nieograniczone wręcz fotografowanie czy filmowanie codzienności – patrz filmy na YouTube – staje się powszechne.

Obraz jest odzwierciedleniem rzeczywistości o treści kształtowanej przez kompozycję dających się wyróżnić a) obiektów, b) wydzielających je konturów i innych krawędzi, c) tekstur opisujących wewnętrzne właściwości obiektów, d) tła, uzupełniającego wymowę sceny.

Rozpoznawanym obiektom i ich cechom przypisywane jest określone znaczenie – semantyka, odnoszące się do wiedzy i doświadczeń użytkownika w kontekście określonych uwarunkowań rejestracji, formowania i prezentacji obrazów, a także zamierzonego sposobu użytkowania obrazów. Na treść przekazu obrazowego mają wpływ: rodzaj występujących obiektów, ich ogólne cechy takie jak rozmiar, kształt, spójność, jednorodność, specyfika tekstur, jakość zarysów, itd., a także relacje przestrzenno-czasowe pomiędzy określonymi obiektami. Niekiedy istotne znaczenie ma również liczba rozpoznanych obiektów, dominujący typ obiektu, charakter tła oraz relacje pomiędzy tłem i obiektami, semantyka innych dostrzegalnych szczegółów i ogólnych cech obrazowanej, w tym zrozumienie ogólnej wymowy całej sceny.

Na podstawie właściwej oceny rozpoznanej treści następuje możliwie dokładny odczyt pełnej informacji obrazowej, najlepiej odpowiednio uporządkowanej i zintegrowanej w swym wyrazie. Stanowi ona przedmiot interpretacji, której efektem są formułowane w etapie końcowym wnioski, czy podejmowane przez użytkownika decyzje.

Warte podkreślenia jest właściwe rozumienie istoty przekazu informacji za pomocą obrazu. Przy formalnej definicji obrazu jako macierzy wartości, każdy z trzech przypadków pokazanych na rys. 1.9 można nazwać obrazem. Jednak w obrazie zawierającym jedynie szum, nie sposób wydzielić obiektów, wyznaczyć konturów, czy sensownie określić tekstury. Nie ma wyrażonej treści, bo nie sposób ustalić jaka jest semantyka danych – nie ma więc żadnego przekazu informacji. Trudno więc w tym przypadku mówić o źródle informacji *sensu stricte* obrazowej, jako o istotnym komponencie multimedialnego wyrazu.

W przypadku wyświetlenia tekstu o określonej treści (w środku rys. 1.9) niewątpliwie mamy przekaz informacji dotyczący relacji Ali i kota. Informacja ta nie jest jednak dostosowana do typowego charakteru obrazu, jest zbyt prosta i uboga na taką formę przekazu, mało ciekawa, nieatrakcyjna. Można w tym przypadku mówić o obrazie uproszczonym, zubożonym, prymitywnym, ale także niejasnym, nieczytelnym. Nie wiemy przecież jak wygląda Ala, a nawet czy przedstawione

Rysunek 1.9: Kiedy obraz jest obrazem? Kolejno: formalny obraz bez przekazu treści (po lewo), formalny obraz z tekstem (w środku) oraz obraz będący przekazem treści typowo obrazowej, czyli przestrzennej – po prawej.

zdanie dotyczy konkretnego zwierzęcia domowego. Z większym zrozumieniem treści mamy niewątpliwie do czynienia w przypadku trzeciego obrazu (po prawej na rys. 1.9), gdzie dodatkowy element służy informowaniu odbiorcy, o jakiego kota chodzi w tym przekazie (Ala pozostaje nadal anonimowa). Dopiero zdjęcie kota jest elementem wykorzystującym specyfikę obrazu, bogactwo detali układających się w realny obiekt o klarownej semantyce. Bogatsza, pełna forma wyrazu przestrzennej informacji obrazowej, jaki ma miejsce przede wszystkim w obrazach naturalnych, stanowi podstawowy strumień przekazu multimedialnego w wielu zastosowaniach.

Bogactwo przekazu obrazowego dotyczy niebanalnej zwykle treści opisującej rzeczywistość, jest jakby oknem na świat, gdzie widać jedynie jego fragmenty, detale o różnym charakterze, skali, chwilowych uwarunkowaniach. Nawet całe mnóstwo takich okien nie daje pełnego obrazu rzeczywistości – trudno jest odtworzyć cały realizm zjawiska opisywanego obrazami. Warto pamiętać, że przy odczytywaniu treści obrazowej warto odwoływać się zarówno do realiów uchwyconej, chwilowej sceny, jak też do zasobów wiedzy, doświadczeń, a niekiedy nawet intuicji. Nie można bagatelizować znanych uwarunkowań procesu rejestracji/rekonstrukcji/formowania obrazu. Wszystko to pomaga uchwycić sens przekazu i odczytać, a następnie zrozumieć czasami bardzo subtelną, trudno dostrzegalną informację. Informacja ta podlega w dalszej kolejności interpretacji i ocenie przez odbiorcę. Takie wykorzystanie informacji obrazowej pozwala formułować wnioski czy podejmować decyzje odnoszące się do przedstawianych realiów.

Obrazy naturalne

Obrazy źródłowe, postrzegane przez ludzki zmysł wzroku, opisywane są generalnie funkcją jasności $\mathcal{C}(x, y, z, t, \lambda)$, reprezentującą przestrzenny rozkład energii

promieniowania widzialnego[1]. Jest to rzeczywista, nieujemna i ciągła funkcja kilku zmiennych niezależnych: współrzędnych przestrzennych x, y, z, czasu t oraz długości fali optycznej λ. Dziedzina ograniczona jest polem rejestracji (tj. umownym polem widzenia obserwatora), skończonym czasem obserwacji oraz zdolnością percepcji określonego zakresu promieniowania.

Odpowiedź w zakresie intensywności i koloru standardowego oka ludzkiego na funkcję jasności obrazu opisywana jest za pomocą pojęć: luminancji, opisującej zróżnicowanie jasności – intensywności oraz chrominancji, charakteryzującej zróżnicowanie odcienia oraz nasycenia kolorów – barw. Odpowiada temu funkcja pola obrazu definiowana jako:

$$\mathcal{F}_i(x, y, z, t) = \int_0^\infty \mathcal{C}(x, y, z, t, \lambda) S_i(\lambda) d\lambda, \quad i = 1, 2, \ldots \tag{1.1}$$

gdzie $S_i(\lambda)$ oznacza widmową odpowiedź czujnika i, nadającą charakter odbieranego, inaczej rejestrowanego obrazu – generalnie mamy do czynienia z kilkoma komponentami barwowymi, odpowiadających różnym podzakresom widma optycznego.

W ludzkim oku są to trzy rodzaje światłoczułych receptorów siatkówki – czopków o odmiennej charakterystyce widmowej (z dominującymi barwami czerwoną, zieloną i niebieską, porządkując według długości fal), pozwalające na zróżnicowany odbiór barw. Dodatkową, istotną zaletą czopków jest zdolność do szybkiej rejestracji obrazów (szybka reakcja na światło) oraz dużą ostrość, a więc wysoka rozdzielczość powstających za ich pomocą obrazów. Ograniczeniem jest natomiast niewielka czułość, wrażliwość jedynie na światło bezpośrednie. Receptą są dodatkowe receptory – pręciki, które umożliwiają widzenie przy bardzo słabym oświetleniu (niestety bez rozróżniania barw), skutecznie rejestrując światło rozproszone. Pręciki odpowiadają także za postrzeganie kształtów i ruchu, chociaż ich reakcja na światło jest wolniejsza niż czopków. Obraz uzyskany za pomocą pręcików daje także ograniczoną ostrość widzenia.

W typowym systemie rejestracji obrazów naturalnych typowym rozwiązaniem jest wykorzystanie czujników "wrażliwych" na trzy podzakresy długości fal promieniowania widzialnego, odpowiednio czerwony (indeksowany przez $i = 1$), zielony ($i = 2$) i niebieski ($i = 3$). Uzyskuje się wówczas informację o przestrzennym i czasowym rozkładzie kolorów przestrzeni RGB: $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3) = (\mathcal{F}_R, \mathcal{F}_G, \mathcal{F}_B)$. Na tej podstawie można wyznaczyć sygnał luminancji oraz inne składowe (komponenty) chrominancji. Za pomocą odpowiednich przekształceń można ustalić także bardziej korzystną w danym zastosowaniu przestrzeń barw.

Chociaż obrazy opisywane są zwykle w przestrzeni RGB, przede wszystkim ze względu na budowę ludzkiego oka oraz działanie typowych urządzeń służących

[1]Część widma promieniowania elektromagnetycznego rejestrowanego przez ludzkie oko, na którą reaguje siatkówka w procesie widzenia; jest to w przybliżeniu zakres fal o długości 380-780 nm, któremu odpowiada pasmo częstotliwości $3,8 - 7,9 \cdot 10^{14}$ Hz.

rejestracji i prezentacji treści obrazowej (kamery, aparaty fotograficzne, monitory, telewizory, skanery), to model percepcji obrazów naturalnych opisujący podobieństwo kolorów, a także wrażliwość na jakość odbieranej informacji wygodniej jest tworzyć w innych przestrzeniach barw. Także przetwarzanie obrazów wygodniej jest zwykle dokonywać w innej przestrzeni barw.

Wśród najczęściej wykorzystywanych przestrzeni barw można wymienić (więcej o kolorach w p. 2.3.4):

- sprzętowe, ważne przy konstruowaniu monitorów (RGB), drukarek (CMY, CMYK), procesorów grafiki komputerowej (HSV);
- percepcji, pozwalające rozróżniać i precyzyjnie definiować barwy widoczne, wyobrażane: XYZ, YIQ, YUV, YCrCb (stosowane np. przy kodowaniu obrazów kolorowych w standardach rodziny JPEG i MPEG).

Wykorzystanie systemów komputerowych w przekazie informacji obrazowej zakłada **cyfrową postać obrazów**. Jest to opisany macierzowo przestrzenny rozkład intensywności – jasności lub zestaw kilku barwowych komponentów o określonej dynamice oraz lokalizacji w przestrzeni i czasie, definiowany procesem rejestracji (akwizycji) w konkretnym systemie obrazowania.

Obrazy cyfrowe rejestrowane są różnorodnych systemach akwizycji z bezpośrednim konwersją funkcji jasności $\mathcal{C}$ na postać cyfrową (np. za pomocą matryc CCD) lub też z konwersją pośrednią (np. skanowanie klasycznego zdjęcia analogowego). Proces akwizycji obejmuje trzy zasadnicze etapy: a) próbkowania (tj. dyskretyzacji chwil czasowych lub współrzędnych przestrzennych, w których rejestrowany jest sygnał), b) kwantyzacji (tj. dyskretyzacji zbioru wartości rejestrowanego sygnału), c) kodowania (tj. ustalenia postaci binarnej reprezentacji kolejnych próbek sygnału). Jeśli

- funkcja jasności obrazu jest rejestrowana w skończonym przedziale czasowym (tj. realnym czasie akwizycji całego obrazu, po którym uśrednione zostają wartości pikseli – zwykle dąży się do minimalizacji tego czasu),
- funkcja jasności jest jednocześnie próbkowana wzdłuż obu współrzędnych przestrzennych,
- wartości tej funkcji są kwantowane w każdym punkcie dyskretnego pola obrazu,
- każdej skwantowanej wartości przypisana jest w sposób jednoznaczny odpowiednia reprezentacja bitowa,

wówczas otrzymujemy obraz cyfrowy reprezentujący rzeczywiste, ciągłe obrazy naturalne (*continuous natural images*). Proces kwantyzacji, próbkowania i kodowania realizowany w określonym systemie akwizycji i powtarzany w kolejnych,

ustalonych chwilach czasowych dostarcza sekwencji obrazów naturalnych nazywanej sekwencją wizyjną. Przykłady takich obrazów zamieszczono na rys. 1.10.

Rysunek 1.10: Przykładowe obrazy naturalne – zdjęcia cyfrowe, skany fotografii, obrazy testowe standardów JPEG, JPEG2000, MPEG (źródło: ogólnodostępne strony internetowe, np. ftp://ftp.ipl.rpi.edu/, http://decsai.ugr.es/ javier/denoise/test_images/index.htm).

Reprezentacja pojedynczego obrazu powstaje na bazie funkcji pola obrazu $\mathcal{F}(x,y,z) = \mathcal{F}_{i=u}(x,y,z,t=t_u)$ zarejestrowanej w danej chwili czasowej t_u, przy ustalonej charakterystyce czujników, określonej indeksem u. Nazwijmy ją, abstrahując od rodzaju komponentu i upraszając, funkcją jasności obrazu zdefiniowaną przestrzennie, czyli trójwymiarowo $f(x,y,z) \triangleq \mathcal{F}(x,y,z)$, rozumiejąc ją przede wszystkim jako odpowiednik rozkładu luminancji. W przypadku płaskim mamy $f(x,y)$.

Rejestrowana cyfrowo funkcja jasności obrazu podlega dyskretyzacji przestrzennej względem współrzędnych x i y (zaś w systemach akwizycji 3W – również z). Ponieważ w zdecydowanej większości systemów rejestracji pole obrazu jest prostokątne (lub prostopadłościenne w systemach 3W), wykorzystywany model obrazu bazuje na prostokątnym (prostopadłościennym) polu obrazu. Nawet jeśli aktywna lub użyteczna powierzchnia detektora ma inny kształt, zwykle wpisywany jest on w referencyjną strukturę prostokątną, stanowiącą układ odniesienia względem bardziej typowych urządzeń i systemów. Binarna maska wskazuje wtedy aktywną dziedzinę przekazu informacji.

Formalnie obraz cyfrowy jest zbiorem pikseli o współrzędnych dziedziny (pola) obrazu, z przypisanymi wartościami funkcji jasności. Niech $f_{\mathcal{O}} : \mathcal{P} \to \mathcal{F}$ oznacza

funkcję jasności obrazu $\mathcal{O}$, która przypisuje dyskretnemu polu $\Omega_f \in \mathbb{Z}^2$ tego obrazu dyskretny zbiór wartości $\mathcal{F} \in \mathbb{R}$. Zakładamy, że dziedzina obrazów jest powłoką (otoczką) wypukłą skończonego zbioru rejestrowanych pikseli $[\Omega_f]$, czyli przyjmuje postać wielokąta, najczęściej prostokąta.

Dziedziną przekształcenia jest wtedy zbiór wszystkich punktów pola (pikseli)

$$\Omega_f = \{(k,l) \in \mathbb{Z}^2 : \ k_{\min} \leqslant k \leqslant k_{\max}, \ l_{\min} \leqslant l \leqslant l_{\max}\} \tag{1.2}$$

gdzie $k_{\min} = \min_{(k,l)\in\mathcal{O}}\{k\}$, $k_{\max} = \max_{(k,l)\in\mathcal{O}}\{k\}$ oraz analogicznie $l_{\min}$, $l_{\max}$ ograniczające odpowiednio szerokość i wysokość obrazu. Stąd szerokość $K = k_{\max} - k_{\min} + 1$, a wysokość: $L = l_{\max} - l_{\min} + 1$. Zwykle, ze względu na konwencję zapisu liczb w rejestrach i komórkach pamięci cyfrowej, kładziemy $k_{\min} = l_{\min} = 0$, wtedy $K = k_{\max} + 1, L = l_{\max} + 1$.

Każda z wartości $f(k,l)$ – elementów uporządkowanego zbioru wartości pikseli $\mathbf{f} = \{f(k,l) : \ k = 0,\ldots,K-1, l = 0,\ldots,L-1,\}$ należy do uporządkowanego rosnąco zbioru (alfabetu) wartości możliwych, tj. $f(k,l) \in A_f = \{a_0,\ldots,a_{M-1}\}$. Zakładając kod dwójkowy jako naturalną regułę reprezentacji rejestrowanych obrazów cyfrowych otrzymujemy $k = \lceil \log_2 M \rceil$. bitowy obraz cyfrowy $\mathcal{O} \in \{0,1,\ldots,2^k - 1\}^{\Omega_f}$. W przypadku obrazów o q składowych barwowych mamy $f_{\mathcal{O}} : \mathbb{R}^2 \to \mathbb{R}^q$, $\mathcal{P} \to \{\mathcal{F}_i\}_{i=1}^q$. Wtedy alfabet wartości pikseli $(f_1, f_2, \ldots, f_q)$ jest iloczynem kartezjańskim zbiorów wartości poszczególnych komponentów obrazu: $A_{f_{1,\ldots,q}} = A_{f_1} \times \cdots \times A_{f_q}$.

Ze względu na postać dyskretnego zbioru wartości pikseli oraz liczbę komponentów, obrazy cyfrowe można podzielić na dwupoziomowe (*bilevel*) z $M = 2$ i $q = 1$ oraz wielopoziomowe (*continuous-tone, multilevel*) przy $M > 2$. Obrazy dwupoziomowe, tj. czarno-białe to przede wszystkim faksy, skanowane dokumenty tekstowe, szkice graficzne. W przypadku obrazów wielopoziomowych wartości pikseli mogą być wyrażone w skali szarości ($q = 1$) – mamy wtedy obrazy monochromatyczne ze skalą szarości – lub też w przestrzeni wielokomponentowej $q > 1$, które zwyczajowo nazywamy obrazami kolorowymi (barwnymi). Podstawowy zakres wartość dla komponentów obrazów wielopoziomowych, naturalnych wynosi 0–255 (obrazy bajtowe, $M = 256$), podczas gdy dla obrazów kolorowych zwykle $q = 3$ (np. w przestrzeniach barw RGB, YUV, YCrCb) lub $q = 4$ (np. CMYK).

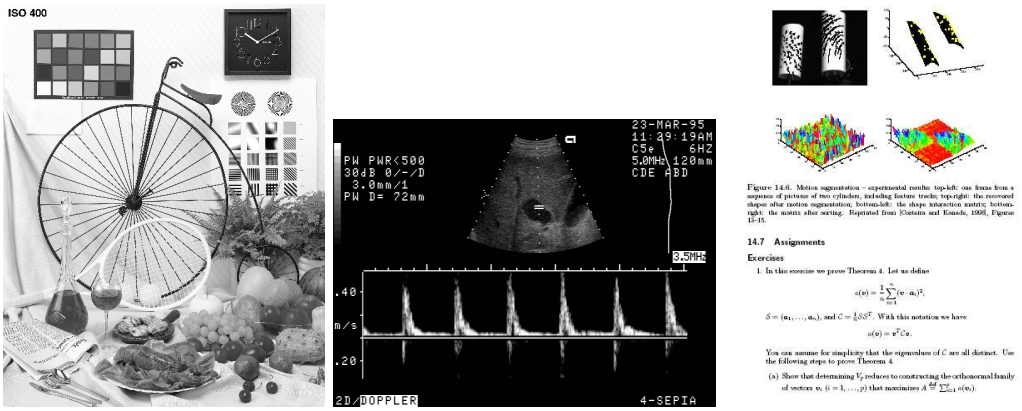
Realizm grafiki komputerowej

Są także obrazy tworzone w systemach grafiki komputerowej – renderingu obiektów przestrzennych, iluminacji, oświetlenia i cieniowania, z atrybutami tekstur nakładanych na przygotowane wcześniej modele, itp., a także w edytorach tekstu, z wykorzystaniem narzędzi do analizy i przetwarzania obrazów itp. – rys. 1.11). Są to tzw. sztuczne obrazy cyfrowe (inaczej obrazy graficzne, grafika).

Rysunek 1.11: Przykładowe obrazy graficzne, realistycznie symulujące obrazy naturalne, ale też niekiedy zawierające uproszczone schematy, obiekty, zarysy koncepcji scen itp.; obrazy te prezentując treść zbliżoną wielokrotnie do treści zawartej w obrazach naturalnych, wymagają zdecydowanie odmiennego podejścia w ich analizie, kodowaniu czy indeksowaniu (źródło:własne, Przemysław Kiciak, Henrik Wann Jensen, strony internetowe, http://www.assassinscreed.com/).

Reprezentacja obiektów przedstawianych w obrazie składa się zazwyczaj z trzech kategorii danych: geometrycznych (określają położenie i kształt składowych obiektu w przestrzeni o ustalonym układzie współrzędnych), topologicznych (definiują relacje pomiędzy składowymi obiektu - np. ustalają kolejność wierzchołków wielokąta na płaszczyźnie) oraz atrybutów (wartości, które ustalają różne właściwości składników obiektu, np. przezroczystość, chropowatość, zamglenie, rozmycie). Ponadto wyróżnia się dwie zasadnicze metody reprezentacji obiektów: brzegowe (konturowe) i obszarowe (objętościowe). Używa się różnych elementów podstawowych konstruujących obraz. Przykładowo mogą to być tzw. prymitywy geometryczne (np. trójkąty, ogólniej proste wielokąty), wielomiany, krzywe Beziera, powierzchnie funkcji sklejanych itd. Często efekty stosowania metod zapewniających realizm generowanych scen wplatane są w obrazy naturalne jako uzupełnienie podstawowej treści czy też ogólnych właściwości tła. Niekiedy wykorzystywane są także fragmenty tekstu, elementy kontrolne, znaczniki, dodane sygnały jednowymiarowe (1W) itp. – zobacz rys. 1.12.

Analiza obrazów graficznych ma odmiennych charakter, chociaż niekiedy w najbardziej zaawansowanych technikach realizmu graficznego o nieznanych algorytmach generacji, bazujących np. na złożonych modelach probabilistycznych obrazu (ukryte modele Markowa, mieszanina uogólnionych rozkładów normalnych) czy też przy nakładaniu drobnych elementów graficznych na sceny naturalne, metody analizy i przetwarzania obrazów mogą być podobne do metod stosowanych



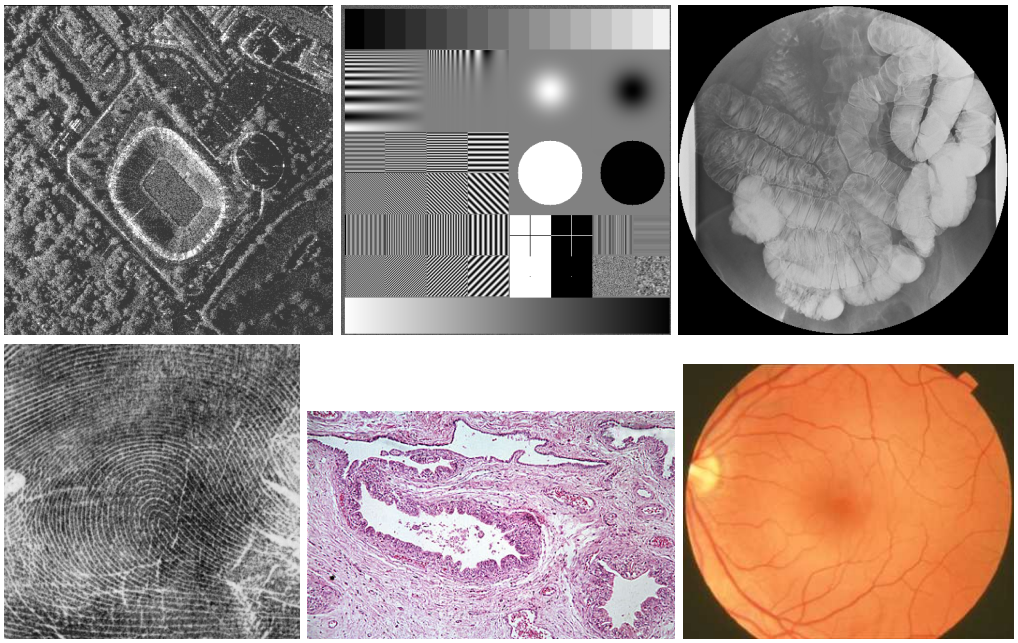
Rysunek 1.12: Przykładowe obrazy złożone z grafiki, zdjęć, kompozycji obiektów, sygnałów dopplerowskich i danych projekcji B badań ultrasonograficznych oraz grafiki i tekstu skanowanego dokumentu (źródło: własne, obrazy testowe standardów JPEG, JPEG2000, zasoby sieciowe).

w przypadku obrazów naturalnych. Zagadnienie kompresji obrazów graficznych ma odmienny charakter, wymaga modelowania przestrzennego (2W, 3W) występujących obiektów, oszczędnego opisu i efektywnego kodowania danych geometrycznych (np. rozkładu wierzchołków siatki) i topologicznych, a także zbioru wartości atrybutów sceny [2, 3].

Obrazy specjalistyczne

Obrazy specjalistyczne stanowią ważny składnik przekazu multimedialnego w kontekście wielu zastosowań. Dotyczy to przede wszystkim tych zastosowań, które czynią z multimediów nie tylko narzędzie rozrywki, ale też bardzo ważny czynnik warunkujący skuteczność najnowszych technologii wykorzystywanych w najbardziej kluczowych obszarach życia społecznego, takich jak opieka zdrowotna, kryminalistyka, różnego typu badania naukowe i wiele innych. Chodzi tutaj o zdjęcia satelitarne, meteorologiczne, wywiadowcze, geodezyjne, odcisków palców w biometrycznych urządzeniach kontrolnych, bazach agencji detektywistycznych, testowe do weryfikacji urządzeń obrazujących oraz metod przetwarzania obrazów, widzenia komputerowego, wykorzystywane w sterowaniu robotów, mikroskopowe, powiększane – siatkówki oka, w podczerwieni, itd. – rys. 1.13.

Ważną grupę stanowią obrazy medyczne – rejestrowane z wykorzystaniem detektorów cyfrowych, rekonstruowane w systemach tomograficznych, tworzone poprzez skanowanie klisz. Jakkolwiek odnoszą się do naturalnych obiektów o regularnych kształtach, to jednak ukazują ich wnętrze niedostępne dla ludzkiego oka. Obrazy te muszą więc być więc często rejestrowane z wykorzystaniem sygnałów o innym charakterze – odmiennej fizycznie naturze niż promieniowanie widzialne.



Rysunek 1.13: Przykładowe obrazy specjalistyczne (kolejno): zdjęcie satelitarne, obraz testowy, medyczny, odcisków palców, mikroskopowe, siatkówki oka (źródło: własne, obrazy testowe standardów JPEG, JPEG2000, zasoby sieciowe).

Jako nośnik zbieranej informacji wykorzystywane jest promieniowanie rentgenowskie, jądrowe czy nawet fal radiowych, ultradźwięki, wiązki elektronów, rozkłady pola elektrycznego, elektrostatycznego, magnetycznego, itp. Techniki optyczne bazują tutaj na zjawisku rozpraszania, a specjalistyczne technologie obrazowania narządów wewnętrznych wykorzystują sygnał z kamery wprowadzanej cewnikiem do wnętrza ludzkiego organizmu (np. badania endoskopowe z obrazowaniem za pomocą wziernika z układem optycznym i własnym źródłem światła – gastroskopia, bronchoskopia, itp.). Pewna cecha charakterystyczna tkanek penetrowanego obszaru (zróżnicowanie oporności akustycznej, poziom osłabienia promieniowania rentgenowskiego, zdolność absorpcji kontrastu, itd.) jest obrazowana w odpowiednio skontrastowanym polu powstającego obrazu.

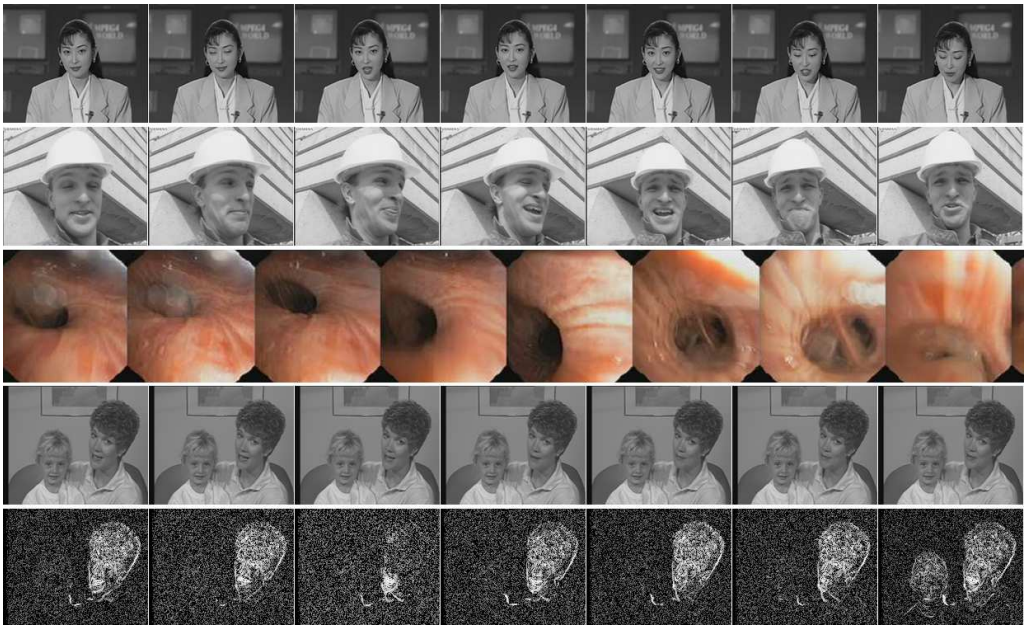
Generalnie, obrazowanie diagnostyczne w medycynie polega na wizualizacji narządów wewnętrznych człowieka metodami nieinwazyjnymi za pomocą specjalistycznej aparatury diagnostycznej i odpowiednich metod badawczych. Obrazy medyczne ukazujące wnętrze ludzkiego ciała pełnią bardzo ważną rolę w diagnostyce pacjentów, ale też wizualizacji i ocenie ilościowej skutków stosowania terapii.

Interpretacja obrazów tej grupy wymaga zwykle specjalistycznej wiedzy danej dziedziny, uwzględnienia specyfiki obrazowanej rzeczywistości, często jest to połączone z zadaniami wykrywania i rozpoznawania patologii czy innych obiek-

tów istotnych diagnostycznie. Ważne jest tutaj wykorzystanie wszelkiej dostępnej informacji (dane kliniczne, wywiad chorobowy, rezultaty innego typu badań diagnostycznych jak oglądanie, obmacywanie, opukiwanie i osłuchiwanie). Szczególny nacisk położony na jakość przekazu informacji, w tym wiarygodność i właściwą percepcję treści obrazowej.

Sekwencje obrazów

Czasowe sekwencje wizyjne (obrazowe) stanowią kluczową formę dynamicznego przekazu multimedialnego, często w realiach czasu rzeczywistego (*on-line*). Filmy telewizyjne, zapis wideo, dane z kamer przemysłowych w systemach monitoringu, analizy scen w stereoskopowym układzie kamer, systemach widzenia komputerowego, a także wiele innych, to informacja reprezentowana w postaci ciągu obrazów z odniesieniem do wymiaru czasu, czasami dźwięku czy uzupełniających informacji tekstowych. W przypadku zastosowań specjalistycznych, np. radiologii, są to przykładowo perfuzyjne badania dynamiczne tomografii komputerowej czy rezonansu magnetycznego, dynamiczne badania ultradźwiękowe czy śledzenie procesów fizjologicznym metodami medycyny nuklearnej. Przykłady podano na rys. 1.14.



Rysunek 1.14: Przykładowe fragmenty naturalnego wideo oraz sekwencji obrazów specjalistycznych – bronchoskopowych. W piątej linii zamieszczono kolejno różnice pomiędzy ramkami (drugą i pierwszą, trzecią i drugą,itd., a jako ostatnią - różnicę pomiędzy klatką siódmą i pierwszą) sekwencji z linii czwartej (źródło: własne, obrazy testowe standardów MPEG).

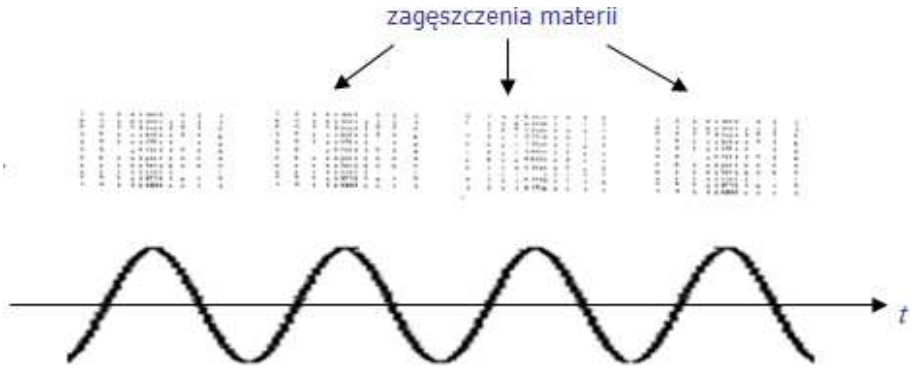
Cechą charakterystyczną sekwencji obrazów jest zwykle silna korelacja wzajemna (ogólniej zależność) kolejnych ramek (ich fragmentów) występująca przynajmniej w najbliższym sąsiedztwie według czasowego porządku rejestracji. Wiele treści zawartej w przestrzeni obrazu (rodzaj i cechy obiektów, przestrzenny rozkład obiektów, tło, oświetlenie, kierunek i prędkość przemieszczania się kamery itp.) powtarza się z kadru na kadr. Fakt ten można i należy wykorzystać konstruując efektywne metody analizy treści, śledzenia ruchu obiektów, wyszukiwania obiektów podobnych, interpolacji czy kodowania, przy czym przedmiotem optymalizacji są modele występujących obiektów, sposoby opisu ich cech (kształtu, tekstury, koloru), ruchu, wzajemnych relacji przestrzennych kilku obiektów, itp. Różnica pomiędzy kolejnymi klatkami, której nie da się w prosty sposób przewidzieć, stanowi o informacji przekazywanej w kolejnych ramkach. Pierwsze trzy sekwencje na rys. 1.14 (licząc od góry) zostały zestawione z odległych od siebie ramek, by dobitniej ukazać różnorodność zmieniającej się treści. Sekwencja czwarta zawiera kolejne ramki, tak jak zostały zarejestrowane (około 30 ramek na sekundę), z silnymi zależnościami treści pomiędzy kolejnymi ramkami. Ostatnia linia zawierające obrazy różnicowe sąsiadujących ze sobą ramek pokazuje zmiany dotyczące przede wszystkim ruchu głowy oraz ust matki. Dopiero różnica klatki ostatniej i pierwszej tej krótkiej sekwencji ukazuje bardzo subtelny ruch głową córki.

1.1.2 Dźwięk, audio, mowa

Drugim (po widzeniu) kluczowym doznaniem, za pomocą którego odbieramy bodźce świata zewnętrznego poznając go, jest słyszenie. Zmysł słuchu umożliwia przekształcenie docierającej do ucha (tj. zasadniczego receptora dźwięku) fali dźwiękowej na doświadczane przez słuchacza wrażenie odbioru dźwięku. Powstaje ono wskutek konwersji mechanicznej energii drgań ośrodka propagacji dźwięku na hemodynamiczną energię płynu przedsionka (ucho środkowe), a następnie na elektryczne impulsy nerwowe (ucho wewnętrzne – ślimak), które za pomocą nerwu słuchowego docierają wreszcie do mózgu (kory słuchowej).

Dźwięk to fala akustyczna rozchodzącą się w sprężystym ośrodku materialnym (ciele stałym, płynie, gazie), zdolna wytworzyć wrażenie słuchowe. Jest falą mechaniczną, czyli rozchodzącym się fizycznym zaburzeniem ośrodka ((rys. 1.15), polegającym na pobudzaniu do drgań kolejnych cząsteczek ośrodka – w uproszczeniu cząstki drgają wokół swego stanu równowagi i przekazują energię cząsteczkom sąsiednim. Do opisu zjawiska propagacji dźwięku wykorzystuje się podstawowe prawa ruchu falowego, formowanie się fali płaskiej i kulistej, zasady odbicia, załamania i rozproszenia fali, tłumienia i absorpcji, itd. Synonimem dźwięku w technologiach multimedialnych jest słowo audio (z łaciny "słyszę").

Prędkość v rozchodzącej się fali dźwiękowej, zależna od gęstości ρ i ściśliwości ośrodka (dokładniej modułu ściśliwości objętościowej ε, inaczej sprężystości czy



Rysunek 1.15: Rozchodzenie się fali akustycznej (tj. zagęszczeń materii powodowanych drgającym ruchem cząsteczek ośrodka) w ośrodku sprężystym.

Younga), wyrażona jest wzorem:

$$v = \sqrt{\frac{\varepsilon}{\rho}} \tag{1.3}$$

Zestawienie wartości prędkości propagacji dźwięku w różnych ośrodkach (jednostka [m/s]): 340 w powietrzu, 1500 w wodzie, 4600 w miedzi, 6320 w aluminium czy 60[2] w gumie pokazuje ich bardzo silne zróżnicowanie. Odmienny jest też charakter fal dźwiękowych – podczas gdy w powietrzu i cieczach są to fale podłużne (kierunek drgań jest zgodny z kierunkiem propagacji fali), to w ciałach stałych dominuje poprzeczny charakter drgań.

Całościowa charakterystyka tych właściwości ośrodka, które są istotne w propagacji fal dźwiękowych, definiowana jest za pomocą oporności (impedancji) akustycznej, której wartość:

$$Z = \rho \cdot v = \sqrt{\varepsilon \cdot \rho} \tag{1.4}$$

Zróżnicowanie wartości oporności akustycznej różnych ośrodków wpływa na występowanie takich zjawisk jak odbicie, załamanie, rozpraszanie czy absorpcja.

Psychofizyczne efekty oddziaływania drgań cząstek ośrodka na zmysł słuchu ludzi i zwierząt rządzą się także swoimi prawami. Przyjmuje się, że typowo ludzkie wrażenia słuchowe powodowane są zaburzeniem o częstotliwościach z zakresu 16Hz-20kHz, co definiuje zakres fal dźwiękowych. Drgania o częstości mniejszej od 16Hz nazywane są infradźwiękami, a o częstotliwości ponad 20kHz – ultradźwiękami. Jednak wrażliwość ludzkiego ucha na różne częstotliwości dźwięku nie jest jednakowa. Istotne są także czasowe następstwo dźwięków o różnej intensywności (energii drgań). Różnymi aspektami percepcji dźwięku zajmuje się psychoakustyka.

[2]Są to wartości orientacyjne, gdyż prędkość dźwięku zmienia się w funkcji temperatury, ciśnienia i innych czynników, charakterystycznych dla danego materiału

Szczególnym, wydzielanym w aplikacjach multimedialnych rodzajem dźwięku jest ludzka mowa, czyli sposób porozumiewania się (w szczególności przekazywania informacji) za pomocą głosu. Specyfika dźwięku mowy obejmuje szereg elementów, m.in. wąski zakres częstotliwości pozwalający zrozumieć przekaz głosowy (zwykle zawężany do pasma 300Hz-3kHz), charakterystyczne fonemy (tj. najmniejsze jednostki mowy, charakterystyczne dla danego języka) oraz składane z fonemów elementy znaczeniowe, czyli morfemy (tj. grupa fonemów o określonej semantyce, której nie można podzielić na mniejsze jednostki znaczeniowe), wyrazy, zwroty i wyrażenia.

Zapisy dźwięku i mowy zawierają wartości kolejno próbkowanych sygnałów analogowych, których analiza i ewentualne przetwarzanie najczęściej uwzględnia percepcyjne zdolności ucha ludzkiego, jak też sam sposób generacji pojedynczego tonu[3], dźwięku złożonego[4] czy wypowiadanego słowa. Odmienny charakter danych będących zapisem mowy i łagodniejsze kryteria jakościowe (koncentrujące się głównie na zachowaniu zrozumiałości wypowiedzi) implikują inną klasę rozwiązań algorytmów przetwarzania, analizy i kompresji, a także inne modele odbiorcy i kryteria dopuszczalnych zniekształceń. Odmiennych metod obróbki wymagają zbiory danych zawierające dźwięk lub mowę generowane syntetycznie.

Dźwięk można opisać za pomocą cech obiektywnych, takich jak widmo częstotliwościowe (fourierowskie), czasowo-częstotliwościowe czy rozkład natężenia[5].

Poziom natężenia dźwięku I wyrażany jest zwykle we względnej skali logarytmicznej, odpowiadającej charakterowi odczuwania głośności, jako:

$$L = 10 \log \frac{I}{I_0} [dB] \tag{1.5}$$

gdzie wartość odniesienia I_0 wynosząca $10^{-12} W/m^2$ odpowiada minimalnemu poziomi słyszalności.

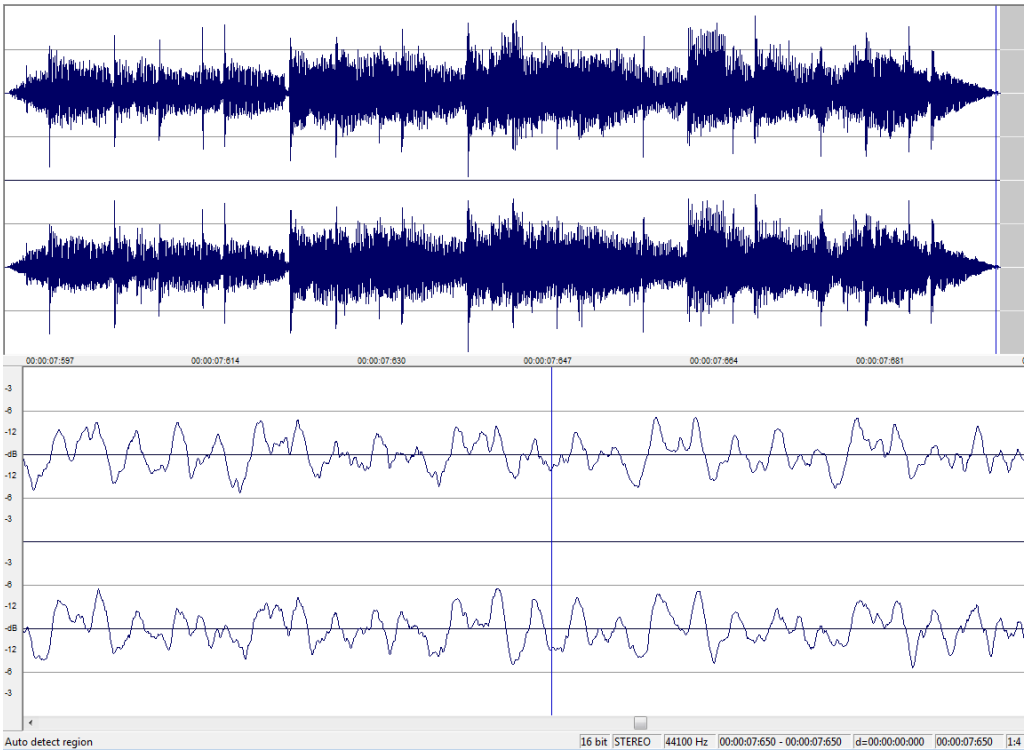
Dźwięk opisywany jest częściej za pomocą subiektywnych wrażeń percepcji dźwięku, czyli cech psychoakustycznych, takich jak:

- wysokość dźwięku prostego czy wielotonu harmonicznego, zależna od częstotliwości podstawowej;
- barwa dźwięku złożonego, pozwalająca odróżnić brzmienie instrumentów czy głosu, wynikająca z ukształtowania widma częstotliwościowego sygnału – rozkładu częstotliwości i intensywności poszczególnych tonów;

[3] Elementarny dźwięk pojedynczej sinusoidy o ustalonej częstotliwości, może być wytworzony przez kamerton

[4] Jest to wieloton harmoniczny, czyli drganie będące sumą sinusoid o częstościach będących wielokrotnościami częstości podstawowej lub też wieloton nieharmoniczny, czyli drganie będące sumą drgań o nieregularnym widmie

[5] Natężenie dźwięku jest miarą energii fali akustycznej przepływającej w jednostce czasu przez jednostkowe pole powierzchni prostopadłej do kierunku rozchodzenia się fali (jednostką jest W/m^2)



Rysunek 1.16: Przykładowe zapisy dźwięku – kilkusekundowy zapis utworu Stinga (u góry), z jego dokładniejszą analizą rozciągniętego fragmentu na dole (wykorzystano narzędzie Wavosaur, http://www.wavosaur.com/).

- czas trwania dźwięku, zależny od czasu występowania drgań źródła dźwięku, a także od ewentualnego pogłosu (tj. stopniowego zanikania energii odbieranego dźwięku, związane z występowaniem dużej liczby fal odbitych);

- głośność dźwięku, czyli wrażenia słuchowego pozwalającego różnicować dźwięki cichsze i głośniejsze, które zależy od natężenia dźwięku, ciśnienia, struktury widmowej, czasu trwania (jednostki – fony i sony).

Czułość ucha potwierdzająca skuteczność słyszenia zależna jest w dużym stopniu od częstotliwości odbieranego dźwięku. Zakres najlepszej słyszalności to przedział częstotliwości od ok. 1 do 5kHz, przy czym minimum czułości wypada na około 3,5kHz. Zakres natężenia odbieranych dźwięków (tj. dynamika słuchu) wynosi ok. 120dB i jest określony jest przez dwie wartości graniczne. Dolna granica słyszalności (próg czułości słuchu) odpowiada minimalnej wartości ciśnienia akustycznego, przy której można uzyskać ledwie postrzegalne wrażenie słyszenia dźwięku o określonej charakterystyce częstotliwościowej. Górna granica słyszalności, analogicznie zależna od częstotliwości, jest określona przez minimalną wartość ciśnienia akustycznego dźwięku, przy której ucho reaguje bólem. Dźwięki o

poziomie natężenia przekraczającym 140dB mogą nawet uszkodzić słuch.

Dla porównania, dynamika natężenia dźwięku mierzona w czasie kilkugodzinnego koncertu dużej orkiestry symfonicznej nie przekracza 80dB, zaś dynamika poszczególnych instrumentów nie przekracza 50 dB. Maksymalna dynamika mowy to 50dB.

1.1.3 Inne dane

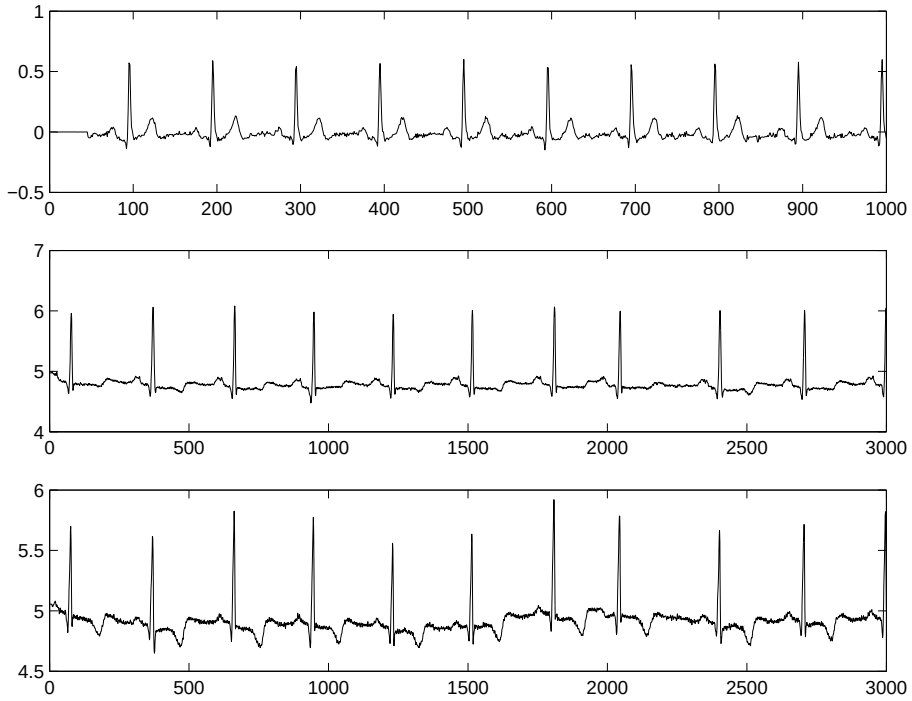
Wśród innych danych multimedialnych, najczęściej spotykane są zbiory tekstowe, zawierające zwykle ciągi bajtowo zdefiniowanych znaków w kodzie ASCII (podstawowym lub rozszerzonym), dwubajtowych znaków kodu UCS-2, 24-bitowych znaków kodu Unicode czy innych [9]. Znaki te łączą się w pojedyncze słowa, zdania, wyrażenia uzupełniane znakami formatującymi, określającymi sposób ich interpretacji – np. jako opracowania powstające w różnych edytorach tekstu, rozkazy w danym języka programowania, wektory danych arkuszy kalkulacyjnych, słowa kluczowe, tagi, itp.

W niektórych przypadkach informacje mogą przenosić dodatkowe dane sterujące, przesyłane w ramach protokołów interakcyjnych pomiędzy nadawcą i odbiorcą, dane tekstowe objaśniające przekaz audio-wideo, dane definiujące sposób interpretacji struktur danych, opisujące relacje bazodanowe, itd.

Przykładem danych pomiarowych, rejestrowanych za pomocą specjalnych czujników, elektrod, układów odbiorczych, będących rejestracją zmian w czasie określonych wielkości fizycznych może być zapis EKG elektrycznej aktywności serca, mierzony za pomocą kilku odprowadzeń – rys. 1.17. Innym rodzajem danych są dane rejestrujące aktywność sejsmiczną w różnych punktach naszej planety, zapis temperatury i ciśnienia centrów meteorologicznych, liczba rozpadów jąder zachodzących w jednostce czasu, w określonym materiale z izotopem promieniotwórczym, itd.

1.1.4 Integralność strumieni przekazu

Wielość strumieni przekazu multimedialnego zakłada ich wzajemną zależność, komplementarność, uzupełnienie czy dopełnienie przesyłanej treści. Istotna są więc wzajemne powiązania, synchronizacja, zbieżność czasowa, zachowanie względności, odpowiedniej kolejności czy też hierarchii w strukturze odtwarzanej treści. Przykładowo, typowy film zawiera ścieżkę dźwiękową, która jest dopełnieniem treści wizyjnej, a czasowa synchronizacja tych strumieni jest warunkiem koniecznym dobrego odbioru filmu. Na ile dźwięk jest istotny, zależy od konkretnej produkcji. Czy jest mało ważnym uzupełnieniem – tłem dla scen rozgrywających się przed oczami widza, czy też stanowi kluczowy element przekazu – kiedy wypowiadane treści lub muzyka decydują o sile wyrazu całego filmu, przekaz powinien dawać efekt synergii obu strumieni.



Rysunek 1.17: Przykładowe przebiegi EKG zapisu elektrycznej aktywności serca – po 4000 próbek dla każdego odprowadzenia.

W przypadku materiału edukacyjnego, tekst przesyłany razem ze ścieżką dźwiękową pozwala uchwycić charakterystyczne elementy wymowy, błędy w wymowie, dykcji, niedopowiadanie końcówek, itp. – wtedy względna rola każdego ze strumieni stwarza konieczność łącznej analizy obu strumieni, a istota przekazu polega na zestawianiu czy różnicowaniu treści wyrażonej tekstem oraz zapisaną formą wymowy.

Kilka kamer monitoringu obiektu przemysłowego jest źródłem strumieni synchronizowanych w czasie, częściowo zależnych w treści, np. poprzez podobny wpływ pory dnia, pogody, niektórych wspólnych elementów dużego obiektu, ale w niektórych elementach treści - niezależnych (np. osoby pojawiające się w polu widzenia jednej kamery, są niewidoczne w innych).

Problem czasowej synchronizacji strumieni przekazu multimedialnego musiał zostać rozwiązany przy opracowaniu standardów multimedialnych rodziny MPEG. Zależności danych czy treści wykorzystane zostały w metodach kodowania, np. w estymacji i kompensacji ruchu w sekwencjach wizyjnych, wzajemnej predykcji wartości próbek przy wielokanałowym zapisie dźwięku. Podejmowane są też próby multipleksowania strumieni informacji o tym samych charakterze, pochodzących z różnych źródeł, a niekiedy próby łącznego kodowania danych z różnych strumieni, po opisaniu ich za pomocą źródeł zintegrowanej informacji o

nowym, wspólnym alfabecie kodowanych danych. Niekiedy w tym celu stosowane są binarne kodery arytmetyczne o adaptacyjnych algorytmach szybkiej modyfikacji modelu źródła, gdzie bez względu na charakter danych, wszystkie strumienie traktowane są jako ciągi bitowe (dostarczane przez źródła nad alfabetem binarnym).

Warto w tym kontekście wspomnieć o pojęciu multimodalności, które w sposób bezpośredni można odnieść do multimediów. Bogactwo świata rzeczywistego, sposób istnienia rzeczy lub zachodzenia zjawisk jest odbierany przez człowieka za pomocą szeregu zmysłów, na kilka uzupełniających się sposobów. Postrzeganie świata obejmuje jakby kilka jego modalności, których połączenie daje synergiczny obraz rzeczywistości. Różnorakie aspekty zachodzących zjawisk znajdują odbicie w multimodalnych systemach rejestracji na potrzeby multimedialnego przekazu informacji kształtowanego potrzebami odbiorcy. Jednym z istotnych zastosowań jest obrazowanie multimodalne w medycynie, łączące np. morfologię struktur z obrazów tomografii komputerowej z ich funkcjonalnością opisaną w rezultatach badań PET[6].

[6] *Positron Emission Tomography*, czyli pozytonowa tomografia emisyjna dająca przestrzenny obraz radiofarmaceutyku o krótkim czasie połowicznego rozpadu izotopów, wprowadzonego do organizmu.

1.2 Rejestracja danych

Wierny zapis (rejestracja) stanu rzeczywistości, unikalnej, będącej źródłem ważnych informacji w przekazie, rzeczywistości z natury wielomodalnej, jest podstawowym elementem sprawnych systemów multimedialnych.

Rejestracja danych multimedialnych jest bardzo ważnym etapem pozyskiwania informacji, przekazywanej w kolejnych etapach do odbiorcy. Różnorakie aspekty zjawisk fizycznych o odmiennej charakterystyce są rejestrowane za pomocą dostosowanych czujników, rejestratorów, złożonych systemów akwizycji, czyli ogólnie specjalistycznych urządzeń pozyskiwania danych. Zapewnienie możliwie wysokiej jakości pozyskiwanych danych oraz wyznaczenie efektywnej ich reprezentacji decyduje często o użyteczności całej aplikacji multimedialnej.

Zapis rejestrowanej treści powinien być wierny, specyficzny, dogodny w dalszej obróbce oraz dostosowany do przewidywanych form odtwarzania czy ogólniej użytkowania. Warto też uwzględnić przewidywane formy kształtowania przekazu informacji, możliwą regulację jakości danych czy też selekcji treści użytecznych.

Różnice pomiędzy systemami akwizycji danych dotyczą przede wszystkim takich dwóch podstawowych aspektów jak:

- fizyczne podstawy wykorzystywanych w rejestracji zjawisk:
  - wykorzystanie właściwych zjawisk fizycznych (pomiaru cech obiektów), odpowiednich materiałów, zasad i innych uwarunkowań pomiaru;
  - wybór właściwych technologii, konstrukcja urządzeń i systemów;
  - projektowanie zestawu czujników/detektorów wraz z mechanizmami odczytu danych;
  - kontrola jakości rejestracji;
- zasady uzyskania sygnałów cyfrowych:
  - dyskretne, przestrzenno-czasowe struktury rejestracji danych;
  - przetworniki A/C;
  - mechanizmy próbkowania, kwantyzacji i kodowania;
  - formowanie/rekonstrukcja sygnału rejestrowanego;
  - wstępne przetwarzanie, ustalanie reprezentacji wyjściowej.

Poniżej dokonano krótkiej charakterystyki zagadnienia rejestracji obrazów i dźwięku.

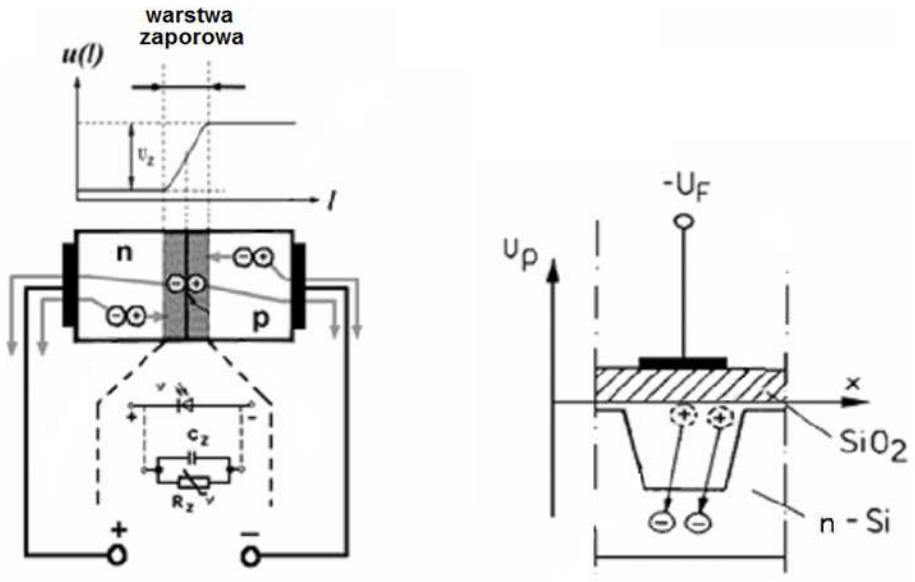
1.2.1 Obraz

Typowe urządzenia służące do rejestracji obrazów, tj. konwersji energii promieniowania optycznego (obrazu optycznego) na energię elektryczną sygnału wizyjnego

wykorzystują podstawowe zjawisko fizyczne — efekt fotoelektryczny uwalniania elektronów (przenoszenie z pasma podstawowego do pasma przewodzenia) z atomów poprzez absorpcję energii fotonów. Jeśli uwolnione elektrony pozostają w materiale detektora, mamy do czynienia z efektem wewnętrznym, zwykle wykorzystywanym w rejestratorach obrazów naturalnych. Efekt fotoelektryczny zewnętrzny, polegający na uwalnianiu elektronów np. z materiału fotokatody, jest wykorzystywany we wzmacniaczach obrazu, które są ważnym elementem rejestracji obrazów za pomocą promieniowania rentgenowskiego czy gamma.

Sam proces rejestracji obejmuje zwykle konwersję energii optycznej na elektryczną w materiale światłoczułym. Jako materiał światłoczuły wykorzystuje się zazwyczaj półprzewodniki (najczęściej monokrystaliczny krzem) o możliwie małej wartości prądu ciemnego (brak wolnych ładunków w paśmie przewodzenia przy braku oświetlenia). Są to struktury wielowarstwowe o spolaryzowanych zaporowo złączach, gdzie rejestrowany jest ruch uwolnionych elektronów oraz dziur.

Pojawiający się ładunek elektryczny jest zbierany i gromadzony za pomocą przyłożonego pola elektrycznego w punktach obrazu - pikselach o konkretnych wymiarach fizycznych. Istotnym elementem jest czas gromadzenia i odczytu ładunku z poszczególnych pikseli. Akumulacja ładunku w strukturze materiału światłoczułego także w czasie pomiędzy odczytami pozwala zdecydowanie zwiększyć liczbę ładunków odczytywanych z poszczególnych pikseli. Wykorzystuje się do tego tzw. kondensatory złączowe – bipolarny lub MOS (zobacz rys. 1.18).



Rysunek 1.18: Elementy przetwarzająco-akumulujące stosowane w analizatorach obrazów: kondensator bipolarny – złącze p-n (po lewej), fotodioda oraz kondensator MOS (źródło: na podstawie rysunku z [13]).

Systemy akwizycji obrazów są silnie zróżnicowane, ale w ogólności można je

podzielić na:

- kamery analogowe ze zmienna szybkością rejestracji obrazu, niskoszumowe, tanie, z sygnałem ucyfrawianym za pomocą urządzeń-kart typu *framegrabber*;
- cyfrowe aparaty fotograficzne i kamery, z wysokoczułymi obiektywami, macierzami CCD[7] lub CMOS[8] (TFT[9]), szybkimi układami sczytywania próbek obrazu, gromadzenia, a często - kodowania;
- skanery z wysoką zdolnością rozdzielczą, liniową charakterystyką w środkowym, możliwie szerokim zakresie przenoszenia kontrastu;
- systemy specjalistyczne:
  - sensory teledetekcyjne[10] – lotnicze czujniki obrazowe (zdjęcia ziemi z wysokości do 35 km) oraz satelitarne czujniki obrazowe,
  - mikroskopy optyczne[11];
  - medyczne systemy obrazowania – rentgenowski, USG, tomografii komputerowej i magnetycznego rezonansu jądrowego, medycyny nuklearnej (SPECT, PET);
  - inne.

Rejestratory obrazów można także podzielić na a) aktywne, wysyłające własną wiązkę promieniowania i rejestrujące jej odbicie (np. urządzenia radarowe i laserowe, lidar) oraz b) pasywne, rejestrujące promieniowanie zewnętrzne (np. aparaty fotograficzne, kamery). Ze względu na szerokość zakresu, w jakim rejestruje się promieniowanie elektromagnetyczne, wyróżnia się rejestratory szerokopasmowe i wąskopasmowe. Dodatkowo, jeśli obraz jest jednocześnie zapisywany w kilku, kilkunastu bądź kilkudziesięciu zakresach promieniowania, wówczas takie analizatory określa się mianem wielospektralnych lub hiperspektralnych (np. w teledetekcji).

Warto zauważyć, że podstawowym źródłem strumienia wizyjnego w przekazie multimedialnym jest kamera. "Sercem" kamery jest przetwornik wizyjny - analizator obrazu jako urządzenie służące bezpośrednio do rejestracji obrazów. W

[7] Charge Coupled Device

[8] Complementary Metal Oxide Semiconductor

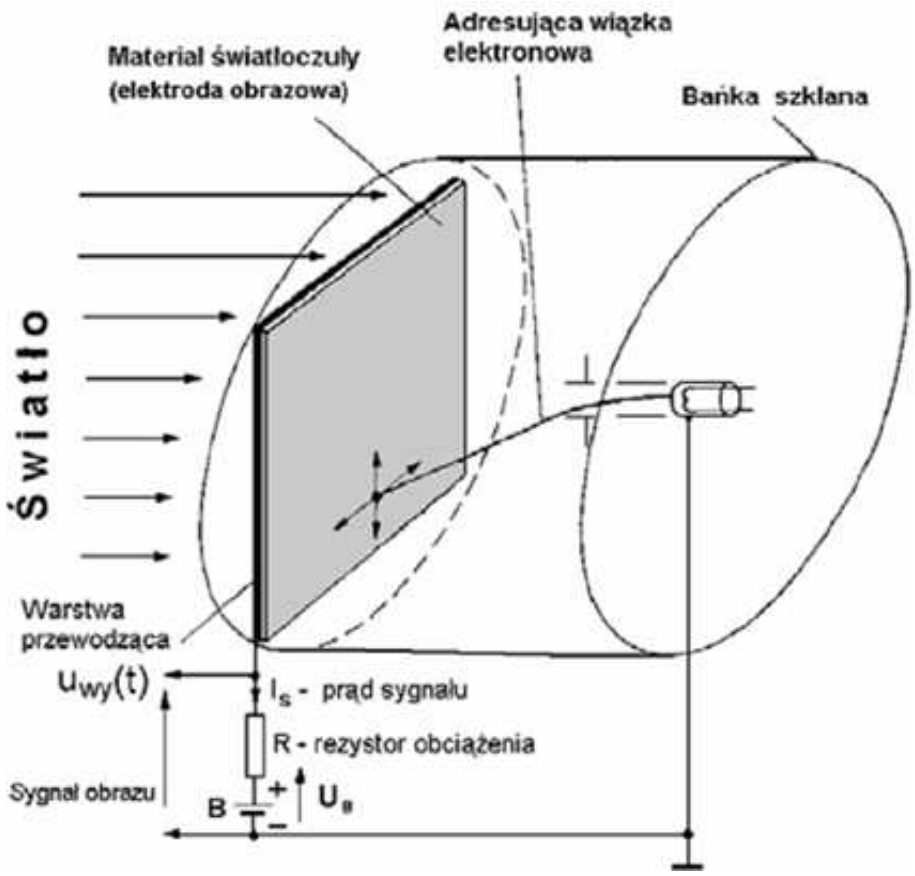
[9] Thin-Film Transistor

[10] Teledetekcja (*remote sensing*) to badanie wykonane z pewnej odległości (zdalne) w celu rozpoznawania obiektów i zjawisk poprzez wykrywanie i analizę promieniowania elektromagnetycznego w różnym zakresie widmowym. Badania teledetekcyjne są wykonywane z samolotów, przestrzeni kosmicznej lub z powierzchni ziemi.

[11] Mikroskopy optyczne służą silnemu powiększeniu obrazów, wykorzystując światło (naturalne lub sztuczne, niekiedy spolaryzowane) przechodzące przez specjalny układ optyczny składający się zazwyczaj z zestawu od kilku do kilkunastu soczewek optycznych. Do rejestracji obrazów stosowane są wysokiej klasy aparaty i kamery cyfrowe. Uzyskiwane powiększenia sięgają 3500x.

analizatorach realizowane są trzy podstawowe procesy: konwersji energii, gromadzenia energii elektrycznej oraz jej odczytu z zachowaniem informacji o położeniu.

Reprezentantem analogowego sposobu adresowania detektora w celu odczytu obrazu jest lampa analizująca z anodą w postaci elektrody obrazowej oraz katodą emitującą wiązkę elektronów. Większemu natężeniu promieniowania oświetlającego elektrodę obrazową, czyli detektor towarzyszy większa ilość uwolnionego ładunku, co może być także analizowane jako zmniejszenie dużej rezystancji skrośnej (w objętości) materiału nieoświetlonego. Poprzez przemiatanie wiązką elektronów (adresującą, sterowaną za pomocą pół magnetycznych i elektrycznych) powierzchni detektora wykonanej z materiału światłoczułego, wytwarzany jest sygnał proporcjonalny do rezystywności danego obszaru - rys. 1.19.



Rysunek 1.19: . Orientacyjny szkic opisujący lampę analizującą wraz ze schematem jej działania (źródło: na podstawie rysunku z [13]).

Cyfrowe rejestratory obrazów

Cyfrowy analizator obrazu charakteryzuje się takimi parametrami jak czułość widmowa (rejestracja określonych przedziałów długości fal), czułość świetlna (re-

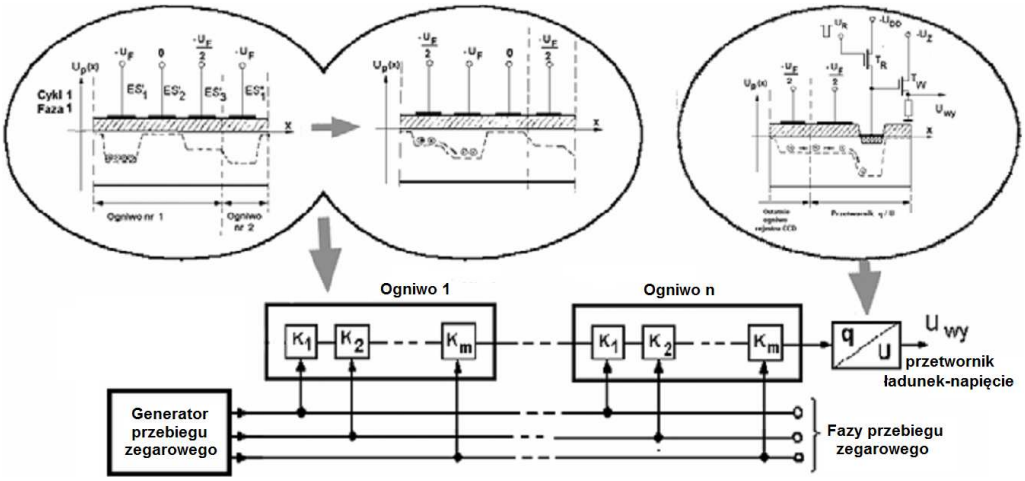
jestracja sygnałów świetlnych o możliwie małym natężeniu), zdolność rozdzielcza (możliwość rozdzielenia na obrazie dwóch leżących blisko siebie punktów), bezwładność (zdolność do rejestracji szybkich, dynamicznych scen). Stosowanym dziś powszechnie analizatorem (m.in. w kamerach i aparatach fotograficznych) pozwalającym uzyskać cyfrowe obrazy są zestawy elementów ze sprzężeniem ładunkowym CCD, które zastąpiły w większości popularnych zastosowań analogową lampę analizującą (przytłaczającą swoim rozmiarem i wagą, chociaż pozwalającą na uzyskanie obrazu jakości HDTV[12]). Jakkolwiek detektor obrazu składany jest z szeregu linijek CCD, które stanowią *de facto* jednowymiarowy detektor analogowy sygnału i wymagają taktowanego zegarem próbkowania w procesie odczytu, to pełna, bezpośrednia, układowa integracja procesu próbkowania i kwantyzacji pozwala traktować detektor CCD obrazu jako cyfrowy z punktu widzenia wielu zastosowań. Koszt rejestratorów CCD w znacznym stopniu zależy od częstotliwości ich taktowania.

Rejestratory CCD budowane są w strukturze monolitycznej z naniesioną półprzezroczystą elektrodą polikrzemową, jako matryca fotoczułych komórek, w których przy odpowiedniej polaryzacji gromadzony jest ładunek, zależnie od energii oświetlenia. Zderzenie fotonu z atomem komórki światłoczułej może spowodować przeskok elektronu na wyższą powłokę lub uwolnienie nośnika ładunku — fotoefekt wewnętrzny. Uwolnione nośniki są gromadzone w kondensatorach, a następnie przesuwane za pomocą impulsów elektrycznych, wyłapywane, a powstający sygnał wzmacniany, kondensatory opróżniane. Ilość nośników zebranych w jednostce czasu odzwierciedla natężenie padającego światła.

Odczyt cyfrowy z dyskretnym adresowaniem uzyskiwany jest za pomocą rejestrów przesuwnych, tj. kaskady komórek pamięci z sekwencyjnie przesuwaną, zapamiętaną informacją w odstępach czasowych wyznaczanych przebiegiem taktującym, jednakowym dla wszystkich komórek. Komórki pamięci rejestru przesuwającego są grupowane w zwykle równoliczne ogniwa (najczęściej 2-4 komórek). Uwolnione nośniki (ładunki), które zostały zgromadzone w poszczególnych punktach obrazu ładunki (w kondensatorach), są przesuwane za pomocą impulsów elektrycznych w kierunku wzmacniacza ładunkowego, przetwarzającego sygnał ładunkowy na napięciowy (rys. 1.20). Zostają one tam wyłapane, a powstający sygnał napięciowy – wzmocniony, zaś kondensatory – opróżnione. Ilość nośników zebranych w jednostce czasu odzwierciedla natężenie padającego światła.

Dokładniej, jeżeli w czasie rejestracji oświetlenia (tzw. okresie akumulacji) jedynie pierwsze elektrody w ogniwach rejestru zostaną spolaryzowane napięciem stałym, wówczas fotoładunki będą wytwarzane i akumulowane jedynie pod elektrodami ES1 każdego ogniwa. Stąd pod pierwszymi elektrodami wszystkich ogniw wytworzy się ładunkowa replika rozkładu oświetlenia wzdłuż rejestru (linii

[12] *High Definition TV – telewizja wysokiej rozdzielczości, w pełnej wersji 1920×1080 pikseli bez przeplotu.*



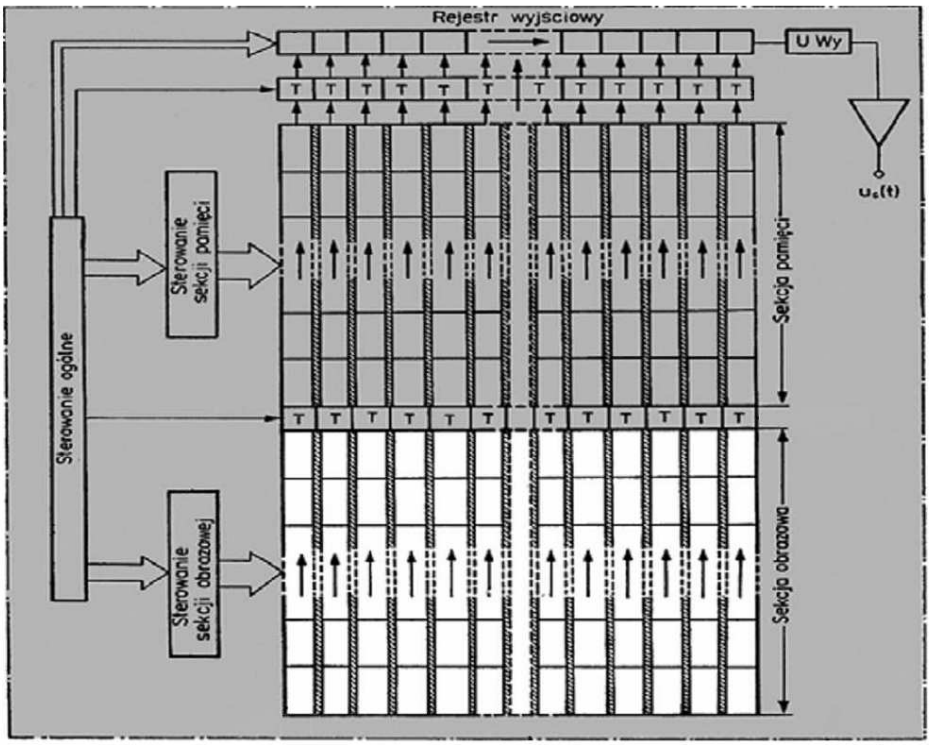
Rysunek 1.20: Odczyt informacji obrazowej w macierzach CCD: trójfazowy rejestr przesuwny ze studnią potencjału, zegarem taktującym i zbierającym sygnał przetwornikiem ładunkowym (źródło: na podstawie rysunku z [13]).

analizy). Załączenie przebiegu zegarowego i poprzez zmianę potencjału komórek sekwencyjne przesuwanie zgromadzonych fotoładunków do wyjściowego przetwornika ładunek/napięcie pozwoli uzyskać na wyjściu napięcie proporcjonalne do liczby zebranych ładunków. Ścisłe powiązanie liczby okresów przebiegu zegarowego z liczbą koniecznych przesunięć (liczbą ogniw w linii) precyzyjnie lokalizuje miejsce pobrania ładunku. Rejestr zbudowany z odpowiedniej linii ogniw może stanowić więc analizator całej linii (wiersza) obrazu.

Pewnym problemem jest możliwość akumulacji dodatkowego ładunku w czasie przesuwu, tzw. efekt "zaplamienia". Podczas przesuwu fotoładunków przez kolejne komórki rejestru dodawane są nowe, pasożytnicze ładunki zbierane podczas ich "transportowej polaryzacji" zakłócając faktyczny rozkład natężenia oświetlenia w linii. Pojawia się smużenie oraz charakterystyczne artefakty - plamki. Zaplamienie jest strukturalną wadą rejestru CCD. Można je wyeliminować jedynie przez zaciemnienie (zasłonięcie) przetwornika podczas transferu ładunków, np. za pomocą przesłony mechanicznej lub ciekłokrystalicznej. Oba sposoby spowalniają jednak proces akwizycji i nie zawsze można je zastosować. W takich przypadkach natężenie zaplamienia można jedynie zmniejszyć minimalizując czas transferu ładunków, a więc zwiększając częstotliwość przebiegu zegarowego, co podwyższa koszt przetwornika.

Aby uzyskać obraz cyfrowy za pomocą analizatora linii, konieczne jest skanowanie rejestrowanego obrazu linią przesuwaną w kierunku prostopadłym do linii analizatora, np. w skanerach, telefaksach, kopiarkach czy kamerach do meteorologicznych zdjęć satelitarnych. W celu równoczesnej akwizycji całego obrazu konieczne jest uzupełnienie zestawu analizatorów linii rejestrem wyjściowym ze

wzmacniaczem ładunkowym, zbierającym cyklicznie przesuwane ładunki zestawu analizatorów linii w kierunku prostopadłym do kierunku przesuwu. Ze względu na "zaplamienie" uzyskiwanego obrazu konieczna jest konstrukcja wykorzystująca dwa zestawy analizatorów linii - otwartych na światło, akumulujących ładunki gromadzone w naświetlanym polu obrazu (sekcja obrazowa) oraz zamkniętych na światło, magazynujących ładunki oczekujące na sczytanie w rejestrze wyjściowym (sekcja pamięci) - rys. 1.21.



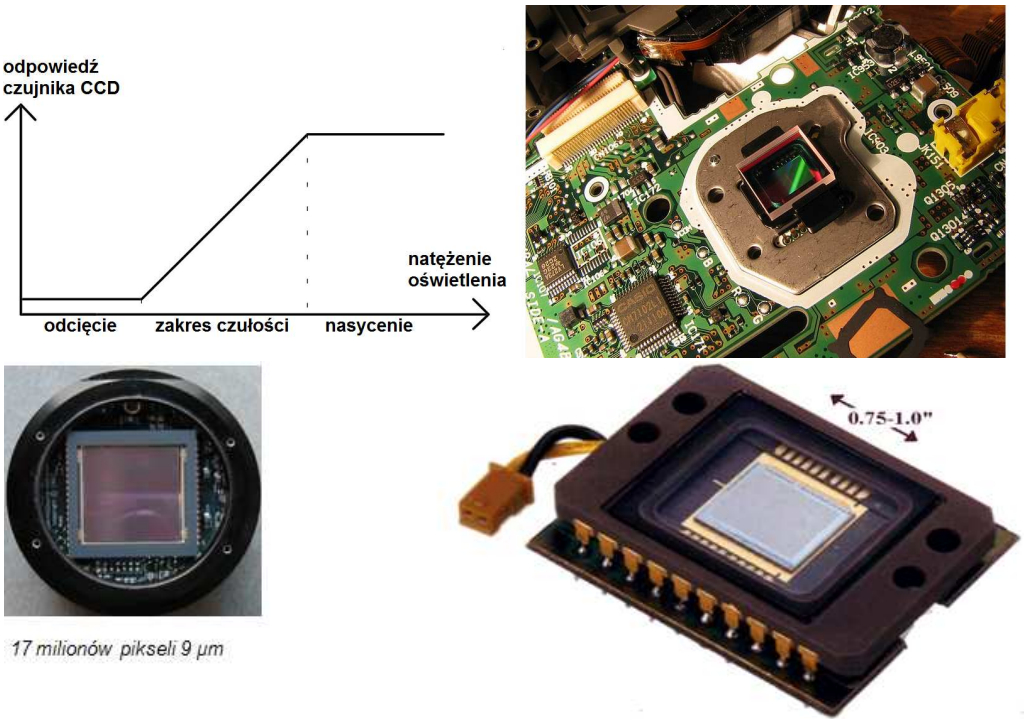
Rysunek 1.21: Koncepcja budowy analizatora obrazów CCD z przesuwem ramki, gdzie: T – bramki transferowe; na biało zaznaczono naświetlane pole obrazu, składające się z pionowo ustawionych – jako kolumny – analizatorów linii (sekcja obrazowa); powyżej - pole sekcji pamięci składające się z identycznych analizatorów linii, zestawu rejestrów przesuwających zarejestrowany w poprzednim kroku obraz do rejestru wyjściowego, sekwencyjnie sczytującego kolejne wiersze obrazu (źródło: na podstawie rysunku z [13]).

W pierwszym etapie odczytu całego obrazu następuje akumulacja fotoładunku w sekcji obrazowej. Elektrody sterujące ES1 (lub kolejne) wszystkich ogniw wszystkich analizatorów kolumn sekcji obrazowej są polaryzowane napięciem stałym, natomiast pozostałe elektrody ogniw pozostają niespolaryzowane – tworzona jest ładunkowa replika rejestrowanego obrazu. W tym czasie ładunki umieszczo-

ne poprzednio w rejestrach pamięci są sekwencyjnie przesuwane i za pomocą drugiego zestawu bramek transferowych przekazywane do rejestru wyjściowego. Sczytywany (adresowany) jest obraz zarejestrowany w poprzednim etapie akumulacji.

Drugi etap to przesuwanie ładunku w obu sekcjach sterowanych tym samych zegarem taktującym. Jest to czas wygaszenia pola rejestrowanego obrazu (brak akumulacji). Fotoładunki zgromadzone w ogniwach każdego analizatora kolumny sekcji obrazowej są w całości transferowane do odpowiadającej mu linii sekcji pamięci za pomocą bramek transferowych. Przesuw zarejestrowanego pola obrazu do sekcji pamięci ładunku kończony jest zdjęciem sygnału zegarowego oraz zamknięciem bramek transferowych.

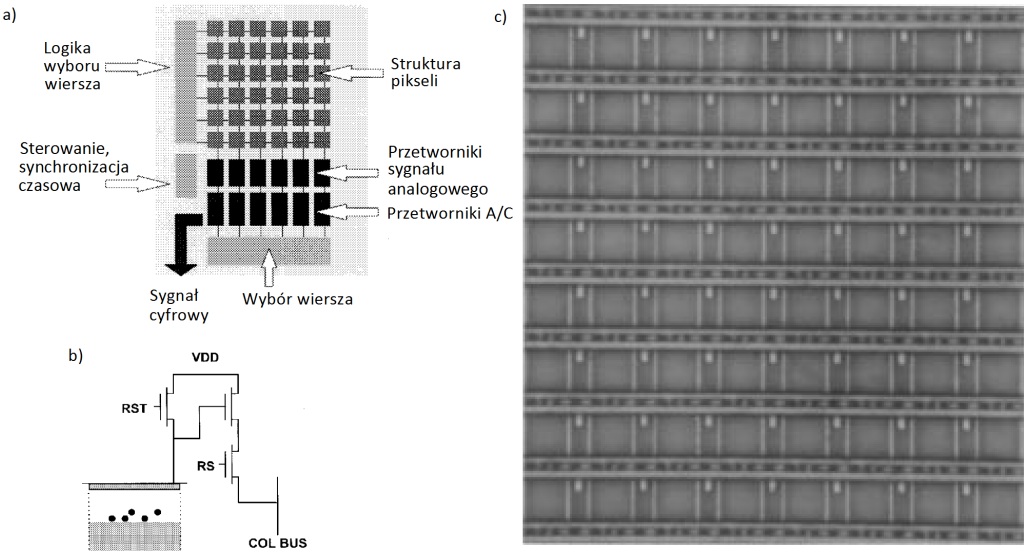
Liniową charakterystykę czujnika obrazu oraz przykładowe matryce CCD przedstawiono na rys. 1.22.



Rysunek 1.22: Analizator obrazu z matrycą CCD: liniowa odpowiedź czujnika CCD na pobudzenie świetlne w zakresie aktywnej pracy rejestratora (u góry po po lewej), sensor CCD zamontowany w kamerze cyfrowej (u góry po prawej) oraz inne przykłady czujników CCD (źródło: własne, Wikipedia, Internet).

Alternatywnym rozwiązaniem są detektory CMOS, stosowane m.in. w aparatach cyfrowych i kamerach internetowych. Detektory te należą do grupy czujników z aktywnymi pikselami (*active pixel sensors*), bazują na w pełni dyskretnej strukturze powierzchniowej i są konkurencyjne w stosunku detektorów CCD. Ma-

tryce CMOS organizowane są w postaci dwuwymiarowej struktury TFT, gdzie każdy element tranzystorowy ma aktywną elektrodę powierzchniową zbierającą sygnał z określonego obszaru pojedynczego piksela. Jako element światłoczuły wykorzystywana jest sprzężony ze strukturę TFT zestaw fotodiod - rys. 1.23.



Rysunek 1.23: Przykładowe rozwiązanie detektora CMOS: a) ogólny schemat struktury; b) czujnik aktywnego piksela z fotodiodą, RS - sygnał selekcji piksela, RST - tranzystor do resetu fotodiody; c) widoczna struktura detektora 1024×1024 aktywnych pikseli (tranzystor CMOS wzmacnia sygnał z fotodiody w ramach komórki piksela) o fizycznym rozmiarze piksela $11{,}9 \mu m$ (na podstawie pracy [10]).

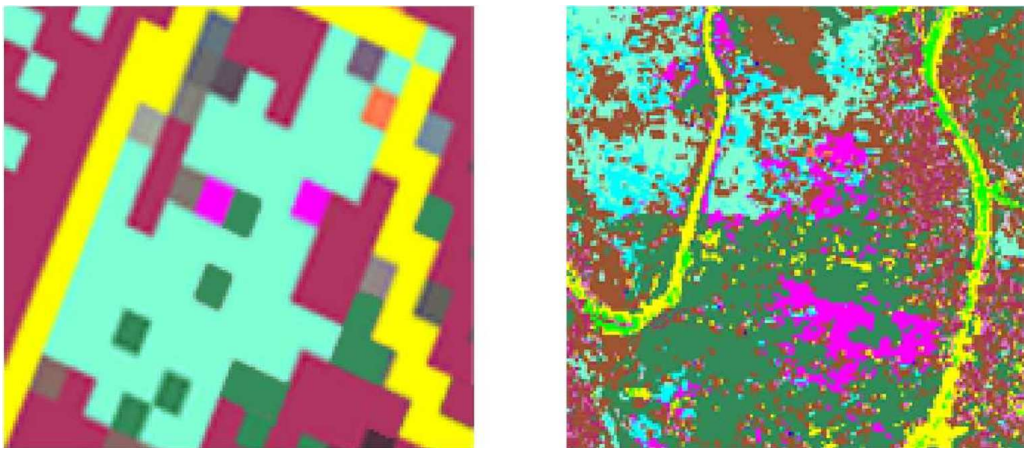
Dzięki dyskretnej w obu wymiarach strukturze elementów gromadzących ładunek, odczyt przestrzennej informacji obrazowej jest prostszy, szybszy, selektywny (można odczytać tylko wybrany region), z redukcją zakłóceń typu "zaplamienie". Jednocześnie niski koszt i niewielki pobór mocy stanowią o atrakcyjności takich detektorów. Wśród wad można wymienić pewne ograniczenia w zakresie uzyskiwanej światłoczułości i dynamiki rejestrowanego sygnału, a także pewne niejednorodności odczytu z całego pola obrazu i stosunkowo duży prąd ciemny. Jednak kolejne udoskonalenia detektorów klasy CMOS prowadzą do coraz wyższej jakości uzyskiwanych obrazów.

Systemy specjalistyczne

Jako przykład współczesnego rozwiązania w obszarze **sensorów satelitarnych** można wymienić sensor TM (emphThematic Mapper) rejestrujący obraz w siedmiu kanałach spektralnych. Jego następca -– sensor ETM (*Enhanced Thematic Mapper*), umieszczony na pokładzie satelity Landsat 7 (rozdzielczość naziemna 15m w kanale panchromatycznym, 30m w kanałach spektralnych i 60m w kana-

le termalnym). Rozdzielczość naziemna danych dostarczanych z innego satelity (SPOT) zmienia się od 2,5m do 20m. Dla rejestratorów lotniczych rozdzielczość naziemna jest wyższa. Przykładowo, niemieckie sensory DAIS oraz ROSIS przy 79 kanałach spektralnych pozwalają uzyskać rozdzielczość odpowiednio 5m i 1m [11].

Sensory hiperspektralne umieszczone na pokładach satelitów cechują się aktualnie niską rozdzielczością naziemną. Przykładowo, sensor MODIS (http://modis.gsfc.nasa.gov/data/) rejestrujący obrazy w 36 zakresach widma (pomiędzy 0,405 i 14,385 μm długości fali), zainstalowany na amerykańskim satelicie Terra, charakteryzuje się rozdzielczością wynosząca zależnie od zakresu spektralnego od 250 m, przez 500 m do 1000 m. Z kolei dane z sensora AVIRIS umieszczonego na samolocie NOAA Twin Otter to obrazy rejestrowane w 224 zakresach widma o rozdzielczości przestrzennej 2,2m przy robieniu zdjęć z wysokości 2000m oraz 20m – z wysokości 20.000m [12] - zobacz rys. 1.24. Efektem doskonalenia systemów wielorozdzielczych jest rozdzielczość uzyskana za pomocą rosyjskich kamer KVR-1000 i KVR-3000, sięgająca 2m (technika skanowania zdjęć analogowych). Panchromatyczne wysokorozdzielcze zdjęcia satelitarne, rejestrowane cyfrowo, uzyskiwane z satelity IKONOS mają rozdzielczość 1m w punkcie podsatelitarnym, natomiast obrazy z Quick Birda2 – 0,61m.



Rysunek 1.24: Zdjęcia satelitarne, spektralne o małej – 20m (z lewej) i dużej – 2m rozdzielczości naziemnej (na podstawie rysunku z [12]).

Szczególnie istotne parametry zdjęć lotniczych za pomocą fotogrametrycznych[13] aparatów (kamer) fotograficznych, wyposażonych w specjalne obiektywy

[13]Fotogrametria to dziedzina nauki i techniki zajmująca się odtwarzaniem kształtów, rozmiarów i wzajemnego położenia obiektów w terenie na podstawie zdjęć fotogrametrycznych (fotogramów). Typowe zastosowania to opracowywanie map, pomiary dużych obszarów i odległości, wyznaczanie wysokości obiektów. Obejmuje fotogrametrię naziemną (terrofotogrametrię) oraz fotogrametrię lotniczą (aerofotogrametrię). Zależnie od sposobu wykorzystania zdjęć rozróżnia się fotogrametrię płaską (jednoobrazową) i fotogrametrię przestrzenną (dwuobrazową),

pozbawione aberracji, to czułość i rozdzielczość przestrzenna.

Rozdzielczość uzyskiwanych obrazów fotograficznych zależy od układu optycznego kamery oraz zdolności rozdzielczej rejestratora (błony filmowej lub detektora cyfrowego – oba te rozwiązania są stosowane). Przeciętna zdolność rozdzielcza typowych niskokontrastowych (bo takich się najczęściej używa) klisz lotniczych waha się od 100 do 150 par linii w milimetrze, co odpowiada rozdzielczości od 7500 do 11500 dpi. Możliwości skanerów wykorzystywanych do uzyskania postaci cyfrowej z negatywów są porównywalne. Zdjęcia lotnicze, wykonane najnowocześniejszymi aktualnie kamerami fotograficznymi (klisze plus skaner) przy odpowiedniej wysokości i prędkości lotu samolotu, osiągają rozdzielczość naziemną na poziomie 0,5–1cm. Z kolei za pomocą kamery cyfrowej osiągnięto rozdzielczość naziemną równą 2,5cm.

Rejestracja obrazu w kamerach cyfrowych odbywa się w dwojaki sposób: za pomocą liniowej (problemy z czasem naświetlania i rozmazem rejestrowanych obrazów) lub prostokątnej macierzy CCD. Akumulacja obrazu wykonywana jest za pomocą kilku oddzielnych macierzy (np. 4 lub 8), z których składany jest jednolity obraz o łącznym wymiarze sięgającym przykładowo 8 tys. na 14 tys. Wymiar elementów światłoczułych stosowanych w kamerach cyfrowych wynosi zazwyczaj od 12 do 14 μm, co odpowiada rozdzielczości 2500 dpi.

W przypadku **obrazów medycznych** mamy do czynienia z szeroką skalą metod i urządzeń rejestrujących obrazy. Wykorzystywane są między innymi następujące nośniki informacji oraz właściwości struktur i tkanek, które są obrazowane:

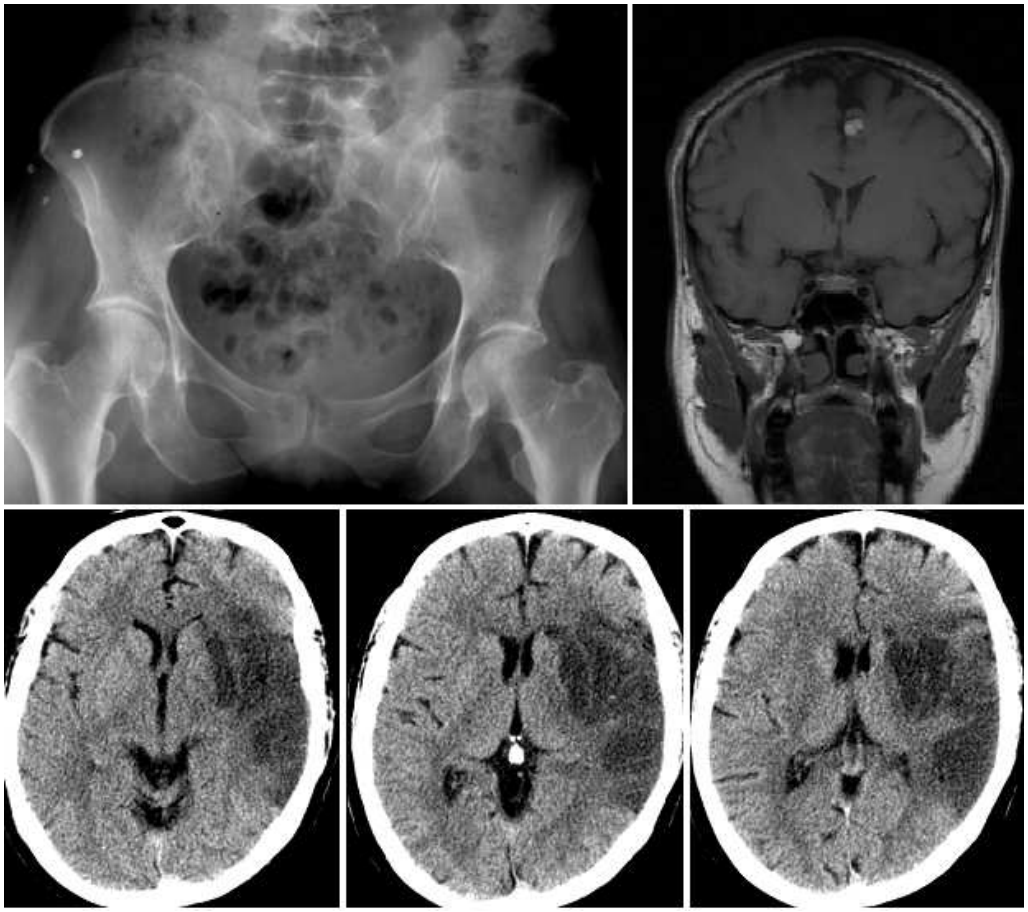
- promieniowanie rentgenowskie, zarówno w zakresie promieniowania hamowania jak i charakterystycznego, pozwalające uwidocznić przestrzenny (dwuwymiarowy) rozkład współczynnika absorpcji (uśrednionego) tkanek penetrowanych wzdłuż wiązki promieniowania; efektem są radiogramy; w przypadku tomografii komputerowej rejestrowane są poprzeczne profile danej warstwy obiektu; na ich podstawie rekonstruowany jest zbiór obrazów kolejnych warstw (uśrednionych jedynie po grubości warstwy) składających się na wolumen obrazowanych struktur wewnętrznych; efektem są tomogramy – rys. 1.25;

- fale ultrasonograficzne (mechaniczne fale o częstotliwościach ponadakustycznych) pozwalające wyznaczyć rozkład oporności akustycznej w płaszczyźnie obrazu; wykorzystywane jest zjawisko odbicia fali na granicy ośrodków o różnej oporności – rejestrowany jest czas penetracji od głowicy do powierzchni odbijającej i z powrotem oraz osłabienie amplitudy fali nośnej; efektem są charakterystyczne obrazy USG typowej projekcji B; wykorzy-

zwaną też stereofotogrametrią, w której przestrzenny obraz obiektu lub terenu uzyskuje się za pomocą stereogramu – pary zdjęć wykonanych z dwóch punktów przestrzeni.

stując zjawisko Dopplera w zbiorze ultradźwięków rozproszonych na czerwonych krwinkach szacowana jest prędkość przepływu krwi w naczyniach i komorach (na podstawie przesunięć w częstotliwościowym widmie odbieranego sygnału); efektem zaś są kolorowe mapy przepływu krwi nałożone na morfologię zobrazowanych struktur (rys. 1.1);

- fale radiowe o niskich częstotliwościach oraz silne, zewnętrzne pole magnetyczne pozwalają uporządkować, a następnie zobrazować rozkład spinów jąder atomów wodoru w danej przestrzeni 3W; metody trójwymiarowej rekonstrukcji zróżnicowań nasycenia tkanek jądrami wodoru pozwalają tworzyć obrazy w dowolnej płaszczyźnie przecięcia obiektów, dając efekt zobrazowań tomografii rezonansu magnetycznego – rys. 1.25.



Rysunek 1.25: Przykładowe obrazy medyczne: radiogram struktur kostnych miednicy (u góry po lewej), obraz głowy w tomografii rezonansu magnetycznego (u góry po prawej) oraz trzy kolejne warstwy z badania głowy w tomografii komputerowej.

Szczególnie dynamiczny rozwój notowany jest w zakresie cyfrowym metod akwizycji radiogramów. Obok tradycyjnych klisz stosowane są również coraz doskonalsze detektory cyfrowe. Wyróżnia się dwa zsadnicze ich rodzaje:

1. z konwersją pośrednią (X → światło → ładunek), zbudowane ze scyntylatora (zwykle jest to jodek cezu CsI) konwertującego promieniowanie X na widzialne, wykorzystujące następnie:

   a) macierz CCD z układem optycznym do rejestracji obrazu;

   b) macierz fotodiod TFD[14], przylegającą do warstwy cienkich tranzystorów TFT z amorficznego krzemu;

   c) obrazowe płyty fosforowe (PSP, *photostimulable storage phosphor*), magazynujące energię absorbowanych fotonów w metastabilnych stanach wzbudzonych (pułapki elektronowe w powłoce walencyjnej) atomów fosforu – wykorzystuje się najczęściej związki BaFI:Eu^{2+}, BaFCl:Eu^{2+}, BaFBr:Eu^{2+}; sczytywane zmagazynowanej energii odbywa się za pomocą stymulacji czerwonym światłem laserowym, gdzie powracające do stanu podstawowego atomy wyświecają fotony w zakresie światła widzialnego (niebieskiego) – co jest rejestrowane za mocą detektorów cyfrowych, przede wszystkim CCD;

2. z konwersją pośrednią (X → ładunek), bazujące na dielektrycznych elektrodach z warstwą amorficznego selenu, gdzie fotony promieniowania X generują pary ładunków elektron dziura dryfujące w przyłożonym polu elektrycznym; ładunki docierające do dodatniej elektrody są zbierane za pomocą struktury TFT.

Orientacyjne wymagania stawiane systemom radiografii cyfrowej przedstawiono w Tabeli 1.1. Przyjmuje się, że zdolność rozdzielcza w radiologii analogowej wynosi 15-20 pl/mm (par linii na milimetr), co odpowiada zdolności rozdzielczej sięgającej 25 μm. Uzyskanie takiej zdolności rozdzielczej w systemach cyfrowych jest sprawą być może niedalekiej przyszłości. Natomiast większe możliwości detektorów cyfrowych w zakresie dynamiki rejestrowanych obrazów wynikają przede wszystkim z liniowej charakterystyki konwersji przestrzennie rejestrowanej wartości dawki (proporcjonalnej do liczby fotonów zarejestrowanych w danym pikselu) na poziom jasności obrazu (gęstości optycznej obrazu na kliszy czy też wartości funkcji jasności przypisane pikselowi w obrazie cyfrowym) oraz większego zakresu wielkości rejestrowanych dawek przetwarzanych na obraz. Dodatkowo, w rozwiązaniu cyfrowym istnieje możliwość dodatkowej poprawy kontrastu metodami przetwarzania obrazów.

[14]TFD – *Thin Film Diode*

Parametr	Radiografia ogólna	Mammografia
rozmiar obrazu	40×40cm	18×24cm
rozmiar piksela	$\sim 150\mu$m	$50 - 100\mu$m
typowa liczba fotonów na piksel	~ 1000	~ 5000
dawka	$2,5\mu$Gy	100μGy
zakres energii	30-120keV	$\sim$ 20keV
dynamika	12bitów	12bitów
czas ekspozycji/odczytu	0,5/1s	1/5s

Tabela 1.1: Wymagania stawiane cyfrowym systemom detekcji w radiografii ogólnej i mammografii.

1.2.2 Dźwięk

Dźwięk wysokiej technicznie jakości rejestrowany jest w studiach nagrań, aczkolwiek spontaniczne zapisy niskiej klasy mikrofonami, np. aparatów komórkowych, mogą być bogate w wyjątkową, zaskakującą, informacyjną treść.

Do interesujących zagadnień należy analiza cech sygnału dźwiękowego, jak też procedura nagrywania dźwięku z wykorzystaniem odpowiedniej sali nagrań, zestawu mikrofonów, stołu mikserskiego i całego zestawu metod przetwarzania, edycji, masteringu dźwięku.

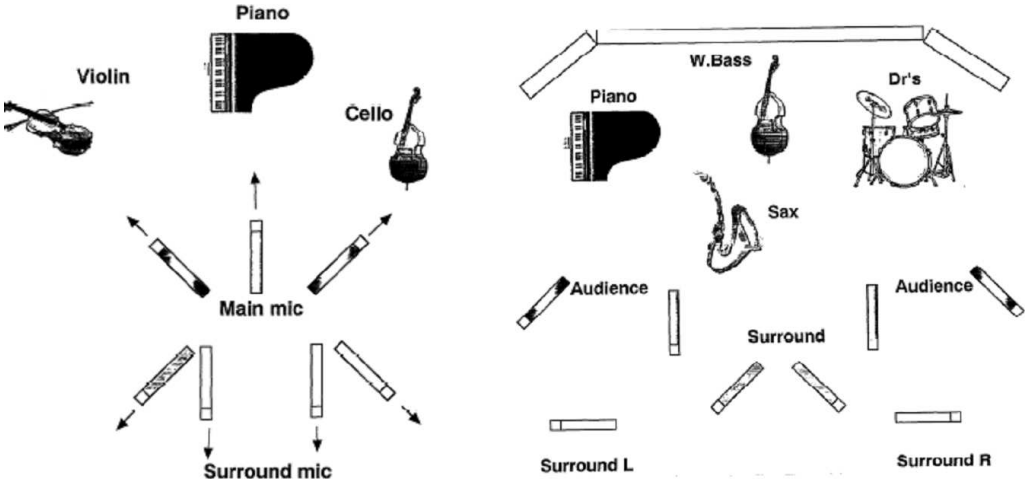
Studio nagrań to przede wszystkim sala nagraniowa oraz reżysernia. Typowe studia nagrań mają uniwersalne, bądź bardziej specjalistyczne przeznaczenie, zależenie przede wszystkim od rozmiarów sali nagrań oraz czasu pogłosu, a także rodzaju i klasy mikrofonów oraz systemów mikrofonowych. Nagrania muzyki rozrywkowej, orkiestrowej czy filmowej wyposażone są w reżysernie ze stołem mikserskim, monitorami do odsłuchu obrabianego dźwięku, zewnętrzne, dodatkowe procesory dźwięku i wzmacniacze, urządzenia do zapisu sygnałów dźwiękowych oraz archiwizacji nagrań. Wśród procesorów dźwięku warto wymienić:

- regulatory głośności, tłumiki;
- korektory barwy dźwięku: graficzne, parametryczne;
- procesory dynamiki: automatyczna regulacja wzmocnienia, kompresor, limiter, ekspander, bramka szumowa;
- procesory efektowe: pogłos, opóźnienie, studnia, zmiana wysokości dźwięku, różne efekty przestrzenne.

Szczególnie istotnym elementem systemów rejestracji dźwięku są mikrofony, które przetwarzają odbierany sygnał (energię) akustyczny na elektryczny, następnie wzmacniany i przetwarzany. Współczesne studia nagrań wykorzystują całe

zestawy mikrofonów o dobranych charakterystykach częstotliwościowych i kierunkowych (techniki mikrofonowe), skutecznie rejestrując przestrzenny rozkład fali akustycznej – rys. 1.26.

Skuteczność mikrofonu w polu akustycznym swobodnym jest to stosunek napięcia na nieobciążonym wyjściu mikrofonu do wartości ciśnienia akustycznego przy określonej częstotliwości i kierunku padania dźwięku. Zaś charakterystyka kierunkowa to rozkład względnej skuteczności mikrofonu w funkcji kąta padania dźwięku na powierzchnię mikrofonu, odniesionej do maksymalnej skuteczności przy kierunku padania fali dźwiękowej 0°. Wśród podstawowych technologii wymienić należy mikrofony węglowe (działającej na zasadzie zmian rezystancji proszku węglowego ściskanego przez membranę, powodujących zmiany przepływającego prądu), dynamiczne (cewkowe, wstęgowe, wykorzystujące zjawisko indukcji elektromagnetycznej), pojemnościowe (zmiany pojemności), elektretowy (zmiany pola elektrycznego) czy optyczny.



Rysunek 1.26: Nagrania dookólne w instytucie NHK w Japonii (Fukada): po lewej - nagranie muzyki kameralnej, po prawej - nagranie muzyki jazzowej (źródło: zaczerpnięte z http://sound.eti.pg.gda.pl/student/tn/techniki_mikrofonowe2.pdf).

1.3 Prezentacja danych

Przekaz multimedialny jest skierowany do odbiorcy informacji, dlatego też skuteczna prezentacja dostarczonych danych jest bardzo ważnym czynnikiem, warunkującym powodzenie całego przedsięwzięcia.

Chodzi tutaj w pierwszej kolejności o wizualizację, czyli przedstawianie informacji w postaci graficznej lub obrazowej. Kluczowa rolę odgrywają tutaj systemy wizyjne, wysokiej klasy monitory LCD, plazmowe, czy najnowszej generacji supercienkie monitory w technologii OLED (*Organic Light Emitting Diode*), budowane na bazie diod organicznych. Zalety OLED w stosunku do najbardziej dziś rozpowszechnionych LCD (*Liquid Crystal Display*) to przede wszystkim większa skala barw i jasności, wysoki kontrast z prawdziwą czernią, bardzo krótki czas reakcji, wynoszący znacznie poniżej 1 milisekundy oraz bardzo duży kąt widzenia. Brak tylnego podświetlenia obniża pobór energii oraz koszty produkcji, która została uproszczona. Do produkcji wyświetlaczy OLED wykorzystuje się diody emitujące światło zielone, czerwone i niebieskie, przy czym istotnym parametrem jest ich żywotność oraz czystość barwy, szczególnie trudne do uzyskania dla diod niebieskich.

Przetworniki obrazów konwertujące elektryczny sygnał wizyjny na obraz świetlny (wyświetlany, prezentowany) to wyświetlacze obrazów, realizujące wyświetlanie sterowane sygnałem elektrycznym, adresowanie miejsca świecenia oraz podtrzymanie emisji do czasu ponownej adresacji. Wyróżniamy wyświetlacze aktywne, będące samoistnym źródłem wyjściowego strumienia świetlnego oraz bierne, jedynie modulujące natężenie promieniowania świetlnego o stałym natężeniu i barwie, wytwarzane przez zewnętrzne źródło światła. Trzy zjawiska fizyczne, wykorzystywane najczęściej we współczesnych wyświetlaczach obrazów to:

- elektroluminescencja, czyli emisja światła pod wpływem napromienienia substancji emitującej światło (luminoforu) szybką wiązką elektronową, stosowana w lampach obrazowych (kineskopowych monitorach CRT), a więc wyświetlaczach aktywnych;

- wyładowanie jarzeniowe w plazmie, czyli samoistne świecenie zjonizowanego gazu (plazmy), związane z przepływem prądu elektrycznego przez gaz, stosowane w monitorach plazmowych (także wyświetlaczach aktywnych);

- efekt Schadta-Helfricha, polegający na zmianie kąta skręcenia płaszczyzny polaryzacji transmitowanego światła pod wpływem zewnętrznego pola elektrycznego; efekt ten zachodzi w ciekłych kryształach i jest wykorzystywany w biernych wyświetlaczach ciekłokrystalicznych LCD.

Drugim ważnym multimedialnym sposobem prezentacji informacji są systemy odtwarzania dźwięku, odsłuchu, generacji przestrzeni dźwiękowej, wykorzy-

Rysunek 1.27: Nowe generacje wyświetlaczy obrazów (od lewej do prawej): monitor CCD podświetlany diodami LED (Samsung LED LCD 8000), monitor plazmowy Panasonic G10 TC-P54Z1 oraz monitor OLED XEL-1 firmy Sony (źródło: materiały reklamowe dostępne w Internecie).

stywane we współczesnym kinie cyfrowym, kinie domowym, studiach nagrań, na koncertach, w terapii dźwiękowej,[15] itp.

Rozwój systemów rejestracji i odtwarzania dźwięku, zaczynając od monofonii i stereofonii dwukanałowej, poprzez kwadrofonię aż po współczesne systemy wielokanałowe cechuje dążenie do uzyskania realnego efektu przestrzennego obrazu dźwiękowego. Służy temu zwiększanie liczby kanałów pełnopasmowych, choć subiektywnie oceniana jakość dźwięku nie poprawia się proporcjonalnie ze wzrostem liczby kanałów. Także rozmieszczenie głośników o odpowiedniej charakterystyce, uzupełnienia konfiguracji systemu dźwięku wielokanałowego o niskoczęstotliwościowe kanały efektowe oraz kanały dookólne przynosi poprawę doświadczenia obrazu dźwiękowego. Wśród najbardziej popularnych systemów wymienić należy:

- 5.1 (Dolby Digital z koderem Audio Coding AC-3 oraz Digital Theater System DTS z koderem Coherent Acoustic Coding CAC) – trzy niezależne przednie i dwa tylne kanały pełnopasmowe z głośnikami (zestawami głośnikowymi) rozmieszczonymi na okręgu, którego środek wyznacza referencyjną pozycję słuchacza, uzupełnione kanałem efektowym.

- 6.1 (Dolby EX oraz DTS-ES) – z dodatkowym, centralnym głośnikiem dookólnym, kodowanym matrycowo z kanałów dookólnych - lewego i prawego;

- 7.1 (Dolby Laboratories) – z czterema głośnikami dookólnymi: bocznym lewym, bocznym prawym, tylnym lewym i tylnym prawym;

[15] Terapia dźwiękowa jest wypróbowaną formą terapii, która działa poprzez słuch bezpośrednio na stan psychiczny, mózg i system nerwowy człowieka. Terapia ta jest instrumentem prowadzącym od stymulacji do harmonii, witalności, energii życia i zdolności koncentrowania się.

- inne, wprowadzające rozmieszczenie głośników na kilku poziomach (wysokościach), gdzie liczba kanałów sięga liczby 21.

1.3.1 Formy wizualizacji i odtwarzania

Najczęściej spotykane formy wizualizacji to przede wszystkim:

- wizualizacja statycznej, czyli ilustracje, wykresy, diagramy, mapy, rysunki oraz zdjęcia; aby pokazać zmiany w czasie stosowało się np. przedstawienie różnych faz na jednym rysunku lub serię rysunków jeden obok drugiego [217];.
- wizualizacja dynamiczna, spotykana przede wszystkim w klasycznej telewizji oraz prezentacjach zapisów wideo;
- wizualizacja komputerowa, czyli połączenie wizualizacji statycznych i dynamicznych z dodatkową możliwością interakcji, gdzie istotne jest dokładne kontrolowanie danych oraz zapewnienie informacji zwrotnej (typowe dla gier decyzyjnych i symulacyjnych); ciekawą, bezpośrednią formę interakcji zapewniają monitory dotykowe.

Z kolei do najbardziej popularnych form odtwarzania dźwięku (odsłuchu) należą:

- klasyczny, czyli monofoniczny lub stereofoniczny (dwukanałowy),
- przestrzenny, z rosnącym udziałem głośników dookólnych ora efektów specjalnych;
- słuchawkowy, bardzo popularny dziś sposób, który umożliwia indywidualny odbiór muzyki praktycznie w każdych warunkach.

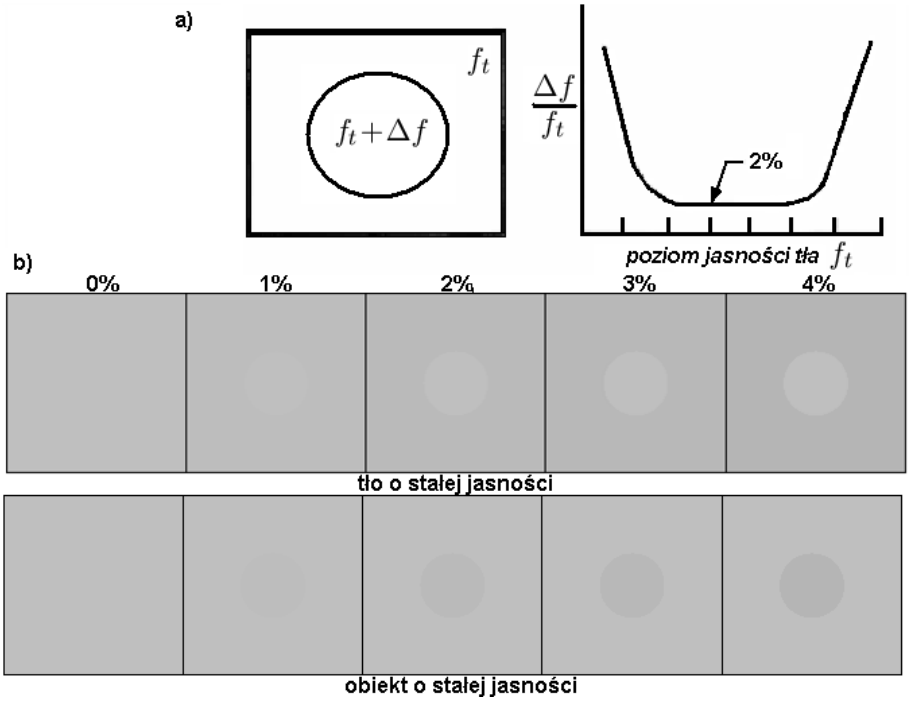
1.3.2 Percepcja i poznanie

Percepcja informacji obrazowej - model HVS

Znane właściwości ludzkiego systemu widzenia (*Human Visual System*), tj. niedoskonałość ludzkiego wzroku, określone zasady percepcji treści obrazowej, odczuwalne kryteria jakości, subiektywizm i zmienność ocen (po czasie i po różnych realizacjach), a także określony sposób pracy z obrazem wymagają dostosowania metod wizualizacji (inaczej prezentacji) zarejestrowanej informacji obrazowej do charakterystyki odbiorcy.

Czułość kontrastu decyduje o możliwości rozróżniania obiektów poprzez postrzeganie różnic w poziomie jasności obiektu f_o względem jednorodnego tła f_t. Próg widoczności $\Delta f = f_o - f_t$ nie jest stały w całym zakresie wartości f_t - czułość spada przy ciemnym i jasnym tle, tak jak to pokazano na rys. 1.28. Okazuje

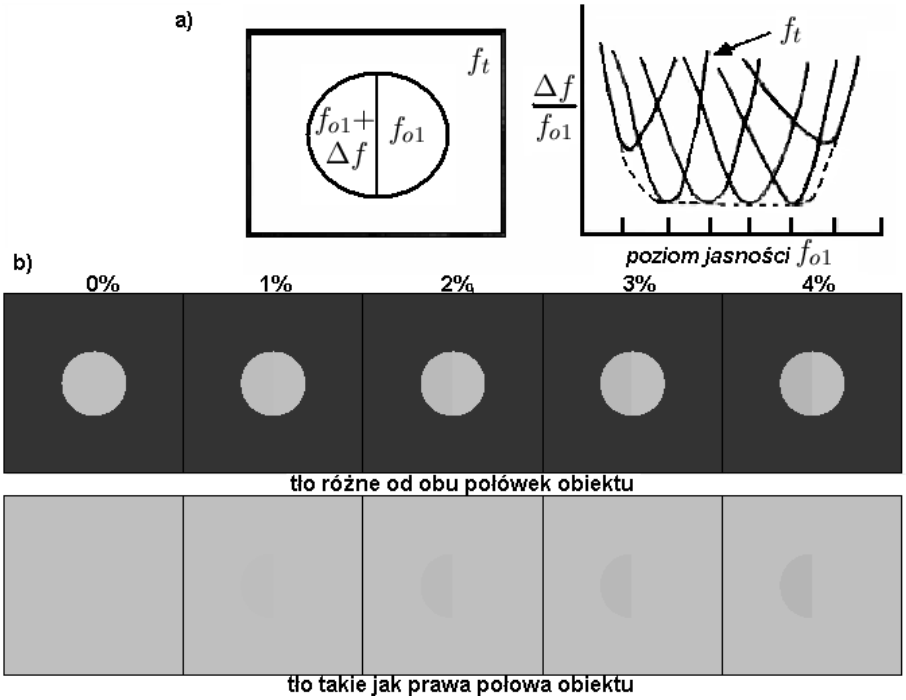
się, że jeszcze trudniej jest dostrzec zróżnicowanie poziomów jasności w obiekcie $\Delta f = f_{o2} - f_{o1}$ przy odmiennej jasności tła - zobacz rys. 1.29.



Rysunek 1.28: Określenie czułości kontrastu: a)krzywe Webera uzyskane przy jednorodnym obiekcie wyróżnianym z tła, b)przykładowe obrazki z eksperymentu. Typowa czułość kontrastu JND (*Just Noticeable Difference*) w środkowym zakresie wartości funkcji jasności wynosi 2% zarówno w teście ze stałą jasnością obiektu, jak też ze stałą jasnością tła.

Mózg ludzki podczas analizy obrazu uwzględnia otoczenie w jakim występuje dany obiekt. Zależnie od cech kontekstu może pojawić się wrażenie deformacji, zmiany koloru, czy wręcz pojawiania się nowych elementów w obrazie albo znikania drobnych obiektów. Przykładem mogą być dwa odcinki jednakowej długości przedstawione na rys. 1.30a), które są odbierane jako dłuższy i krótszy (efekt rozciągania i ściskania przez groty strzałek). Innym przykładem są dwie równoległe linie (rys. 1.30b)), których wzajemna relacja fałszowana jest poprzez wyraźną preferencję kierunku rozchodzenia się symulowanej fali.

Nieobiektywność ludzkiego wzroku potwierdzają więc różnego typu iluzje (rys. 1.31) będące przede wszystkim efektem adaptacji zdolności widzenia do zróżnicowanego kontrastu w otoczeniu obiektu czy też odmiennej jasności otaczającego tła. Dodatkowo wzmacniane są niewielkie różnice cech porównywanych obiektów (kolor, kontrast, rozmiar). Symetria widzenia obszarów jasnych i ciemnych powoduje wrażenie występowania obiektów pozornych, które niekiedy w określonych warunkach mogą zakłócić proces poprawnej, tj, odpowiadającej realnym cechom



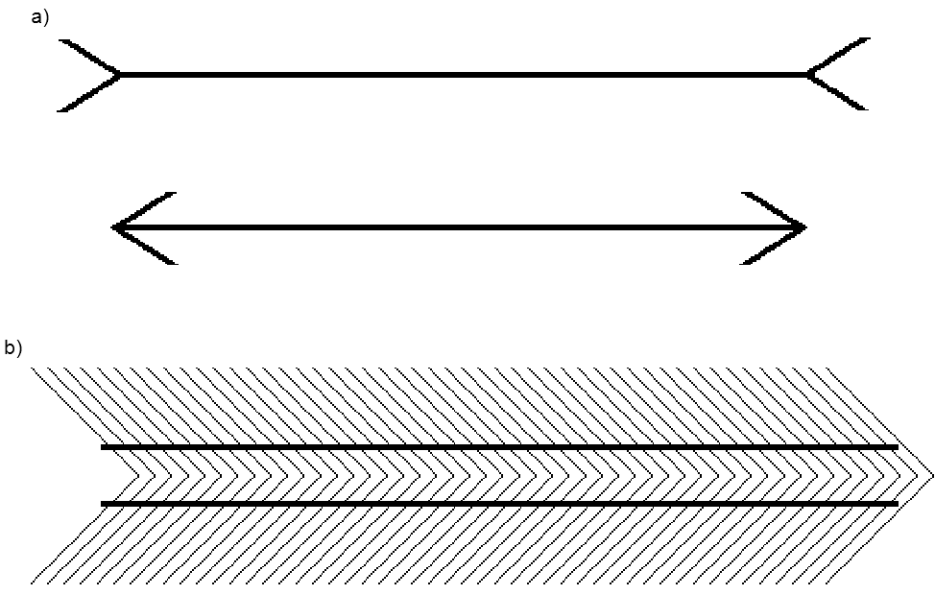
Rysunek 1.29: Określenie czułości kontrastu: a)krzywe Webera uzyskane przy niejednorodnym obiekcie wyróżnianym z tła, b)przykładowe obrazki z eksperymentu. Typowa wartość JND w środkowym zakresie wartości funkcji jasności wynosi 4% przy jasności tła odmiennej od obiektu oraz 2% przy równej jasności tła i części obiektu.

obiektów, interpretacji obrazów.

Kolejny przykład przedstawia okręgi w centralnym punkcie obrazu będące podmiotem obserwacji, otoczone okręgami wyraźnie mniejszymi lub większymi. Przy odbiorze informacji obrazowej trudno jest określić czy interesujące okręgi mają jednakowe rozmiary - zobacz rys. 1.32.

Percepcja dźwięku

W odbiorze dźwięku podstawowe są zróżnicowane wrażenia słuchowe, co do wysokości i głośności. Percepcja dźwięku związana jest z miejscem pobudzenia błony podstawnej, amplitudą odkształcenia i jej powierzchni. Zależy także od liczby impulsów przechodzących przez włókna nerwowe – głośność dźwięku jest proporcjonalna do liczby impulsów nerwowych powstałych w danej jednostce czasu. Gdy rośnie poziom ciśnienia (do ok. 60 dB) - rośnie liczba impulsów w jednym neuronie, zaś przy bardzo dużym poziomie pobudzanych jest więcej neuronów. Najmniejsza zmiana poziomu ciśnienia powodująca zmianę wrażenia głośności, tzw. próg różnicy głośności, wynosi około 0,2 dB dla tonu 1000 Hz.

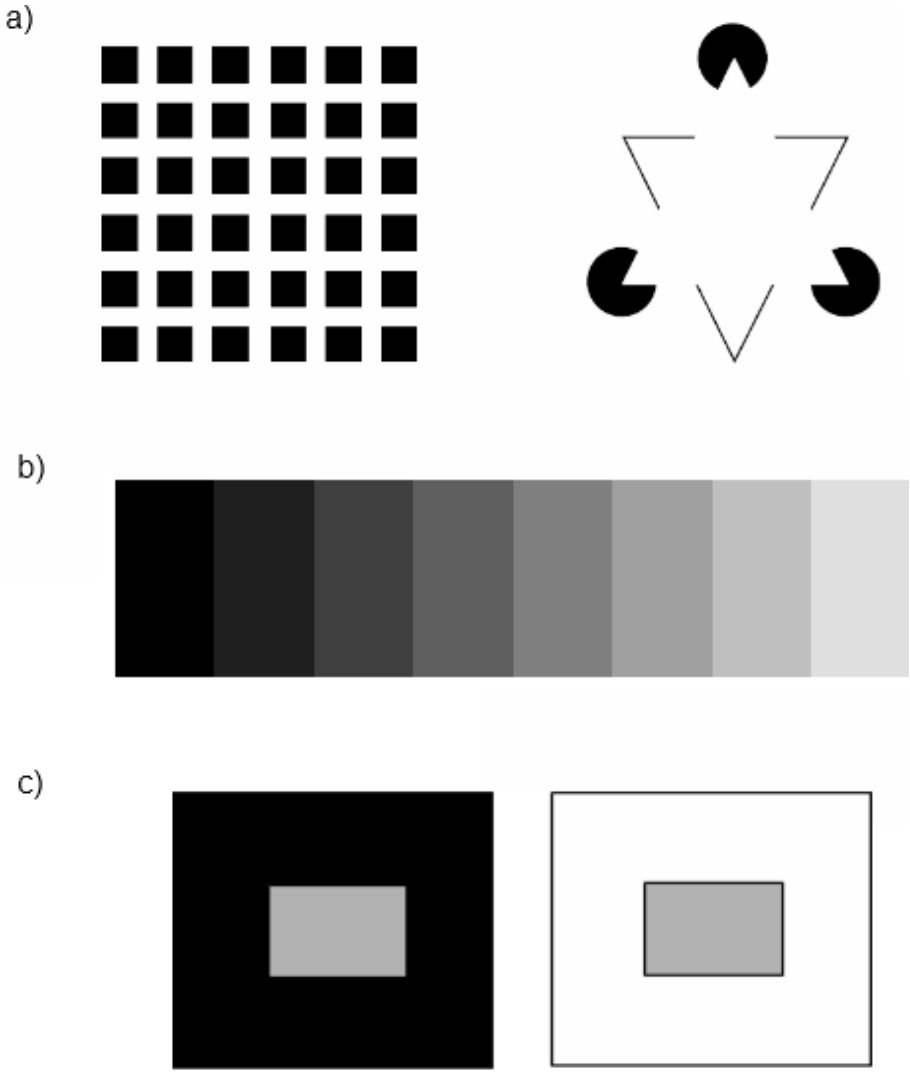


Rysunek 1.30: Iluzje wynikające z określonych cech ludzkiego wzroku: a) odcinki o jednakowej długości wydają się różne, b) linie odbierane są jako nierównoległe.

Odczuwanie wysokości przejawia się przy słuchaniu jednoczesnym tonów złożonych, jako zdolność do wyodrębnienia składowych dźwięku (tonów harmonicznych). Ponadto, odczuwana jest także przy słuchaniu kolejnych dźwięków, jako zdolność do rozróżniania zmian częstotliwości. Percepcja wysokości zależy od zarówno od miejsca odkształcenia błony podstawnej, jak i od czasowego rozkładu impulsów w neuronach. Wrażenie wysokości zależy przede wszystkim od częstotliwościowego widma dźwięków, czasu trwania i poziomu ciśnienia.

Głośność i wysokość są cechami wrażeniowymi jednowymiarowymi, natomiast barwa ma charakter wielowymiarowy. Barwa dźwięku zależy głównie od względnych amplitud składowych widmowych i ich zmienności w czasie. Barwa to wrażenie, które pozwala słuchaczowi rozróżnić dwa dźwięki złożone o tej samej głośności i wysokości.

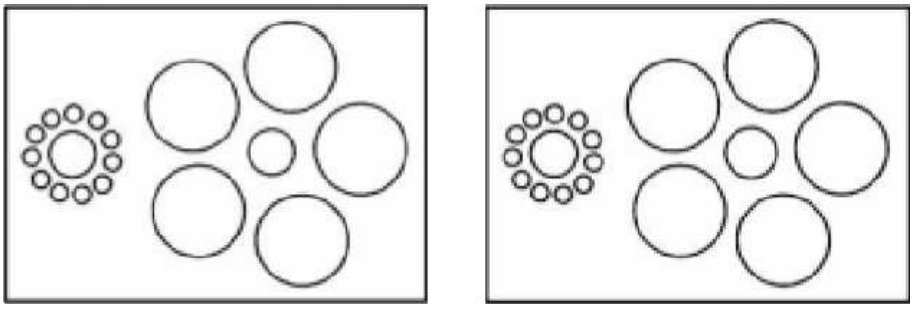
Warto wspomnieć o kilku efektach, które charakteryzują ludzkie zdolności percepcji dźwięku. Efekt precedensu polega na tym, że jeśli krótkie dźwięki (impulsy, transjenty) docierają w małym odstępie czasu (1 - 5 ms dla impulsów tonu, do 40 ms dla impulsów mowy lub muzyki) słyszane są jako jeden dźwięk, którego lokalizacja jest zdeterminowana przez kierunek pierwszego. Dźwięk wtórny ma niewielki wpływ na lokalizacje, jeśli dociera z kierunku bardzo odległego od kierunku dźwięku pierwotnego. Jeśli odstęp czasu miedzy impulsami jest mniejszy



Rysunek 1.31: Iluzje wynikające z określonych cech ludzkiego wzroku: a)pojawiają się pozorne obiekty, b) lewa część każdego z pasków wydaje się jaśniejsza, c) prawy prostokąt (na jaśniejszym tle) jest pozornie bardziej szary.

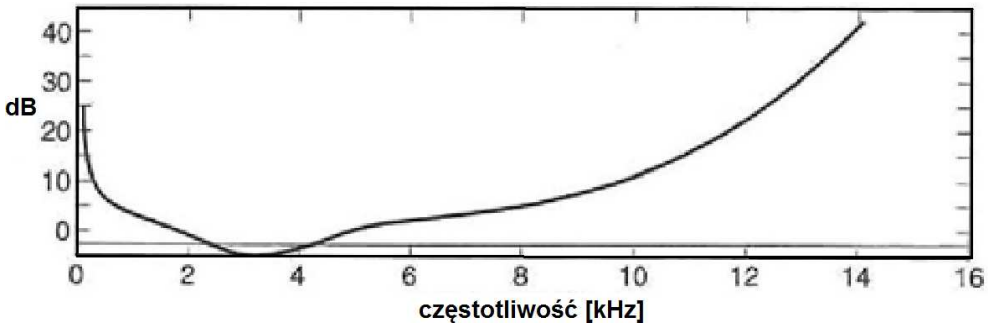
od 1 ms, to efekt precedensu nie występuje, a źródło lokalizowane jest pomiędzy kierunkiem dźwięku pierwotnego i wtórnego. Efekt precedensu zanika, gdy poziom dźwięku wtórnego przewyższa poziom pierwotnego o 10-15 dB.

Ważne są też zjawiska maskowania w pasmach krytycznych słuchu i krzywej progowej słyszenia dwuusznego. Ucho ludzkie nie reaguje na dźwięki o natężeniu mniejszym od pewnej wartości zwanej progiem słyszenia dwuusznego. Wartość progowa zależy od częstotliwości dźwięku – największa czułość słyszenia występuje w zakresie od 2 kHz do 4 kHz (1.33). Dźwięki usytuowane na krzywej progowej są ledwie słyszalne. Słuchacz nie odczuwa pogorszenia jakości percepcji sygna-



Rysunek 1.32: Iluzja Ebbinghausa. Obserwator odnosi wrażenie, że na rys. po lewej stronie środkowy okrąg otoczony mniejszymi okręgami jest większy od okręgu otoczonego okręgami dużymi, natomiast na rys. po prawej oba centralne okręgi wydają się jednakowe. W rzeczywistości na lewym rysunku interesujące obiekty są jednakowych rozmiarów, a na prawym większy jest okrąg otoczony dużymi okręgami.

łu dźwiękowego, jeśli z widma tego sygnału zostaną usunięte składowe, których poziom jest niższy od progu słyszenia.

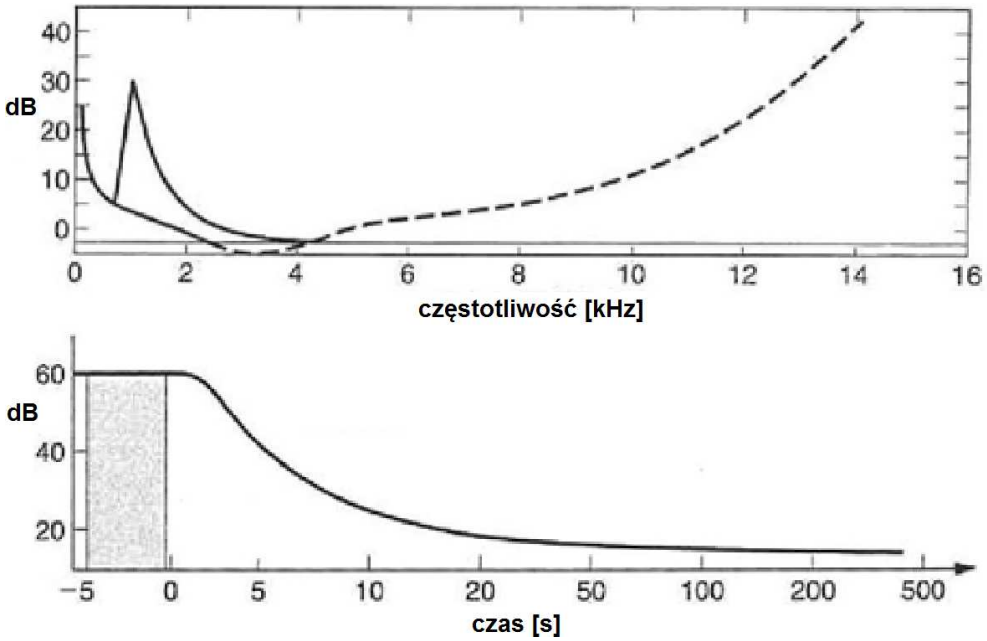


Rysunek 1.33: Przebieg krzywej progowej słyszenia dwuusznego.

Efekt maskowania jest to zjawisko polegające na podniesieniu progu słyszalności dźwięku maskowanego wskutek obecności dźwięku maskującego (maskera), tzn. dźwięk maskowany może być wtedy słyszany słabo albo wcale. Zjawisko to jest ściśle związane z adaptacją układu słuchowego. Rozróżnia się maskowanie równoczesne i nierównoczesne, związane odpowiednio z właściwościami częstotliwościowymi i czasowymi dźwięków.

Maskowanie równoczesne polega na tym, że dźwięk maskowany znajduje się w najbliższym czasowo sąsiedztwie (razem – przy jednoczesnej percepcji) tonu maskującego. Skuteczność maskowania zależy od natężenia dźwięków: maskującego i maskowanego oraz wzajemnej relacji ich widm częstotliwościowych. Zasadniczo najsilniej maskowane są dźwięki o zbliżonych częstotliwościach, a tony o mniejszych częstotliwościach maskują dźwięki o częstotliwościach wyższych – na rys. 1.34 widoczny jest efekt maskowania pojedynczym tonem 1 kHz.

W przypadku maskowania niejednoczesnego (rys. 1.34) spada słyszalność dźwięku maskowanego, gdy masker występuje bezpośrednio przed lub bezpośrednio po sygnale maskowanym. Czas maskowania po zaniku tonu maskującego może sięgać do około 200 ms i zależy od natężenia tonu maskującego oraz czasu jego trwania.



Rysunek 1.34: Efekty maskowania: zmiana krzywej progowej wskutek maskowania równoczesnego tonem 1 kHz (góra) oraz efekt maskowania nierównoczesnego (dół).

1.3.3 Ocena jakości

Kryteria oceny jakości obrazów nie są jednoznaczne. Zwykle obraz jest dobrej jakości, gdy według percepcji wzrokowej wygląda "przyjemnie", czyli jest odpowiednio skontrastowany, z dobrze widocznymi szczegółami, z naturalną paletą barw, nie ma rzucających się w oczy zniekształceń, artefaktów, a treść jest rozpoznawalna, czytelna, zrozumiała. Niekiedy o wysokiej jakości obrazu świadczy jego specyficzna użyteczność w określonym zastosowaniu, niekoniecznie związana z wysokiej jakości wrażeniem ogólnym, np. dobrze są rozpoznawalne istotne szczegóły obrazowanych struktur, których detekcja jest zasadniczym celem stosowania metod obrazowych (tak jest w radiologii, przy rozpoznawanie zmian patologicznych w obrazach medycznych).

Nie istnieje niestety jedna skuteczna miara pozwalająca określić jakość obrazu. Można generalnie wyróżnić następujące metody oceny jakości obrazów:

- obliczeniowe miary zniekształceń – wielkości liczbowe skalarne bądź wekto-

rowe, formy graficzne, wyznaczane automatycznie według ustalonych reguł matematycznych, a więc powtarzalne, porównywalne, obiektywne, niekiedy uwzględniające modele ludzkiego widzenia i postrzegania treści obrazowej;

- miary obserwacyjne, będące efektem testów subiektywnej oceny jakości – psychowizualne testy oceny jakości przeprowadzane przy pomocy typowych użytkowników, bądź też grona specjalistów danej dziedziny według ustalonych procedur, z wykorzystaniem liczbowej skali ocen, z przypisanym znaczeniem poszczególnych poziomów ocen, lub też mechanizmu porządkowania obrazów w kolejności zgodnej z postrzeganą jakością obrazów;
- obliczeniowo-obserwacyjne miary jakości – wektorowe, złożone miary obliczeniowe optymalizowane wzorcem z testów subiektywnej oceny jakości obrazów.

Ze względu na sposób oceny, metody te możemy podzielić na:

- miary absolutne, bezwzględne (*univariate*),
  - liczbowe, takie jak miary kontrastu, parametry histogramów, przekrojów rozkładu funkcji jasności, estymatory poziomu szumów z widma spektralnego, kierunkowe współczynniki korelacji, parametry modeli regresji itp.,
  - graficzne, przykładowo miara Eskicioglu [5], będąca słupkowym rozkładem dynamiki, odchyleń standardowych oraz liczebność klas bloków diadycznych o rozmiarze od 2×2 do 16×16, na jakie dzielony jest obraz.
- miary porównawcze, względne (*bivariate*)
  - liczbowe, takie jak błąd średniokwadratowy, stosunek sygnału do szumów itp.,
  - graficzne, przykładowo wykresy Hosaki [4], będące wykreślonym na płaszczyźnie wielokątem ukazującym różnice wartości średnich oraz odchyleń standardowych w kilku klasach bloków o różnym rozmiarze, na jakie dzielony jest obraz.

Wśród metod oceny jakości obrazów przede wszystkim ze względu na ich użyteczność, wymienić należy przede wszystkim specjalistyczne testy użyteczności obrazów – złożone, dotyczące konkretnej aplikacji testy obserwacyjne bazujące na możliwie wiernej symulacji rzeczywistych warunków pracy z obrazami (detekcji określonych elementów, interpretacji treści), opiniach obserwatorów-specjalistów wyrażanych w formach liczbowych, możliwie wiernie symulujących realia oceny treści obrazów oraz wnikliwej analizie statystycznej odpowiednio opracowanych wyników testów klasyfikacyjnych.

Przykładem takiej oceny mogą być testy detekcji patologii w medycznych badaniach diagnostycznych z udziałem specjalistów-radiologów. Obrazy medyczne określonej modalności (ultrasonografii, tomografii komputerowej, rentgenowskie itp.) interpretowane są ze względu na obecność określonego rodzaju podejrzanych zmian patologicznych, a efekty podjętych decyzji diagnostycznych poddawane są analizie statystycznej z wykorzystaniem krzywej ROC (*Receiver Operating Characteristics*) [6].

Obliczeniowe miary jakości

Do najbardziej pożądanych cech miary obliczeniowej należy zaliczyć w pierwszej kolejności: duży poziom korelacji z subiektywną oceną obserwatorów – najczęściej ostateczną weryfikacją przydatności miary liczbowej jest jej zgodność, a przynajmniej niesprzeczność z oceną psychowizualną – oraz wysoką podatność w analizie obliczeniowej, tj. łatwość obliczeniową, prostotę aplikacji, bogactwo narzędzi do analizy i optymalizacji oraz łatwość interpretacji. Połączenie tych dwóch oczekiwań okazuje się w praktyce bardzo trudne.

Szczególnie w przypadku **miar skalarnych** uzyskanie dobrej korelacji z oceną subiektywną jest niełatwe. Miary te dają jednak łatwość interpretacji i analiz porównawczych. Niech oryginalny obraz cyfrowy, wielopoziomowy ze skalą szarości, o szerokości M i wysokości N będzie opisany funkcją jasności $f(k,l)$, $0 \leqslant m < M$, $0 \leqslant n < N$. Wartości pikseli obrazu przetworzonego w tej samej dziedzinie oznaczono przez $\tilde{f}(k,l)$. Do najbardziej użytecznych skalarnych miar jakości obrazów, należących do kategorii metod porównawczych liczbowych, zaliczyć należy przede wszystkim takie miary jak:

- maksymalna różnica (*maximal difference*):

$$MD = max_{k,l}\{|f(k,l) - \tilde{f}(k,l)|\} \tag{1.6}$$

- błąd średniokwadratowy (*mean square error*):

$$MSE = \frac{1}{KL}\sum_{k,l}[f(k,l) - \tilde{f}(k,l)]^2 \tag{1.7}$$

- szczytowy stosunek sygnału do szumu (*peak signal to noise ratio*):

$$PSNR = 10 \cdot log\frac{KL \cdot [max_{k,l}f(k,l)]^2}{\sum_{k,l}[f(k,l) - \tilde{f}(k,l)]^2} \tag{1.8}$$

- średnia różnica (*average difference*):

$$AD = \frac{1}{KL}\sum_{k,l}|f(k,l) - \tilde{f}(k,l)| \tag{1.9}$$

- znormalizowany błąd średniokwadratowy (*correlation quality*):

$$CQ = \frac{\sum_{k,l} f(k,l)\tilde{f}(k,l)}{\sum_{k,l} f(k,l)} \tag{1.10}$$

- dokładność rekonstrukcji obrazu (*image fidelity*):

$$IF = 1 - \frac{\sum_{k,l}[f(k,l) - \tilde{f}(k,l)]^2}{\sum_{k,l}[f(k,l)]^2} \tag{1.11}$$

- miara chi-kwadrat (*chi-square measure*):

$$\chi^2 = \frac{1}{KL} \sum_{k,l} \frac{[f(k,l) - \tilde{f}(k,l)]^2}{f(k,l)} \tag{1.12}$$

Do obliczeniowych miar poprawy kontrastu można zaliczyć:

- indeks poprawy kontrastu CII [7, 8], obliczany na podstawie skontrastowania obiektu i tła (DR):

$$DR = \frac{\overline{f}_O - \overline{f}_B}{\overline{f}_O + \overline{f}_B} \tag{1.13}$$

  gdzie $\overline{f}_O$ – średni poziom szarości obiektu, $\overline{f}_B$ średni poziom szarości tła obiektu;

$$CII = \frac{DR_p}{DR_o} \tag{1.14}$$

  gdzie DR_p i DR_o to kontrasty obiektu, liczone odpowiednio na obrazie przetworzonym i oryginalnym.

- miara separacji rozkładów DSM *'distribution separation measure'*):

$$DSM = (|\overline{f}_O^p - \overline{f}_B^p|) - (|\overline{f}_O^o - \overline{f}_B^o|) \tag{1.15}$$

  gdzie $\overline{f}_O^p$ i $\overline{f}_B^p$ odpowiada średniej intensywności obiektu i tła po przetworzeniu, a $\overline{f}_O^o$ i $\overline{f}_B^o$ to średni poziom intensywności obiektu i tła na oryginałach (przed przetwarzaniem).

- inne miary poprawy kontrastu to stosunki średnich intensywności obiektu i tła oraz ich odchylenia standardowe i entropie.

Porównawcze miary skalarne mogą być stosowane jako dodatkowa informacja, opisująca stopień wprowadzanych w obrazie zmian, jednak nie dają informacji o kierunku tych zmian – poprawie percepcji zmian czy ich zniekształceniu.

Miary obserwacyjne

Miary obserwacyjne są naturalnym sposobem oceny jakości obrazu. Opierają się na opiniach odbiorców informacji obrazowej, którzy określają ich jakość według własnych kryteriów, na podstawie doświadczenia, osobistych preferencji, czy też – w przypadku specjalistów – według obowiązującej wykładni danego zastosowania.

Podstawowymi problemami, związanymi ze stosowaniem miar obserwacyjnych, są: czasochłonność testów, subiektywizm (każdy ze specjalistów ocenia dany obraz w nieco inny sposób, brakuje obiektywnych, jednoznacznych kryteriów oceny), konieczność angażowania kilku niezależnych ekspertów, czynniki ludzkie, takie jak zmęczenie, możliwości pomyłek.

Skala ocen dla miar subiektywnych powinna mieć zdefiniowany zakres liczbowy oraz skojarzony z nim opis słowny. Opis ten ma charakter bezwzględny (miary absolutne) i względny (miary porównawcze). Przykłady skal z opisem podano w tabelach 1.2 i 1.3. Na podstawie ocen cząstkowych poszczególnych

Kategoria k	Wartość skali ocen s_k	Opis słowny, charakteryzujący względną jakość obrazu
1	3	Zdecydowanie (bezwzględnie) lepsza
2	2	Lepsza
3	1	Nieznacznie lepsza
4	0	Porównywalna z oryginałem
5	-1	Nieznacznie gorsza
6	-2	Gorsza
7	-3	Zdecydowanie (bezwzględnie) gorsza

Tabela 1.2: Przykładowa skala ocen jakości obrazów, stosowana w psychowizualnych testach porównawczych do oceny metod poprawy percepcji. Służy do opisu ogólnego, subiektywnego wrażenia obserwatorów, porównujących obraz przetworzony z oryginalnym.

Kategoria k	Wartość skali ocen s_k	Opis słowny, charakteryzujący bezwzględną jakość obrazów
1	5	Wyśmienita
2	4	Dobra
3	3	Akceptowalna
4	2	Słaba
5	1	Nie do przyjęcia

Tabela 1.3: Przykładowa skala ocen jakości obrazów, stosowana w psychowizualnych testach miar subiektywnych. Zawiera opis słowny w kategoriach bezwzględnych (jedynie na podstawie obserwowanego obrazu).

osób, biorących udział w teście, jest obliczana średnia ocena grup obrazów przez obserwatorów według zależności:

$$S = \frac{\sum_{k=1}^{K} s_k n_k}{\sum_{k=1}^{K} n_k} \tag{1.16}$$

gdzie K– liczba kategorii w przyjętej skali, s_k – wartość oceny, związanej z kategorią k, n_k – liczba ocen z danej kategorii.

Analogicznie do obserwacyjnych metod subiektywnej oceny jakości obrazów przez odbiorców informacji obrazowej (obserwatorów), można mówić o subiektywnej ocenie jakości dźwięku przez odbiorców (słuchaczy) – odsłuchowe testy subiektywne. Ogólniej, chodzi o subiektywne testy odbioru informacji multimedialnej przez osoby weryfikujące jakość przekazu multimedialnego. Ogólne zasady tych testów są takie same jak w przypadku obrazów, natomiast dobierana skala ocen powinna oddawać zasadniczy cel oceny jakości przekazu (dominującą rolę któregoś ze strumieni danych, synchronizację treści przekazu, wskazanie na strumień o najwyższej jakości, ogólne wrażenie percepcji całości informacji, itp.).

1.4 Podsumowanie

Do najistotniejszych zagadnień poruszonych w tym rozdziale należy charakterystyka specyfiki przekazu multimedialnego – zintegrowanej wymiany informacji z podmiotową rolą odbiorcy oraz innowacyjnymi intencjami nadawcy, uwzględniającą szczególną wagę reprezentowania informacji przy różnorodnej treści i formie przekazywanych danych. Przytoczono szereg aplikacji, zwrócono uwagę na rosnącą skalę i różnorodność zastosowań – od tych typowych, powszechnie znanych, po wysoce specjalistyczne, nawiązujące do doświadczeń komputerowej inteligencji.

Zarys metod rejestracji i prezentacji danych multimedialnych, krótki opis stosowanych urządzeń sygnalizuje olbrzymią rolę systemów wejściowych, dostarczających danych źródłowych oraz wyjściowych, efektowych, służących postrzeganiu i rozumieniu treści przekazu przez użytkownika. Zresztą trudno przecenić coraz wnikliwszą i zindywidualizowaną rolę użytkownika, która staje się normą semantycznych technologii współczesnych multimediów.

Warto zwrócić więc uwagę na modele percepcji, oceny i rozumienia informacji przez człowieka. Nawiązują do nich zasygnalizowane metody oceny jakości i użyteczności obrazów. Istotne jest, jakie czynniki wpływają na przyjazny odbiór informacji obrazowej, jakie są ludzkie ograniczenia, a jakie możliwości postrzegania określonych obiektów, na czym polega subiektywizm ocen oraz formy ich obiektywizacji. Zrozumienie mechanizmów opisu, przetwarzania, analizy i syntezy treści multimedialnej staje się coraz bardziej niezbędne w twórczym wykorzystywaniu potencjału współczesnych multimediów.

Zadania do tego rozdziału podano na stronie 359.

Ćwiczenie pozwalające na eksperymentalną weryfikację zasad rejestracji oraz postrzegania multimediów zamieszczono odpowiednio na stronach: rejestracja – str. 373, postrzeganie – str. 377.

Rozdział 2

Reprezentowanie informacji

Reprezentowanie informacji jest zagadnieniem istotnym w każdym niemal zastosowaniu, a w różnych formach i postaciach jest także obecne w rozważaniach wielu teorii abstrakcyjnych i stosowanych. W uproszczeniu, można posłużyć się następującym schematem. By przekazać informację, istotny jest przede wszystkim sposób jej wyrażenia – zrozumiały dla odbiorcy, ale niezbędny jest także fizyczny nośnik przekazu oraz zorganizowana forma technicznego czy technologicznego zapisu danych, czyli jej reprezentacja. Wymagana jest określona reguła tworzenia reprezentacji danych, czyli kod ustalony na etapie akwizycji (dający źródłową reprezentację danych), bądź też w dalszym procesie przekazywania i przetwarzania danych przenoszących informację (np. reprezentacja kodowa uzyskiwana wskutek kompresji danych). W sposób jawny lub niejawny, wprost lub pośrednio reprezentacja ta zawiera elementy opisu danych, w tym pewne odnośniki do ich semantyki. Zależnie od zastosowań, celów wykorzystania danych czy charakteru zawartej treści poszukiwane są skuteczne formy reprezentacji informacji.

Zarys teorii użytecznych przy wyznaczaniu efektywnej reprezentacji informacji to przede wszystkim:

- podstawy teorii sygnałów, przy czym sygnały rozumiane są przede wszystkim jako nośnik informacji czy też urealnienie przekazu informacji w naturalnych warunkach akwizycji, ucyfrowienia i kodowania sygnałów;

- zarys teorii informacji, zarówno w jej probabilistycznej koncepcji składniowej (prace Shannona), jak też w rozszerzeniu do semantycznej teorii informacji;

- podstawy teorii aproksymacji, jako poszukiwanie przybliżeń treści istotnych przekazu informacji, nawiązujących zarówno do liczbowej charakterystyki

sygnału, jak też jego warstwy znaczeniowej (semantyki), odwołującej do określonego modelu odbiorcy;

- charakterystyki odbiorcy informacji, zdolności percepcji treści, np. ludzkiego systemu widzenia – ang. *human visual system* czy też pracy odbiorcy ze źródłami informacji, rozpoznawanie treści, jej interpretacja – metoda ROC (ang. *Receiver Operating Characteristic*);
- metody inteligencji obliczeniowej, przede wszystkim w zakresie wyszukiwania, ekstrakcji czy rozpoznawania informacji, a przede wszystkim jej interpretacji.

Rozdział ten obejmuje podstawy teorii informacji, zarówno w sensie matematycznej koncepcji informacji rozumianej jako poziom niepewności odbiorcy, jak też jej semantycznych rozszerzeń, tak istotnych w aplikacjach multimedialnych. Tzw. matematyczna teoria informacji abstrahuje od specyfiki zastosowań nadawanej semantycznymi modelami informacji, wymaganiami dotyczącymi charakterystyki użytkownika (odbiorcy informacji), określonej specjalistyczną wiedzą dziedzinową. Dzięki temu dostarcza modele źródeł informacji, które są przejrzyście sformalizowane, obiektywne, bardziej uniwersalne i podatne na zaawansowane rozwiązania numeryczne, optymalizacyjne, porównawcze. Poprzez semantyczne rozszerzenia tych modeli możliwa jest selekcja informacji w metodach kompresji, ekstrakcja treści istotnej z zaszumionego sygnału źródłowego, a za pomocą deskryptorów treści tworzone są mechanizmy przeszukiwania zawartości rozległych zasobów danych multimedialnych zgodnie z oczekiwaniami użytkownika.

2.1 Wprowadzenie

Informacja służy odbiorcy w realizacji określonego celu. Przekaz danych dokonuje się zawsze w kontekście określonej treści, tj. funkcji semantycznej oraz jej wartości dla odbiorcy, czyli użyteczności. Precyzyjnie określając semantykę, śledząc jej zmiany przy selektywnej lub zakłóconej komunikacji danych pośrednio definiujemy również użyteczność tych treści. Pozwala to na automatyczne wyznaczenie ilości informacji z uwzględnieniem jej semantycznych właściwości.

Podstawowy schemat przekazu informacji jest następujący:

a) dane (w reprezentacji źródłowej, nad określonym alfabetem),

b) znaczenie danych (przypisane pojedynczym symbolom, grupie symboli),

c) treść (rozpoznanie obiektów, integracja znaczenia obiektów, znaczenie relacji pomiędzy obiektami, efekt synergii),

d) informacja (relacja rozpoznana treść (gdzie pewną rolę odgrywają także źródło informacji, zdolność percepcyjna oraz interpretacyjna odbiorcy) – odbiorca źródło, wiedza i doświadczenie)

2.1.1 Informacja

Definicja 2.1 *Informacja*
Informacją nazywamy to wszystko, co przekazane – okazuje się użyteczne dla odbiorcy. Informacja służy realizacji zamierzonego celu, zaspokaja określone potrzeby, ale też uświadamia, buduje wiedzę, ukazuje nowe możliwości, weryfikuje domniemania. □

Punktem odniesienia przekazu informacji jest odbiorca, jego cele i poczucie użyteczności. Przesyłane dane mają określone znaczenie, opisane funkcją semantyczną, które kształtuje treść przekazu. Odbiorca rozumiejąc treść danych, weryfikuje ich użyteczność. Odbiera informację lub uznaje przesłane dane za bezużyteczne. Nadawca formując przekaz stara się zaspokoić domniemane potrzeby odbiorcy.

Wymiana informacji, o możliwie atrakcyjnej treści oraz stosowanej formie (reprezentacji), jest podstawową funkcją szeroko rozumianych multimediów. Zarówno sposób – bezpośredni przekaz (komunikacja) lub pośrednicząca archiwizacja, jak i forma – uzupełniających się strumieni danych o charakterze zróżnicowanym w sensie sposobu percepcji przekazywanej treści.

Zakładając sensowność procesu wymiany danych, należy doszukiwać się występującej tam informacji, przyjmując ogólny schemat nadawcy i odbiorcy spiętych ustaloną formą kanału transmisyjnego o charakterze pozytywnym. Pozytywny znaczy choćby w minimalnym stopniu użyteczny, gdzie obok danych i treści

nadmiarowych pojawia się choćby ślad informacji nadającej sens całemu przedsięwzięciu. Informacja jest wtedy sensem i istotą przekazu bez względu na jego charakter. Dlatego efektywne reprezentowanie informacji stanowi podstawowe zagadnienie wszystkich aplikacji multimedialnych.

Przekaz informacji poprzedzony jest procesem pozyskiwania informacji – niekiedy kosztownym, innym razem dość przypadkowym, – bazującym na złożonych, kosztownych technologiach lub przede wszystkim na ludzkiej spostrzegawczości. Pozyskanie treści jest niekiedy bardzo trudne i musi być uzupełnione złożonym procesem wydobywania treści z nadmiaru rejestrowanych danych. Istotnym okazuje się wtedy problem ekstrakcji czytelnej postaci informacji z jej formy niejawnej, subtelnej, zniekształconej, itp. Rejestrowany sygnał – ciąg danych staje się nośnikiem określonej treści, która rozpoznawana jest jako informacja w kontekście jej użyteczności. Przekaz treści stanowiącej informację w szerokiej skali społecznej koncentruje dziś uwagę twórców najbardziej ambitnych rozwiązań w obszarze mediów cyfrowych, multimediów, telewizji, internetu, technik komputerowych i wielu innych.

Doskonalenie form przekazu oraz rosnąca cena wartościowej informacji charakteryzują współczesny rozwój sieciowego społeczeństwa informacyjnego, niezbyt szybko (a może wcale?) zmierzający w kierunku wizji nowoczesnego społeczeństwa wiedzy. Wyraźny nadmiar powielanych, sztampowych treści i pseudotreści, które, rozsyłane, pretendując do miana informacji "szukają łatwego zysku", powoduje stały wzrost znaczenia wolności wyboru w wymiarze osobistym i społecznym. Intencje nadawcy nie są zwykle jednoznaczne, a korzyści odbiorcy są często rozumiane "interesownie". Nadawca formując przekaz stara się spełnić domniemane oczekiwania odbiorcy, albo je biznesowo kreować. Rozpoznanie informacji bazującej na przekazie prawdziwych, otwartych treści staje się sztuką, ale i koniecznością. Czas odbiorcy staje się cenny dla nadawcy, ale przede wszystkim dla samego odbiorcy. Rośnie znaczenie filtrów, automatycznego rozpoznawania treści użytecznych, innych form preselekcji przekazu.

Wymagający odbiorca korzysta z wolności wyboru źródeł przekazu, interesuje się wiarygodnością otrzymywanych danych, oddziela ewentualny komentarz czy narzuconą interpretację. Wybiera przyjazne, sprawdzone formy, by dotrzeć do istoty przekazu, rdzenia odczytywanych treści, weryfikuje ich prawdziwość. Skuteczna weryfikacja warunkowana jest dostateczną jakością danych, czytelnością treści, jej uporządkowaniem, klarownością.

Kluczowym zagadnieniem, które służy odbiorcy jest efektywne reprezentowanie informacji, czyli konwersja przekazu danych w prezentację informacji z kryterium maksymalnej użyteczności odbiorcy. Przedstawione w tej pracy zagadnienia służą przede wszystkim zrozumieniu teoretycznych i praktycznych podstaw pojęcia reprezentowanej informacji. Według założonej koncepcji wspomagania procesu przekazu, wykorzystanie "podanej" informacji leży w gestii odbiorcy.

2.1.2 Reprezentacja danych

Reprezentacja danych to sposób przedstawienia lub inaczej organizacji danych. Dane w maszynach cyfrowych mają swoją reprezentację w postaci sekwencji bitów kodu dwójkowego, kodów bardziej złożonych, łączonych w bajty, wielobajtowe słowa, bloki. Są one interpretowane w terminach wewnętrznych typów danych określonej dziedziny, struktury, za pomocą operacji na liczbach lub znakach tekstu, jako liczby całkowite i ułamki, itp.

Reprezentacja danych może być rozumiana na różnych poziomach abstrakcji. Uwzględniając znaczenie i charakter danych może być orientowana na określoną treść, hierarchię istotności, wydzielenie sygnału i redukcję szumu, uporządkowanie według przyjętych kryteriów. Sposób kształtowania reprezentacji danych może być różnorodny, zwykle jednak przebiega według typowego, nie zawsze pełnego schematu:

a) reprezentacja źródłowa, opisana najprostszym kodem, np. dwójkowym;

b) reprezentacja wstępnie przetworzona, z redukcją szumu i poprawionym kontrastem;

c) reprezentacja estymowanego sygnału, z wydzieloną treścią użyteczną;

d) reprezentacja upakowana, rzadka, po usunięciu nadmiarowości, uporządkowana, ze strukturą hierarchii, skalowalna;

e) reprezentacja morfologiczna, z wydzieleniem składników, semantyczną kompozycją treści przekazu informacji.

2.1.3 Reprezentacja informacji

W różnego typu zastosowaniach teleinformatycznych, multimedialnych, widzenia maszynowego, obrazowania medycznego, przemysłowych, itd. metody reprezentacji danych obrazowych nabierają szczególnego znaczenia. Reprezentacja źródłowa, czyli pozyskana w procesie akwizycji/rejestracji danych, jest z natury nadmiarowa, bo zakłada *a priori* maksymalny zakres dopuszczalnych zmian , zgodnie z naturalnie zróżnicowaną dynamiką rejestrowanego sygnału oraz realiami systemu akwizycji. Przykładowo, reprezentacja danych obrazowych ma zwykle postać ciągu słów kodu dwójkowego o rozmiarze 8 bitów/piksel przy założeniu skali szarości lub 24 bitów/piksel dla formy obrazu w skali barw RGB. Odpowiada to dynamice przetworników a/c, ośmiobitowych dla każdego komponentu, często stosowanych w urządzenia rejestracji obrazów. Rejestracja dźwięku przy typowej częstości próbkowania 44 lub 96 kHz daje typowo ciąg 16 lub 24 bitowych próbek zapisanych w kodzie dwójkowym. Przy ograniczonej dynamice rejestrowanego sygnału redundantna reprezentacja danych utrudnia ich przekaz, archiwizację, analizę, a nawet wizualizację czy odsłuch. Taką nadmiarowość nazywamy syntaktyczną.

Kody, czyli reguły tworzenia nowych, bardziej upakowanych sekwencji bitowych reprezentujących dane, pozwalają uzyskać nowe formy reprezentacji danych – o zredukowanym rozmiarze, o większej odporności na zakłócenia, porządkujące występowanie danych w strumieniu (np. w formie progresji od ogółu treści do szczegółu) itp.

Metody kodowania wykorzystują proste mechanizmy modelowania danych, jak powtarzające się serie identycznych symboli (metoda kodowania długości serii) czy też zróżnicowana częstość występowania poszczególnych symboli alfabetu źródła danych (kod Huffmana). Bardziej zaawansowane kody bazują na transformacji danych do nowej dziedziny, dającej reprezentacje upakowaną, skalowalną, a nawet naturalnie uporządkowaną w sensie przyjętego kryterium progresji jakości danych (dziedzina falkowa w algorytmie kodowania standardu JPEG2000[1]). Możliwa jest też ingerencja odbiorcy w proces kodowania danych, gdzie za pomocą interaktywnego protokołu nadaje on kształt przekazu strumienia informacji definiując swoje potrzeby (interaktywny protokół JPIP[2]).

Rozumienie danych, czyli treść

Treść przypisana zbiorowi czy strumieniowy danych, odgrywająca kluczową rolę w przekazie informacji, związana jest bezpośrednio z naturą danych, techniką akwizycji i formowania postaci wyjściowej, określonym przeznaczeniem, intencjami nadawcy czy specyfiką rejestrowanego zjawiska. **Treść** rozumiana jest jako sens przekazu danych, jego wymowa koncepcyjna, ideologiczna. To wszystko, co można odkryć, zrozumieć, odczytać, analizując określony ciąg danych. Odczytanie znaczenia słów, w które układa się forma danych, właściwe ich skojarzenie w znaczenie, semantykę przekazu stanowi podstawę właściwej interpretacji danych.

Warunkiem rozumienia treści jest rozpoznanie szczegółów przekazu, percepcja wszystkich istotnych właściwości występujących elementów składowych, detekcja obiektów o rozpoznanym znaczeniu czy też grupy obiektów wraz z ich wzajemnymi odniesieniami. Rozpoznanie komputerowe naśladuje ludzkie poprzez wstępne wydzielenie obiektów i opisanie ich właściwości za pomocą dobranych deskryptorów, a następnie algorytmiczną realizację rezonansu poznawczego. Chodzi tutaj o skojarzenia parametrycznych charakterystyk obiektów ze sformalizowaną wiedzą specjalistyczną danej dziedziny, doświadczeniem gromadzonym latami w podobnych okolicznościach.

Jeśli rozpoznanie treści dokonujące się w głowach odbiorców nie sposób przełożyć na formalny model wiedzy i doświadczenia, obiektywny opis znaczeniowy treści staje się praktycznie niemożliwy. Rola jaką odgrywa intuicja czy intelekt odbiorcy przy czytaniu treści nie zostało opisane formalnie. Pozostaje jedynie naśladowanie rozumowego wnioskowania.

[1] http://www.jpeg.org/jpeg2000/
[2] http://www.jpeg.org/jpeg2000/j2kpart9.html

Odbiór informacji bazuje na rozumieniu treści, przy czym ważną role odgrywa także właściwa jej interpretacja. Znajdująca się na wyższym poziomie abstrakcji interpretacja treści przekazu, czyli ocena zasadniczej wymowy odczytanej treści stanowi jedno z najbardziej ambitnych zadań inteligencji obliczeniowej, a właściwie obliczeniowej mądrości [14].

Informacja, czyli chciana treść

Kluczowym warunkiem udanego przekazu informacji jest znaczenie przesyłanych za pomocą danych treści, której reprezentacja winna umożliwić skuteczny jej odbiór na sposób zgodny ze zdolnościami percepcji treści przez odbiorcę. Semantyka, czyli znaczenie danych formułuje treść przekazu, a ta w mniejszym lub większym stopniu staje się użyteczną dla odbiorcy informacją. Informacja z założenia stanowi istotę każdego sensownego przekazu danych, służy odbiorcy w zaspokojeniu określonych potrzeb. Personifikowany nadawca zaspokaja potrzeby odbiorcy realizując swoje cele. Gdy nadawcą jest "natura", podglądana, rejestrowana – odkrywamy wtedy jej tajemnice zdobywając informacje i budując wiedzę. Wymiana informacji jest podstawową funkcją życiową, wydaje się warunkiem koniecznym istnienia każdej społeczności, która trwa.

Informacja wynika z treści przekazu strumienia danych, która okazuje się znacząca dla odbiorcy. Znacząca, czyli coś daje, do czegoś się przydaje, zaspokaja określone potrzeby. Nie zawsze chodzi tutaj o dostarczenie nowych wiadomości, zobaczenie nowego filmu czy spektaklu, wideorozmowę z osobą, której nie widzieliśmy kilka lat. Czasami chcemy posłuchać ulubionej muzyki, przypomnieć sobie wzruszający serial sprzed lat, powtórzyć czy odświeżyć wiedzę, bo tego właśnie nam potrzeba, bo taki jest nastrój czy wymóg chwili. Można także dokonać wyboru treści znaczących w sposób arbitralny, niekiedy nawet wbrew woli odbiorcy, by uświadomić mu pewne fakty, pouczyć, narzucić konieczność konfrontacji z określoną tematyką, itp.

Takie **subiektywne rozumienie informacji** jest w dużym stopniu niejednoznaczne, z trudem poddaje się formalizacji zobiektywizowanego opisu, algorytmicznej procedurze ustalania warunków przekazu np. multimedialnego[3]. Selekcja treści, uporządkowanie, ustalenie względności używanych pojęć i liczb, hierarchia opisu wymaga przyjęcia pewnego modelu odbiorcy, który z natury musi być uproszczony, uogólniony, schematyczny. Brakuje formalnych rozwiązań, które dostosowują się do potrzeb indywidualnego odbiorcy. Przekaz jest więc często wspomagany różnymi formami interakcji.

Nieco inne rozumienie informacji, zakładające pewne ujednolicenie opinii dotyczących wartości przesyłanych danych, bazuje na fakcie, że pozyskanie informa-

[3]Przekaz multimedialny znaczy wielostrumieniowy, ze znacznikami czasu rzeczywistego, synchronizacją treści poszczególnych strumieni, naśladujący w pewnym stopniu uwarunkowania przekazu ludzkiego.

cji związane jest z pewnym kosztem. Koszt ten, wynikający z charakteru przekazywanych treści oraz przyjętej reprezentacji danych, jest zazwyczaj mniejszy od korzyści wynikających z jej użytkowania. Zysk mierzony różnicą wartości korzyści uzyskanych wskutek przekazu informacji w odniesieniu do poniesionych kosztów jest miarą ilości informacji. Stąd jeśli koszty przerosły zyski, przekazane dane nie były informacją. Ocena ilości informacji jest w tym przypadku możliwa jedynie w analizie retrospektywnej.

Matematyczna teoria informacji, której podstawy sformułowano pod koniec lat czterdziestych zeszłego wieku [21], zakłada zobiektywizowane pojęcie informacji, umożliwiające ilościową charakterystykę informacji, tworzenie modeli źródeł informacji oraz zasad zniekształceń tych źródeł, a także konstruowanie kodów dopasowanych do specyfiki danych. Claude E. Shannon, uznawany za twórcę matematycznej teorii informacji, wprowadził rozdzielenie pojęcia informacji oraz semantyki przekazu twierdząc, że semantyka nie jest istotna przy rozwiązywaniu inżynierskich problemów komunikacji. Informacja przekazu dotyczy jedynie wyboru jednej z dostępnych możliwości źródłowych. Stąd informacja została zdefiniowana jako **poziom niepewności odbiorcy** dotyczącej przekazywanych danych. Wśród transmitowanych danych tylko te zawierają informacje, które pozostają nieokreślone czy nieprzewidywalne (odbiorca nie ma pewności, jakie dane otrzyma). Po ich otrzymaniu poziom niepewności odbiorcy maleje.

W matematycznej teorii informacji rozważany jest więc jedynie transmisyjny (syntaktyczny, z analizą postaci informacji), a nie semantyczny aspekt informacji. Znaczy to, że nie prowadzi się formalnych rozważań dotyczących prawdziwości czy znaczenia tego, co jest przesyłane. Informacja rozumiana jest wtedy jako ciąg danych – symboli nad ustalonym alfabetem, z określonym prawdopodobieństwem ich występowania. Przyjęto więc probabilistyczny model źródła informacji. Podstawy matematycznej teorii informacji określają metody opisu źródeł informacji, kodowania tych źródeł oraz teorie zniekształceń źródeł informacji.

Źródło informacji opisane jest w pierwszym przybliżeniu parą zbiorów (A_S, P_S). A_S jest alfabetem źródła, czyli zbiorem wszystkich symboli – postaci danych, jakimi wyrażana jest informacja (inaczej zbiorem informacji elementarnych), a P_S to rozkład wartości prawdopodobieństw wystąpienia poszczególnych symboli alfabetu źródła o liczności odpowiadającej liczbie symboli alfabetu dla źródeł określonych, $|A_S| = |P_S|$.

Współczesny rozwój technologii teleinformatycznych oraz coraz bardziej istotna rola przekazu informacji w życiu społecznym prowadzi do rosnącej liczby zastosowań, które odwołują się do semantyki przekazu, a uproszczony model probabilistyczny przekazu informacji staje się niewystarczający. Wśród wielu przykładów można wskazać wprowadzenie wspomnianego protokołu JPIP w ramach standardu JPEG2000, czy też wymagania zastosowań medycznych, przede wszystkim kodowania obrazów w celach archiwizacji lub transmisji w systemach telediagnozy

z zachowaniem wiarygodności diagnostycznej obrazów. Znaczenie pojedynczych pikseli, grup pikseli, obiektów i wzajemnych relacji definiujących treść jest tutaj kluczowe.

Semantyka przekazywanej informacji odgrywa na tyle znaczącą rolę w jej użytkowaniu przez odbiorcę, zrozumieniu, ocenie, interpretacji, że winna stanowić ważny element modelowania źródeł informacji. Przykładowe rozszerzenie definicji źródła informacji o alfabet znaczeń poszczególnych symboli Σ_S prowadzi do modelu (A_S, P_S, Σ_S). Kolejnym, niezwykle istotnym aspektem w przekazie informacji jest jej prawdziwość. L. Floridi zdefiniował pojęcie semantycznej informacji jako ciąg danych dobrze uformowanych (reprezentowanych), znaczących (z niezerowym opisem semantycznym) oraz prawdziwych. Dane określonej treści, mające znaczenie dla odbiorcy tylko wtedy stanowią informację, gdy są prawdziwe. Choć taka definicja wydaje się z inżynierskiego punktu widzenia bardzo wymagająca, niewątpliwie stanowi ona pełny i wiarygodny opis pojęcia informacji [15].

Ustalenie dobrze uformowanej, tj. skutecznej w danym zastosowaniu, reprezentacji informacji powinno więc odwoływać się więc zarówno do znaczenia jak i prawdziwości, czy inaczej wiarygodności danych źródłowych.

Jedną z podstawowych metod optymalizacji przekazu informacji jest dobór efektywnej reprezentacji dostarczanych danych źródłowych. Najlepiej jak jest to reprezentacja informacji semantycznej w postaci zwartej – upakowanej, czyli rzadkiej (ang. *sparse*) w sensie wymiaru dziedziny źródłowej oraz uporządkowanej (skupionej w niewielkim zakresie dziedziny). Taka reprezentacja pozwala na bardziej efektywną realizację procedur kodowania, przetwarzania, analizy, ekstrakcji treści użytecznej, selekcji informacji, itp. W przypadku zastosowań medycznych zwiększa skuteczność systemów komputerowego wspomagania diagnostyki obrazowej, rozpoznawania patologii, wydobywania treści ukrytych, czyli niedostrzegalnych w ocenie radiologa.

Dobór reprezentacji

Reprezentacja źródłowa Jedną z podstawowych metod optymalizacji przekazu informacji jest dobór efektywnej reprezentacji dostarczanych danych źródłowych. Rejestracja informacji z wykorzystaniem określonego sygnału wprowadza w sposób oczywisty zależności pomiędzy ciągami wartości sygnału, bo taka jest natura każdej informacji. Przekłada się to na nadmiarowość reprezentacji źródłowej.

Treść wyrażana jest za pomocą określonych obiektów i wzajemnych relacji. Więcej różnorodnych, stosunkowo niewielkich obiektów luźno ze sobą powiązanych przekłada się na wzrost ilości informacji zawartej w sygnale. Duże, jednorodne, podobne do siebie obiekty będące wyrazem treści oznaczają małą ilość informacji, dużą zależność danych, a więc silną nadmiarowość reprezentacji źródłowej, zwaną nadmiarowością stochastyczną.

Rzeczywistej rejestracji sygnału towarzyszy także zapis szumu, czyli składo-

wej wprowadzającej losowość zmian kolejnych wartości rejestrowanego sygnału. Redukcja zależności pomiędzy danymi powoduje w przypadku wzrostu energii szumów wyraźne zwiększenie entropii, rozumianej w tym przypadku jako miara nieuporządkowania. Niestety, sposób liczenia entropii nie pozwala wskazać przyczyny wzrostu jej wartości - nie wiemy, czy przybywa informacji czy też nieuporządkowanego szumu. Występowanie szumu powoduje nadmiarowość znaczeniową (semantyczną), której poziom można ustalić m.in. za pomocą semantycznych deskryptorów numerycznych, dostosowanych do specyfiki obrazów.

Modelowanie sygnałów w celu ich kodowania, przetwarzania, analizy, ekstrakcji informacji, itp. jest bardziej użyteczne, jeśli bazuje na zwartym opisie sygnału. W przypadku naturalnych źródeł informacji zwarta, czyli upakowana reprezentacja sygnału jest rzadka (ang. *sparse*) w stosunku do wymiaru dziedziny źródłowej.

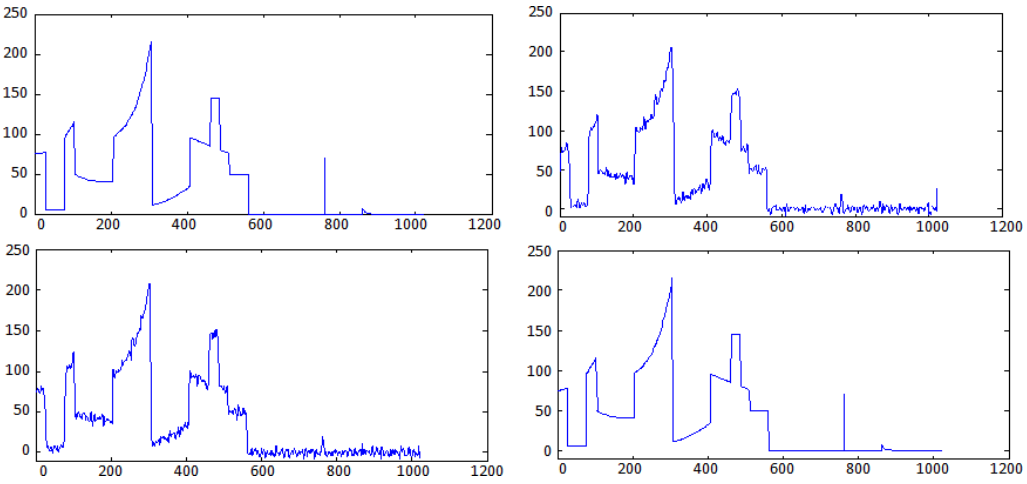
Możliwe jest wykorzystanie przekształceń $\mathcal{P} : f \rightarrow w$ powodujących dekorelację czy nawet dających pełną niezależność danych. Przekształcenia te tworzą upakowaną, jednoznaczną reprezentację, która jest rzadka i uporządkowana w sensie lokalnego skupienia energii sygnału (przenoszącego informację) w niewielkim zakresie dziedziny przekształcenia. Znaczy, to że liczba niezerowych współczynników w obszarze tej dziedziny jest znikomo mała, czyli realny wymiar nowej dziedziny reprezentacji informacji został znacząco zredukowany. Taki zwarty opis sygnału daje zwykle jedynie przybliżoną postać wersji źródłowej, najlepiej przy zachowaniu wszystkich istotnych jego cech, a usunięciu nadmiarowości semantycznej.

Przyjmując bardziej formalnie pewną złożoność rozważanego problemu, można założyć, że sygnał f składa się z K składników o różnej morfologii $f = \sum_{i=1}^{K} f_i$, np. obraz składa się z kilku obiektów o różnej morfologii (charakterystyce teksturowej, kształcie itp.), a ponadto istnieje słownik baz φ, tj. zbiorów wektorów bazowych, służących efektywnej reprezentacji sygnałów. Przyjmuje się, że każdy ze składników f_i może mieć reprezentację $\mathcal{R}_i = \varphi_i^T f_i$, uzyskaną za pomocą określonej bazy φ_i. Celem jest dobranie reprezentacji możliwie rzadkiej w sensie pseudo-normy l_0: $\|\varphi_i^T f_i\|_0$, gdzie $\|x\|$ oznacza liczbę niezerowych współczynników wektora x. Dążymy więc do $\min \sum_{i=0}^{K} \|\varphi_i^T f_i\|_0$ dobierając odpowiednie bazy dla poszczególnych składników. Najprostszy przypadek dla $K = 1$ sprowadza się do poszukiwania bazy dającej możliwie rzadką, czyli maksymalnie upakowaną reprezentację f. Znając charakterystykę zróżnicowanych składowych sygnału, które stanowią informację obrazową, dobieramy bazy maksymalnego upakowania oddzielnie dla każdego z potencjalnie niezależnych komponentów (obiektów) obrazu.

Przykładowo, na rys. 3.33 pokazano przybliżenia obrazów testowych, uzyskane za pomocą upakowanej reprezentacji w kilku różnych bazach – funkcji falkowych, falek geometrycznych - kliników, falek kierunkowych - krzywek oraz funkcji szeregu fourierowskiego. Zalety bazy falkowej widać na rys. 2.2, gdzie uzyskano bardzo

wierny obraz sygnału za pomocą falkowej reprezentacji o wymiarze stanowiącym zaledwie 15% wymiaru dziedziny źródłowej.

Rysunek 2.1: Efekty opisu dwóch obrazów testowych (barbara i goldhill) za pomocą upakowanej reprezentacji; od lewej kolejno obrazy źródłowe o rozmiarze 512×512×8 bitów oraz ich przybliżenia z 13,6% współczynników obrazów, uzyskanych za pomocą bazy falkowej, wedgeletowej (klinków), curveletowej (krzywek) oraz fourierowskiej.



Rysunek 2.2: Przybliżenie sygnału źródłowego za pomocą reprezentacji zredukowanej do zaledwie 15% wymiaru dziedziny źródłowej; kolejno od lewej do prawej, zaczynając od góry - sygnał źródłowy oraz przybliżenia za pomocą baz fourierowskiej, funkcji dyskretnej transformacji kosinusowej oraz bazy falkowej; w przypadku funkcji sinusoidalnych o nieskończonym nośniku widoczne są charakterystyczne oscylacje przy krawędziach o dużym gradiencie, w punktach nieciągłości - efekt Gibbsa.

2.2 Nośniki informacji

Nośniki informacji, czyli ogólnie sygnały są bardzo ważnym elementem konstruowanych aplikacji multimedialnych. Sposób ich definiowania, opisu, kształtowania, modulacji treścią są nierozerwalnie związanie z zasadniczym celem skutecznego przekazu informacji. Dopasowanie sygnałów do charakteru treści, istotnych właściwości przesyłanych danych, ale też do natury opisywanego zjawiska czy faktu jest fundamentalnym, bo bardzo pragmatycznym zagadnieniem inżynierii multimediów.

2.2.1 Wyrażanie informacji

Treść przekazu staje się informacją w określonych okolicznościach. Treść ta może mieć charakter immanentny lub transcendentny. Przekaz immanentny towarzyszy zwykle rejestracji jakiegoś zjawiska fizycznego, obserwacji jego niedostepnej natury, odczytu stanu czujników śledzących przebieg zakrytych przed obserwatorem zdarzeń, wymaga odpowiedniego, często specjalistycznego odczytu przez fachowców, a interpretacja danych ma wtedy zawsze charakter informacji (stwierdzenie, że nic się nie dzieje w interesującym obszarze też jest informacją).

Przekaz transcendentny jest zwykle zamierzony, zbudowany na bazie zewnętrznych, uogólniających obserwacji, treść znamionująca informację jest ogólnie rozpoznawalna, podobnie interpretowalna, rozumiana dość jednozacznie tak przez nadawcę, jak i przez typowego odbiorcę, a wybór formy, sposobu i technologii przekazu jest zwykle dobierany ze względu na charakter i właściwości tej treści.

Możliwe są też rozwiązania hybrydowe, niejednoznaczne, wynikające np. z różnego rozumienia treści przez nadawcę i odbiorcę (to co było celem przekazu i miało stanowić informację okazało się nieistotne, natomiast inna właściwość przekazanej treści może okazać się przydatna odbiorcy). Może to niekiedy powodować zniekształcenie przekazu ze względu na nieodpowiednio dobraną technologię przekazu.

Różna w charakterze i formie treść może być zawarta w sygnale ciągłym, dyskretnym, cyfrowym. W grę może wchodzić zbiór danych o charakterze jednolitym, zbiory danych wzajemnie referujące na siebie, ze znacznikiem upływającego czasu lub też asynchronicznym odwoływaniem się do pewnej sekwencji zdarzeń, itp.

2.2.2 Podstawowe przestrzenie opisu sygnałów

Chcąc przybliżać sygnał poprzez selekcję zawartej w nim informacji, często słabo dostrzegalnej, subtelnej bądź wręcz ukrytej, konieczna jest reprezentatywna charakterystyka klasy sygnałów, które będą analizowane. Służy temu zdefiniowanie przestrzeni obiektów (wektorów), do której należą interesujące nas sygnały i określanie ich właściwości. Stosując terminologię funkcjonalnej analizy sygnałów

możemy mówić o aproksymacji **funkcji celu** w sygnale opisanym jako funkcja źródłowa.

Pokrótce przedstawiono kolejne przestrzenie od najbardziej ogólnych, do tych bardziej użytecznych, które pozwolą zarysować metodologię aproksymacji najbardziej istotnych cech realnych sygnałów, w szczególności informacji obrazowej. Zebrano podstawowe definicje i stosowne formalizmy by dokładniej przedstawić teoretyczne podstawy metod tworzenia reprezentacji informacji.

Przestrzeń liniową tworzą obiekty, które mogą być skalowane i dodawane.

Definicja 2.2 *Przestrzeń liniowa*
Przestrzenią liniową nad ciałem liczb rzeczywistych $\mathbb{R}$ (ogólniej zespolonych $\mathbb{C}$) nazywamy zbiór $\mathcal{L}$ obiektów (wektorów), dla którego określono dwa działania: dodawanie i mnożenie przez liczbę (skalar) tak, że dla dowolnych $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{L}$ oraz $a, b \in \mathbb{R}$ (lub ogólniej $a, b \in \mathbb{C}$) spełnione są następujące aksjomaty:

1. przemienność dodawania
   $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$,

2. łączność dodawania
   $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$,

3. rozdzielność dodawania względem mnożenia przez liczbę
   $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{y} + a\boldsymbol{x}$,

4. element neutralny dodawania
   $\forall_{\boldsymbol{x}\in\mathcal{L}}$: istnieje element $\mathbf{0} \in \mathcal{L}$ taki, że $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$

5. element przeciwny dodawania
   $\forall_{\boldsymbol{x}\in\mathcal{L}}$: istnieje element $-\boldsymbol{x} \in \mathcal{L}$ taki, że $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$,

6. element neutralny mnożenia
   $\forall_{\boldsymbol{x}\in\mathcal{L}}$: $\mathbf{1} \cdot \boldsymbol{x} = \boldsymbol{x}$.

□

Niech liniowa $\mathcal{L}$ będzie przestrzenią ciągów $\boldsymbol{x} = (x_1, x_2, \ldots)$, $\boldsymbol{y} = (y_1, y_2, \ldots)$, $\boldsymbol{z} = (z_1, z_2, \ldots)$ itd. Wtedy dodawanie definiowane jest jako:
$\boldsymbol{x} + \boldsymbol{y} = (x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (x_1 + y_1, x_2 + y_2, \ldots)$,
a mnożenie przez liczbę jako: $a\boldsymbol{x} = a(x_1, x_2, \ldots) = (ax_1, ax_2, \ldots)$.

Ustalmy podzbiór $\mathcal{B}$ składający się z n elementów liniowej przestrzeni $\mathcal{L}$, tak że $\mathcal{B} = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots, \boldsymbol{x}_n\}$ i $\forall_{i=1,\ldots,n}\boldsymbol{x}_i \in \mathcal{L}$. Elementy te są **liniowo niezależne**, jeśli $\sum_{i=1}^{n} a_i\boldsymbol{x}_i = 0$ jedynie wtedy, gdy $\forall_{i=1,\ldots,n}a_i = 0$. Liniowa niezależność zbioru elementów danej przestrzeni oznacza, że żadnego z tych elementów nie może przedstawić za pomocą liniowej kombinacji pozostałych.

Jeżeli dodatkowo $\mathcal{B}$ generuje całą $\mathcal{L}$, tj. za pomocą liniowych kombinacji elementów $\mathcal{B}$ można uzyskać dowolny element $\mathcal{L}$:

$$\mathcal{L} = \left\{ \sum_{i=1}^{n} a_i \boldsymbol{x}_i | a_i \in \mathbb{R} \text{ lub } \mathbb{C}, \boldsymbol{x}_i \in \mathcal{B} \right\}$$

wówczas podzbiór $\mathcal{B}$ tworzy **bazę przestrzeni** $\mathcal{L}$: $\mathcal{B} = \mathcal{B}_{\mathcal{L}}$. Inaczej, zbiór liniowo niezależny o maksymalnej liczbie elementów przestrzeni jest bazą tej przestrzeni, a liczbę elementów (wektorów) bazowych (moc zbioru bazy danej przestrzeni) nazywamy **wymiarem przestrzeni**. Przestrzeń nieskończenie wymiarowa zawiera nieskończony zbiór liniowo niezależnych wektorów.

Dla dowolnego $\boldsymbol{y} \in \mathcal{L}$ mamy więc:

$$\boldsymbol{y} = \sum_{i=1}^{n} a_i \boldsymbol{x}_i$$

gdzie $\{a_i\}_{i=1}^{n}$ jest zbiorem współczynników tego wektora względem elementów bazy $\mathcal{B}_{\mathcal{L}}$, nazywanym **reprezentacją sygnału** względem danej bazy.

O przestrzeni $\mathcal{L}$ mówi się, że jest rozpięta na elementach zbioru (bazy) $\mathcal{B}_{\mathcal{L}}$, czyli $\mathcal{L} = \mathbf{span}(\mathcal{B}_{\mathcal{L}})$. Mówi się też, że baza generuje przestrzeń. Badanie właściwości baz oraz wynikającej stąd specyfiki rozpinanych przez nie przestrzeni stanowi kluczowe zagadnienie teorii aproksymacji. Aby bliżej przyjrzeć się podstawowym własnościom baz przestrzeni konieczne jest dookreślenie użytecznych przestrzeni sygnałów.

W zagadnieniach aproksymacji sygnałów istotnego znaczenia nabierają szczególne przypadki przestrzeni liniowych ze zdefiniowaną normą (długością) i metryką (odległością wektorów), posiadające istotną cechę zupełności i określony iloczyn skalarny.

Definicja 2.3 *Przestrzeń unormowana*
Przestrzenią unormowaną nazywamy przestrzeń liniową nad ciałem liczb rzeczywistych $\mathbb{R}$ (ogólniej zespolonych $\mathbb{C}$), w której dowolnemu $\boldsymbol{x}$ przyporządkowano normę (tj. długość) jako liczbę rzeczywistą nieujemną $\|\boldsymbol{x}\|$ spełniającą następujące warunki:

a) $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = 0$

b) $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$

c) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

□

Ważnym przykładem przestrzeni unormowanej jest służąca opisowi świata rzeczywistego przestrzeń euklidesowa. Norma różnicy dwóch dowolnych wektorów przestrzeni liniowej $\|\boldsymbol{x} - \boldsymbol{y}\|$ generuje **metrykę** ich odległości, spełniającą analogiczne warunki:

a) $\|\boldsymbol{x}-\boldsymbol{y}\|=0 \Leftrightarrow \boldsymbol{x}=\boldsymbol{y}$

b) $\|\boldsymbol{x}-\boldsymbol{y}\|=\|\boldsymbol{y}-\boldsymbol{x}\|$

c) $\|\boldsymbol{x}-\boldsymbol{z}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|+\|\boldsymbol{y}-\boldsymbol{z}\|$

Zbiór wektorów wraz z metryką zdefiniowaną na tym zbiorze nazywamy **przestrzenią metryczną**. Istotną rolę w przestrzeni metrycznej odgrywa zbieżny ciąg jej elementów nazywany ciągiem podstawowym. Ciąg $\{\boldsymbol{x}_n\}$ jest **ciągiem podstawowym**, inaczej **ciągiem Cauchy'ego**, jeśli odległość $\|\boldsymbol{x}_n-\boldsymbol{x}_m\| \to 0$ dla $n, m \to \infty$.

Przestrzeń metryczna, w której każdy ciąg podstawowy elementów tej przestrzeni jest zbieżny w tej przestrzeni (tj. ma granicę należącą do tej przestrzeni) jest **przestrzenią zupełną**. Ponadto, **przestrzeń metryczna jest zwarta**, jeśli jest zupełna i całkowicie ograniczona (tj. dla każdego $\epsilon > 0$ można ją pokryć skończenie wieloma zbiorami o średnicach[4] mniejszych lub równych ϵ).

Definicja 2.4 *Przestrzeń Banacha*
Przestrzeń unormowaną zupełną nazywamy przestrzenią Banacha.
□

W przestrzeni Banacha norma określa metrykę pozwalającą sprawdzić warunek zupełności. W analizie sygnałów ważną rolę odgrywa szczególny przypadek przestrzeni Banacha, to jest przestrzeń Hilberta zawierająca iloczyn skalarny pochodzący od normy. Iloczyn skalarny pozwala między innymi charakteryzować bazy przestrzeni zupełnych i definiować operacje na sygnałach (obiektach, wektorach) tych przestrzeni.

Definicja 2.5 *Przestrzeń unitarna*
Przestrzenią unitarną nazywamy przestrzeń liniową nad ciałem liczb rzeczywistych $\mathbb{R}$ (ogólniej zespolonych $\mathbb{C}$), w której określono iloczyn skalarny $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ jako funkcjonał dwuargumentowy o następujących właściwościach:

a) $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$

b) $\langle a\boldsymbol{x}, \boldsymbol{y} \rangle = a\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ dla $a \in \mathbb{C}$

c) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle^* = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$

d) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$

e) jeśli $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$, to $\boldsymbol{x} = \boldsymbol{0}$

□

[4] Średnica zbioru to supremum odległości wszystkich par elementów tego zbioru

Zdefiniowanie iloczynu skalarnego pozwala określić **normę wektora** jako: $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

Definicja 2.6 *Przestrzeń Hilberta*
Przestrzeń unitarna zupełna, której norma określona jest przez iloczyn skalarny, nazywana jest przestrzenią Hilberta.
□

Przestrzenie Hilberta są podstawowym pojęciem wykorzystywanym w analizie funkcjonalnej[5].

Szczególnym przypadkiem przestrzeni Hilberta są: **przestrzeń funkcji całkowalnych z kwadratem** $\mathrm{L}^2(\mathbb{R})$ opisująca sygnały ciągłe o skończonej energii oraz **przestrzeń funkcji sumowalnych z kwadratem** $\mathrm{l}^2(\mathbb{Z})$ dla sygnałów dyskretnych o skończonej energii. Przestrzenie te dobrze reprezentują źródłowe sygnały rzeczywiste pochodzące z określonych urządzeń akwizycji lub rejestracji (źródeł sygnałów zawierających informację).

Sygnał jest opisany funkcją ciągłą $x(t) \in \mathrm{L}^2(\mathbb{R})$, jeśli $|x(t)|^2$ jest całkowalne, czyli:

$$\int_{t\in\mathbb{R}} |x(t)|^2 dt < \infty$$

W przestrzeni tej iloczyn skalarny jest zdefiniowany jako $\langle x(t), y(t) \rangle_{\mathrm{L}^2} = \int_{t\in\mathbb{R}} x(t)y(t)^* dt$, norma $\|x(t)\|_{\mathrm{L}^2} = \sqrt{\langle x(t), x(t) \rangle} = \sqrt{\int_{t\in\mathbb{R}} |x(t)|^2 dt}$, a metryka $\|x(t) - y(t)\|_{\mathrm{L}^2} = \sqrt{\int_{t\in\mathbb{R}} |x(t) - y(t)|^2 dt}$.

Analogicznie, sygnał dyskretny o wartościach $x_i \in \mathbb{R}$ lub $\mathbb{C}$ jest sumowalny z kwadratem, czyli $x_i \in \mathrm{l}^2(\mathbb{Z})$, jeśli:

$$\sum_{i\in\mathbb{Z}} |x_i|^2 < \infty$$

W przestrzeni tej iloczyn skalarny jest zdefiniowany jako $\langle x_i, y_i \rangle_{\mathrm{l}^2} = \sum_{i\in\mathbb{Z}} x_i y_i^*$, norma $\|x_i\|_{\mathrm{l}^2} = \sqrt{\langle x_i, x_i \rangle} = \sqrt{\sum_{i\in\mathbb{Z}} |x_i|^2}$, a metryka $\|x_i - y_i\|_{\mathrm{l}^2} = \sqrt{\sum_{i\in\mathbb{Z}} |x_i - y_i|^2}$.

Bazy przestrzeni Hilberta

Ważnym przypadkiem wzajemnej relacji dwóch wektorów jest ich zerowy iloczyn skalarny. Mówimy, że wektor $\boldsymbol{x}$ jest **ortogonalny** do wektora $\boldsymbol{y}$ (co zapisujemy jako $\boldsymbol{x} \perp \boldsymbol{y}$), jeśli zachodzi warunek: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. Jeśli dodatkowo długość każdego z ortogonalnych wektorów jest równa 1, czyli $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, wówczas wektory te nazywamy *ortonormalnymi*. Zbiór dowolnej liczby wektorów $\{\boldsymbol{x}_i | i = 1, 2, \ldots\}$ jest ortogonalny, jeśli wszystkie wektory tego zbioru są parami ortogonalne. Analogicznie zbiór ten jest ortonormalny, jeśli spełniony jest warunek jednostkowej długości każdego wektora zbioru. Dla zbioru ortonormalnego

[5] Analiza funkcjonalna to obszar analizy matematycznej zajmujący się badaniem właściwości przestrzeni funkcyjnych (sygnałowych)

mamy więc: $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \delta_{i,j}$[6]. Jeśli bazę przestrzeni sygnałów tworzy zbiór funkcji ortogonalnych (lub ortonormalnych), wówczas mówimy o bazie ortogonalnej (ortonormalnej).

Rozszerzając pojęcie ortogonalności (i analogicznie ortonormalności), $\boldsymbol{y}$ jest ortogonalny do podprzestrzeni unitarnej $\mathcal{P} = \{\boldsymbol{x}_i | i = 1, \ldots, n\}$, czyli $\boldsymbol{y} \perp \mathcal{P}$, jeśli $\forall_{i=1,\ldots,n} \boldsymbol{y} \perp \boldsymbol{x}_i$. Jeszcze bardziej ogólnie, dwie podprzestrzenie unitarne $\mathcal{P}_1$ oraz $\mathcal{P}_2$ są ortogonalne, jeśli wszystkie wektory w jednej przestrzeni są ortogonalne do wszystkich wektorów w drugiej przestrzeni, co zapisujemy jako $\mathcal{P}_1 \perp \mathcal{P}_2$.

Jeśli zbiór funkcji ortonormalnych $\{x_i(t) = x_i\}_{i=1,2,\ldots}$ stanowi bazę przestrzeni Hilberta $\mathcal{H}$, to dowolny sygnał $y(t) = y$ tej przestrzeni ma reprezentację postaci $y = \sum_{i=1}^{\infty} a_i x_i$, a współczynniki a_i odpowiadają ortogonalnym rzutom y na odpowiedni element bazy $a_i = \langle y, x_i \rangle$ (inaczej a_i określa wartość rzutu na kierunek wektora bazowego x_i - przy silniejszej kierunkowej zbieżności sygnału i wektora bazy współczynnik ten ma większą wartość).

Ogólniej, **rzutem ortogonalnym** funkcji (wektora) $y \in \mathcal{H}$ na podprzestrzeń $\mathcal{S} \subset \mathcal{H}$ nazywamy funkcję (wektor) $\tilde{y} \in \mathcal{S}$, dla której różnica $y - \tilde{y}$ jest ortogonalna do $\mathcal{S}$. Jeśli $\mathcal{S}$ jest domknięta, to każdy wektor $y \in \mathcal{H}$ ma rzut ortogonalny na $\mathcal{S}$.

Jeśli $\{x_i(t) = x_i\}_{i=1,2,\ldots,n}$ jest bazą ortonormalną podprzestrzeni $\mathcal{H}_n \subset \mathcal{H}$, to rzut ortogonalny $\tilde{y} = \sum_{i=1}^{n} a_i x_i$, gdzie $a_i = \langle y, x_i \rangle$, jest najlepszym, w sensie minimalnej odległości, przybliżeniem $y \in \mathcal{H}$ w przestrzeni $\mathcal{H}_n$, tj.

$$\tilde{y} = \arg \min_{x \in \mathcal{H}_n} \|y - x\|$$

Rzuty ortogonalne na kolejne podprzestrzenie z bazami ortogonalnymi o skończonym, rosnącym wymiarze stanowią coraz doskonalsze przybliżenia y z nieskończeniewymiarowej $\mathcal{H}$.

Twierdzenie 2.1 *O zbieżności rzutów ortogonalnych*
Jeśli zbiór funkcji ortonormalnych $\{x_i(t) = x_i\}_{i=1,2,\ldots}$ stanowi bazę przestrzeni Hilberta $\mathcal{H}$, to dla każdego sygnału $y(t) = y \in \mathcal{H}$ ciąg $(\tilde{y}_n)_{n \to \infty}$ rzutów przybliżających ten sygnał postaci $\tilde{y}_n = \sum_{i=1}^{n} a_i x_i$, gdzie $a_i = \langle y, x_i \rangle$, jest zbieżny do y. □

Kolejne uzupełnienia aproksymacji y w podprzestrzeniach rozpinanych przez ortonormalne $\{x_i(t) = x_i\}_{i=1,2,\ldots,n}$ można zapisać jako $\tilde{y}_n = \sum_{i=1}^{n-1} a_i x_i + \langle y, x_n \rangle x_n$.

Reprezentacja sygnału w bazach Hilberta prowadzi nas do przybliżeń sygnału w podzbiorach tej przestrzeni w postaci rzutów ortogonalnych, czyli do zagadnienia aproksymacji. Dostarcza ono użytecznych narzędzi w estymacji funkcji ukrytej (niejawnej) - nieznanej, poszukiwanej, docelowej funkcji (*target function*) maskowanej w analizowanym sygnale.

[6]Symbol $\delta_{i,j}$ oznacza deltę Kroneckera – dwuargumentową funkcję równą 1 dla $i = j$ oraz 0 dla $i \neq j$.

Bazy fourierowskie Bazy te tworzone są z wykorzystaniem szeregu Fouriera w postaci trygonometrycznej oraz znajdującej szersze zastosowanie w przetwarzaniu sygnałów postaci wykładniczej. Służą one do reprezentacji sygnałów okresowych $y(t)$, z okresem T przy $t \in (-\infty, \infty)$, całkowalnych z kwadratem. Z okresu wynika częstotliwość podstawowa $f_0 = 1/T$ i pulsacja $\omega_0 = 2\pi f_0$ reprezentacji y. Rozwinięcie $y(t)$ w bazie **trygonometrycznego szeregu Fouriera** wygląda następująco:

$$\tilde{y}(t) = a_0 + \sum_{i=1}^{\infty} \acute{a}_i \cos(i\omega_0 t) + \sum_{i=1}^{\infty} \grave{a}_i \sin(i\omega_0 t) = a_0 + \sum_{i=1}^{\infty} a_i \cos(i\omega_0 t + \vartheta_i) \quad (2.1)$$

gdzie $a_0 = 1/T \int\limits_{|T|} y(t)dt$ (całkowanie w dowolnym przedziale o długości T oraz dla $i = 1, 2, \ldots$: $\acute{a}_i = 2/T \int\limits_{|T|} y(t)\cos(i\omega_0 t)dt$, $\grave{a}_i = 2/T \int\limits_{|T|} y(t)\sin(i\omega_0 t)dt$ oraz $a_i = \sqrt{\acute{a}_i^2 + \grave{a}_i^2}$, zaś faza $\vartheta_i = -\arctan \frac{\grave{a}_i}{\acute{a}_i}$. Taka reprezentacja jest zbieżna do y, gdy spełnione są odpowiednie warunki (tzw. warunki Dirichleta):

1. $y(t)$ jest bezwzględnie całkowalna, czyli $\int\limits_{|T|} |y(t)|dt < \infty$;
2. $y(t)$ jest funkcją o ograniczonej zmienności w każdym ograniczonym przedziale, czyli nie ma ekstremów o nieskończonej wartości i liczba ekstremów jest skończona;
3. $y(t)$ ma skończoną liczbę nieciągłości w każdym ograniczonym przedziale.

Równość $\tilde{y}(t) = y(t)$ zachodzi prawie wszędzie, tj. z wyjątkiem punktów nieciągłości, w których $\tilde{y}$ jest średnią arytmetyczną granic lewo- i prawostronnej sygnału y.

Reprezentacja sygnału ciągłego $y(t)$, okresowego z okresem T w nieskończonej bazie fourierowskiej, składającej się z funkcji – elementów szeregu Fouriera **o postaci wykładniczej** zgodnie z formułą Eulera $x_i = e^{ji\omega_0 t}$ (przyjęto oznaczenie $j = \sqrt{-1}$) z częstotliwością podstawową (pierwszej harmonicznej) równą $\omega_0 = 2\pi/T$, wygląda następująco:

$$y(t) = \sum_{i=-\infty}^{\infty} b_i e^{ji\omega_0 t} \quad (2.2)$$

gdzie współczynniki $b_i = 1/T \int\limits_{-T/2}^{T/2} y(t)e^{-ji\omega_0 t}dt$, przy czym w odniesieniu do współczynników szeregu trygonometrycznego (2.1) jest następująca: $b_0 = a_0$, $b_i = (\acute{a}_i - \grave{a}_i)/2$, a $b_{-i} = b_i^*$ dla sygnałów rzeczywistych.

Przybliżenie $y(t)$ za pomocą skończonej liczby rzutów ortogonalnych na kolejne elementy bazy fourierowskiej wygląda następująco: $y(t) = \sum\limits_{i=-n}^{n} a_i e^{ji\omega_0 t}$. W

przypadku sygnału dyskretnego y_n o okresie N rozwinięcie w szereg Fouriera wygląda następująco:

$$y_n = \sum_{i=0}^{N-1} a_i e^{j2\pi i/Nn} \tag{2.3}$$

Reprezentacja sygnału w bazach fourierowskich według równań (2.2) lub (2.3), nazywana transformacją Fouriera, ma dwa podstawowe ograniczenia: zakładana okresowość sygnału korelująca z okresowością elementów bazy oraz utrata orientacji źródłowej dziedziny czasu czy przestrzeni. Nowa reprezentacja sygnałów za pomocą fourierowskich współczynników a_i ma charakter częstotliwościowy (widma częstotliwościowego), bo odzwierciedla rzuty sygnału na kolejne funkcje bazowe o liniowo rosnącej częstotliwości. Zatracona zostaje natomiast zupełnie informacja o czasowo-przestrzennym położeniu, tak istotnym w opisie wielu sygnałów rzeczywistych.

Reprezentacja sygnałów za pomocą szeregu Fouriera ma ograniczone zastosowanie w analizie lokalnych cech sygnału, które są często kluczowe w przekazie informacji. Rozwiązaniem może być krótkoczasowa (okienkowa) transformacja Fouriera sygnałów (*short-time Fourier transform*), wykorzystująca bazę postaci

$$x_{i,k} = g(t - k\tau)e^{ji\omega_0(t-k\tau)} \tag{2.4}$$

ze stałą funkcją okna $g(\cdot)$, zwykle o charakterze gaussowskim, przesuwana po osi czasu z krokiem τ. Parametr k nowej dziedziny pozwala zachować informację o położeniu przy estymacji lokalnego widma częstotliwościowego reprezentowanego sygnału.

Taka reprezentacja pozbawiona jest jednak bardzo istotnej cechy skalowalności ze względu na stałą postać funkcji okna w całej dziedzinie transformacji. Alternatywnym rozwiązaniem jest stosowanie skalogramów, uzyskanych za pomocą falkowych baz funkcji rozpinających przestrzenie kolejnych przybliżeń z zachowaniem lokalności opisu sygnału. Generalnie, baza ustalonej przestrzeni opisu sygnału narzuca sposób reprezentowania informacji w dziedzinie stosowanego przekształcenia oraz warunki wyznaczenia postaci przybliżenia.

2.3 Opis informacji

Prace C. Shannona sprzed ponad 50 lat określiły matematyczne podstawy statystycznej teorii informacji formalizując m.in. pojęcia statystycznego źródła informacji z modelem procesu losowego o ciągłym zbiorze wartości, entropii jako miary informacji, zniekształceń źródeł informacji.

Zaproponowany przez Shannona opis informacji znalazł powszechne zastosowanie w różnych dziedzinach nauki, m.in. w biologii, medycynie, filozofii. Występuje w nim stochastyczne rozumienie informacji jako "poziomu niepewności" odbiorcy związanej z analizą dostępnych danych. Trudno podważyć ogromne znaczeniem teorii Shannona w rozwoju współczesnej nauki. Jednak warto zwrócić uwagę na ograniczenia tej teorii, szczególnie w zakresie uproszczonych, nierealistycznych założeń statystycznych oraz poprzez pominięcie semantyki w definiowaniu źródeł informacji.

W statystycznej teorii informacji dominuje składniowy (syntaktyczny) aspekt informacji. Nie prowadzi się rozważań dotyczących prawdziwości czy użyteczności danych. Informacja rozumiana jest wtedy jako rozważany ciąg symboli źródła informacji nad ustalonym alfabetem z ogólnie przyjętą wartością semantyczną (znaczeniem).

2.3.1 Teoria informacji według Shannona

Teoria informacji zajmuje się kodowaniem źródłowym i kanałowym bazując na statystycznych właściwościach źródeł informacji, których modele nie uwzględniają aspektu znaczeniowego ze względu na pominięcie aspektu użyteczności. Można przyjąć, że informacja rozumiana jest wtedy jako wiadomość W, tj. ciąg symboli nad ustalonym alfabetem, z przypisaną wartością semantyczną $\Sigma_{\cdot}(W)$ u nadawcy N zgodną z wartością semantyczną u odbiorcy O:

$$\Sigma_N(W) \simeq \Sigma_O(W) \tag{2.5}$$

Przy takiej koncepcji informacja to para $(W, \Sigma_N(W))$, przy możliwej do pominięcia, bo zgodnej, funkcji semantycznej, przy przekazie niezależnym od wymagań odbiorcy.

Ze względu na postać, w jakiej wyrażona jest informacja, można wyróżnić informację ze zbiorem ziarnistym (zbiór o skończonej liczbie elementów) oraz informację ze zbiorem ciągłym (obok jednej informacji dowolnie blisko można znaleźć inne informacje). W kodowaniu danych cyfrowych użyteczne jest pojęcie informacji ze zbiorem ziarnistym. Można dla niej określić tzw. ciągi informacji, które są ciągiem symboli ze zbioru informacji elementarnych (alfabetu), pojawiających się w określonej kolejności, stanowiącej istotę informacji. Przykładowo, źródło opisane alfabetem $A_S = \{\text{a,b,c}\}$ generuje ciąg informacji: $\mathbf{s}(A_S) = (\text{a,a,c,a,b,b,b,c,c,a,c,b,}\ldots)$, tj. sekwencję symboli nad alfabetem A.

W teorii informacji istnieją dwa zasadnicze cele wykorzystania modeli źródeł informacji:

- wyznaczanie fundamentalnych, teoretycznych wartości granicznych wydajności określonej klasy kodów w odniesieniu do ustalonych modeli źródeł informacji,
- opracowanie skutecznych kodów źródeł informacji wiernie przybliżających rzeczywiste strumienie informacji.

Modelowanie źródeł informacji

Źródło informacji jest matematycznym modelem bytu fizycznego, który w sposób losowy generuje (dostarcza, emituje) sukcesywnie symbole. Przyjmując stochastyczny model źródła informacji dobrze opisujący niepewność zakładamy, iż informacja jest realizacją pewnej zmiennej losowej (procesu losowego czy dokładniej łańcucha [7]) o określonych właściwościach statystycznych. Model informacji, w którym zakładamy, że ma on charakter realizacji zmiennej, łańcucha lub procesu stochastycznego o znanych właściwościach statystycznych (istnieją i są znane rozkłady prawdopodobieństwa informacji), nazywany jest modelem z pełną informacją statystyczną lub krócej - modelem probabilistycznym.

Ze względów praktycznych szczególnie interesujące są probabilistyczne modele źródeł informacji, a analiza została ograniczona do dyskretnych źródeł informacji. Dla takich modeli obowiązują twierdzenia graniczne, w tym prawa wielkich liczb, z których wynika, że przy dostatecznie dużej liczbie niezależnych obserwacji częstości występowania określonych postaci informacji będą zbliżone do prawdopodobieństw ich występowania. Częstościowe określanie prawdopodobieństw jest tym dokładniejsze, im liczniejsze zbiory są podstawą wyznaczenia prawdopodobieństw.

W kontekście implementacji metod kodowania wygodniej jest mówić o *wagach symboli*. Jest to liczba wystąpień danego symbolu wogóle lub też w określonym kontekście. Podzielona przez liczbę wystąpień wszystkich symboli jest miarą częstości występowania symbolu, przybliżeniem (estymacją) prawdopodobieństwa.

Proces generacji informacji modelowany za pomocą źródła informacji S polega na dostarczaniu przez źródło sekwencji (ciągu) symboli $\mathbf{s} = (s_1, s_2, \ldots)$ wybranych ze skończonego alfabetu A_S (czyli $s_i \in A_S$) według pewnych reguł opisanych zmienną losową o wartościach s. Bardziej ogólnie probabilistycznym modelem ciągu informacji elementarnych jest sekwencja zmiennych losowych $(S_1, S_2, \ldots)$ traktowana jako proces stochastyczny (dokładniej łańcuch). Właściwości źródła określone są wtedy przez parametry procesu stochastycznego (prawdopodobieństwa łączne, charakterystyka stacjonarności itd.). Stacjonarność źródła w naszych rozważaniach jest rozumiana w kontekście procesu, którego realizacją jest dana

[7] Łańcuchem stochastycznym nazwiemy proces stochastycznym z argumentem ziarnistym

informacja. Rozważmy przestrzeń symboli (dyskretnych próbek) generowanych ze źródła jako zbiór wszystkich możliwych sekwencji symboli wraz z prawdopodobieństwami zdarzeń rozumianych jako występowanie rozmaitych zestawów tych sekwencji. Zdefiniujmy także przesunięcie jako transformację T określoną na tej przestrzeni sekwencji źródła, która przekształca daną sekwencję na nową poprzez jej przesunięcie o pojedynczą jednostkę czasu w lewo (jest to modelowanie wpływu czasu na daną sekwencję), czyli $T(s_1, s_2, s_3, \ldots) = (s_2, s_3, s_4, \ldots)$. Jeśli prawdopodobieństwo dowolnego zdarzenia (zestawu sekwencji) nie ulegnie zmianie poprzez przesunięcie tego zdarzenia, czyli przesunięcie wszystkich sekwencji składających się na to zdarzenie, wtedy transformacja przesunięcia jest zwana niezmienniczą (inwariantną), a proces losowy jest stacjonarny. Teoria stacjonarnych procesów losowych może być więc traktowana jako podzbiór teorii procesów ergodycznych, odnoszącej się do śledzenia zachowania średniej po czasie oraz po próbkach procesów w całej definiowanej przestrzeni [19].

Warto przypomnieć, że w źródle będącym realizacją procesu ergodycznego każda generowana sekwencja symboli ma te same właściwości statystyczne [21]). Momenty statystyczne procesu, rozkłady prawdopodobieństw itp. wyznaczone z poszczególnych sekwencji zbiegają do określonych postaci granicznych przy zwiększaniu długości sekwencji, niezależnie od wyboru sekwencji. W rzeczywistości nie jest to prawdziwe dla każdej sekwencji procesu ergodycznego, ale zbiór przypadków, dla których jest to fałszywe, występuje z prawdopodobieństwem równym 0. Stąd dla stacjonarnych procesów ergodycznych możliwe jest wyznaczenie parametrów statystycznych (średniej, wariancji, funkcji autokorelacji itp.) na podstawie zarejestrowanej sekwencji danych (symboli, próbek) $\{s_i\}_{i=1,2,\ldots}$, co jest wykorzystywane w praktycznych algorytmach kodowania, opartych na przedstawionych poniżej uproszczonych modelach źródeł informacji.

Model bez pamięci - DMS Najprostszą postacią źródła informacji S jest dyskretne źródło bez pamięci DMS (*discrete memoryless source*), w którym sukcesywnie emitowane przez źródło symbole są statystycznie niezależne. Źródło DMS jest całkowicie zdefiniowane przez zbiór wszystkich możliwych wartości s zmiennej losowej, tj. zbiór symboli tworzących alfabet $A_S = \{a_1, a_2, \ldots, a_n\}$, oraz zbiór wartości prawdopodobieństw występowania poszczególnych symboli alfabetu: $P_S = \{p_1, p_2, \ldots, p_n\}$, gdzie $\Pr(s = a_i) = P(a_i) = p_i$, $p_i \geqslant 0$ i $\sum_{s \in A_S} P(s) = 1$.

Można sobie wyobrazić, że źródło o alfabecie A_S zamiast pojedynczych symboli generuje bloki N symboli z alfabetu źródła, czyli struktura pojedynczej informacji jest ciągiem N dowolnych symboli źródła. W takim przypadku można zdefiniować nowe źródło S^N o n^N elementowym alfabecie, zawierającym wszystkie możliwe N - elementowe ciągi symboli. Rozszerzony alfabet takiego źródła jest następujący: $A_S^N = \underbrace{A_S \times A_S \times \cdots \times A_S}_{N}$, czyli $A_S^N = \{(a_{i_1}, a_{i_2}, \ldots, a_{i_N}) : \forall_{j \in \{1,\ldots,N\}} a_{i_j} \in A_S\} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_i, \ldots \mathbf{a}_{n^N})$. Prawdopodobieństwo i-tego ele-

mentu alfabetu wynosi $P(\mathbf{a}_i) = P(a_{i_1})P(a_{i_2}) \cdots P(a_{i_N})$ (mamy do czynienia ze źródłem bez pamięci). Źródło S^N jest nazywane rozszerzeniem stopnia N źródła S.

Model z pamięcią (warunkowy) - CSM Jakkolwiek model DMS spełnia założenia prostej, wygodnej w analizie struktury, to jednak w wielu zastosowaniach jest nieprzydatny ze względu na małą zgodność z charakterem opisywanej informacji. Założenie o statystycznej niezależności kolejnych zdarzeń emisji symbolu jest bardzo rzadko spełnione w praktyce. Aby lepiej wyrazić rzeczywistą informację zawartą w zbiorze danych, konstruuje się tzw. model warunkowy źródła - CSM (*conditional source model*), zwany także modelem z pamięcią w odróżnieniu od modelu DMS. Jest to ogólna postać modelu źródła informacji, którego szczególnym przypadkiem jest DMS, a także często wykorzystywany model źródła Markowa.

Modele źródeł z pamięcią, najczęściej ograniczoną, pozwalają z większą dokładnością przewidzieć pojawienie się poszczególnych symboli alfabetu źródła (strumień danych staje się lepiej określony przez model źródła). Koncepcja pamięci źródła jest realizowana poprzez określenie kontekstu (czasowego), który ma wpływ na prawdopodobieństwo wyemitowania przez źródło konkretnych symboli w danej chwili. W każdej kolejnej chwili czasowej t, po wcześniejszym odebraniu ze źródła sekwencji symboli $\mathbf{s}^t = (s_1, s_2, \ldots, s_t)$, można na podstawie $\mathbf{s}^t$ (tj. przeszłości) wnioskować o postaci kolejnego oczekiwanego symbolu poprzez określenie rozkładu prawdopodobieństw warunkowych $P(\cdot|\mathbf{s}^t)$.

Zbiór wszystkich dostępnych z przeszłości danych $\mathbf{s}^t$ stanowi pełny kontekst wystąpienia symbolu s_{t+1}. Kontekst C wykorzystywany w obliczanym rozkładzie $P(\cdot|C)$ do modelowania lokalnych zależności danych dla różnych źródeł informacji stanowi zwykle skończony podzbiór $\mathbf{s}^t$. Może być także wynikiem redukcji alfabetu źródła, przekształceń wykonanych na symbolach $\mathbf{s}^t$ itp. Reguły określenia C mogą być stałe w całym procesie generacji symboli przez źródło lub też mogą ulegać adaptacyjnym zmianom (np. w zależności od postaci ciągu symboli wcześniej wyemitowanych przez źródło).

Model źródła S z pamięcią jest określony poprzez:

- alfabet symboli źródła $A_S = \{a_1, a_2, \ldots, a_n\}$,
- zbiór kontekstów C dla źródła S postaci $A_C^S = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$,
- prawdopodobieństwa warunkowe $P(a_i|\mathbf{b}_j)$ dla $i = 1, 2, ..., n$ oraz $j = 1, 2, ..., k$, często wyznaczane metodą częstościową z zależności

$$P(a_i|\mathbf{b}_j) = \frac{N(a_i, \mathbf{b}_j)}{N(\mathbf{b}_j)} \tag{2.6}$$

  gdzie $N(a_i, \mathbf{b}_j)$ - liczba łącznych wystąpień symbolu a_i i kontekstu $\mathbf{b}_j$, $N(\mathbf{b}_j)$ - liczba wystąpień kontekstu $\mathbf{b}_j$, przy czym jeśli $N(\mathbf{b}_j) = 0$ dla pew-

nych j (taki kontekst wystąpienia symbolu jeszcze się nie pojawił), wtedy można przyjąć każdy dowolny rozkład przy tym kontekście (wykorzystując np. wiedzę *a priori* do modelowania źródła w takich przypadkach),

- zasadę określania kontekstu C w każdej 'chwili czasowej' t jako funkcję $f(\cdot)$ symboli wcześniej wyemitowanych przez źródło.

Załóżmy, że źródło emituje sekwencję danych wejściowych $\mathbf{s}^t = (s_1, s_2, \ldots, s_l, \ldots, s_t)$, gdzie $s_l \in A_S$. Sekwencja kontekstów wystąpienia tych symboli jest określona przez funkcję $f(\cdot)$ oraz A_C^S i przyjmuje postać: $\mathbf{c}^t = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_l, \ldots, \mathbf{c}_t)$, gdzie $\mathbf{c}_l = f(s_1, s_2, \ldots, s_{l-1}) \in A_C^S$ dla $l = 2, \ldots, t$ (w przypadku symbolu s_1 brak jest symboli wcześniej wyemitowanych przez źródło, można więc przyjąć dowolny kontekst $\mathbf{c}_1$). W ogólności $f(\cdot)$ wyznaczająca kontekst następnego symbolu s_{t+1} musi być określona jako przekształcenie wszystkich możliwych sekwencji symboli z A_S o długości t lub mniejszej w A_C^S. Ponieważ $\mathbf{s}_t$ jest jedyną dostępną sekwencją symboli wygenerowaną przez źródło S prawdopodobieństwa warunkowe określane są na podstawie $\mathbf{s}_t$ według zależności (2.6).

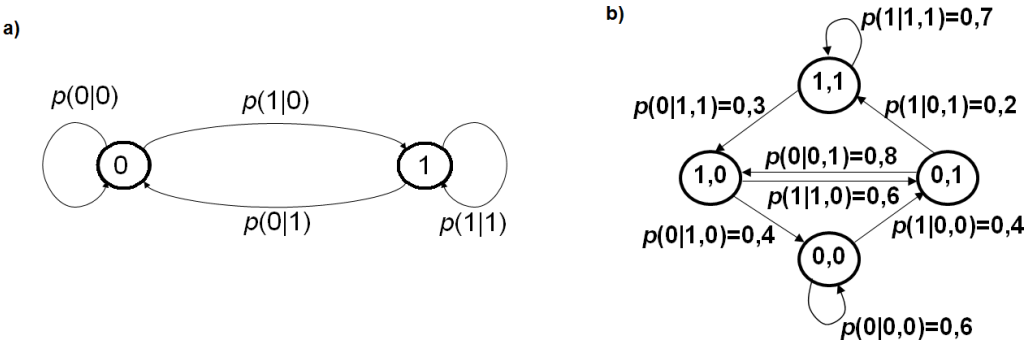
Istotnym parametrem modelu CSM jest rząd kontekstu C, który określa liczbę symboli tworzących kontekst $\mathbf{c}_l$ dla kolejnych symboli emitowanych przez źródło. Rozważmy prosty przykład sekwencji $\mathbf{s}^t$ ze źródła S modelowanego rozkładem $P(a_i|\mathbf{b}_j)$ przy kontekstach kolejnych symboli $\mathbf{c}^l$ rzędu 1. Niech kontekst C stanowi symbol bezpośrednio poprzedzający kodowaną wartość s_l: $\mathbf{c}_l = f(s_1, s_2, \ldots, s_{l-1}) = s_{l-1}$ dla $i = 2, \ldots, t$ oraz $\mathbf{c}_1 = a_r \in A_S$ dla pewnego $1 \leqslant r \leqslant n$. Wtedy $A_C^S = A_S$, a ten model CSM jest modelem źródła Markowa pierwszego rzędu. Generalnie, model źródła Markowa rzędu m jest bardzo powszechnie stosowaną realizacją CSM (model DMS jest modelem źródła Markowa rzędu 0).

Model źródła Markowa Źródło Markowa rzędu m jest źródłem, w którym kontekst $C^{(m)}$ wystąpienia kolejnych symboli s_l generowanych przez źródło S stanowi skończona liczba m poprzednich symboli $\mathbf{c}_l^{(m)} = (s_{l-1}, s_{l-2}, \ldots, s_{l-m})$, czyli dla dowolnych wartości l oraz $t \geqslant m$ $P(s_l|s_{l-1}, s_{l-2}, \ldots, s_{l-t}) = P(s_l|\mathbf{c}_l^{(m)})$. Zatem prawdopodobieństwo wystąpienia symbolu a_i z alfabetu źródła zależy jedynie od m symboli, jakie pojawiły się bezpośrednio przed nim, przy czym określone jest przez zbiór prawdopodobieństw warunkowych (oznaczenia jak przy definicji źródła CSM):

$$P(a_i|\mathbf{b}_j) = P(a_i|a_{j_1}, a_{j_2}, \ldots, a_{j_m}) \tag{2.7}$$

dla wszystkich i oraz $j_1, j_2, \ldots, j_m = 1, 2, \ldots, n$.

Często źródło Markowa analizowane jest za pomocą diagramu stanów, jako znajdujące się w pewnym stanie, zależnym od skończonej liczby występujących poprzednio symboli - zobacz rys. 2.3. Dla źródła Markowa pierwszego rzędu jest n takich stanów, dla źródła rzędu m mamy n^m stanów.



Rysunek 2.3: Diagramy stanów prostych modeli Markowa z alfabetem binarnym: a) ogólny model rzędu 1 z przejściami pomiędzy poszczególnymi stanami, b) model rzędu 2 z przykładowymi wartościami prawdopodobieństw warunkowych. Stany opisane są wszystkimi możliwymi kombinacjami kontekstów, zaś odpowiednie przejścia pomiędzy stanami źródła odzwierciedlają wystąpienie kolejnej danej źródłowej; w kontekście prawy symbol oznacza ten bezpośrednio poprzedzający, zaś lewy - to symbol jeszcze wcześniejszy.

Przy kompresji danych obrazowych efektywny kontekst budowany jest zwykle z najbliższych w przestrzeni obrazu pikseli, przy czym sposób określenia kontekstu może zmieniać się dynamicznie w trakcie procesu kodowania, np. zależnie od lokalnej statystyki. Popularną techniką jest taka kwantyzacja kontekstu (tj. zmniejszanie kontekstu w celu uzyskania bardziej wiarygodnego modelu prawdopodobieństw warunkowych), kiedy to liniowa kombinacja pewnej liczby sąsiednich symboli warunkuje wystąpienie symbolu, dając *de facto* model warunkowy pierwszego rzędu, zależny jednak od kilku- czy kilkunastoelementowego sąsiedztwa.

Miary ilości informacji

Miara ilości informacji dostarczanej (emitowanej) przez probabilistyczne źródło informacji konstruowana jest przy dwóch intuicyjnych założeniach: a) więcej informacji zapewnia pojawienie się mniej prawdopodobnego symbolu, b) informacja związana z wystąpieniem kilku niezależnych zdarzeń jest równa sumie informacji zawartej w każdym ze zdarzeń.

Informacja $I(a_i)$ związana z wystąpieniem pojedynczego symbolu a_i alfabetu źródła S określona jest w zależności od prawdopodobieństwa wystąpienia tego symbolu $p_i = \Pr(s = a_i)$ jako $I(a_i) = \lg(1/p_i)$, $p_i \neq 0$. Jest to tzw. informacja własna (*self-information*).

W przypadku strumienia danych generowanych przez źródło do określenia ilości informacji wykorzystuje się pojęcie entropii. Zasadniczo, dla sekwencji kolejnych symboli s_i, gdzie $i = 1, 2, \ldots$, dostarczanych ze źródła informacji S o

alfabecie $A_S = \{a_1, a_2, \ldots, a_n\}$ entropia określona jest jako

$$H(S) = \lim_{m\to\infty} \frac{1}{m} I_m \tag{2.8}$$

gdzie

$$I_m = -\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_m=1}^{n} P(a_{j_1}, a_{j_2}, \ldots, a_{j_m}) \lg P(a_{j_1}, a_{j_2}, \ldots, a_{j_m}) =$$
$$= -\sum_{j_1,\ldots,j_m=1}^{n} \Pr(s_1 = a_{j_1}, \ldots, s_m = a_{j_m}) \lg \Pr(s_1 = a_{j_1}, \ldots, s_m = a_{j_m})$$

oraz $(s_1, s_2, \ldots, s_m)$ jest sekwencją symboli źródła S o długości m.

Tak określona entropia nosi nazwę entropii łącznej, gdyż jest wyznaczana za pomocą prawdopodobieństwa łącznego wystąpienia kolejnych symboli z alfabetu źródła informacji. Definicja entropii według zależności (2.8) jest jednak niepraktyczna, gdyż nie sposób wiarygodnie określić prawdopodobieństwa łącznego wystąpienia każdej, możliwej (określonej przez alfabet) kombinacji symboli źródła w rzeczywistym skończonym zbiorze danych. Wymaga to albo dużej wiedzy *a priori* na temat charakteru zbioru danych, które podlegają kompresji, albo nieskończenie dużej liczby danych do analizy (nieskończenie długiej analizy). Należałoby więc zbudować model źródła informacji określający prawdopodobieństwo łącznego wystąpienia dowolnie długiej i każdej możliwej sekwencji symboli tegoż źródła. Bardziej praktyczne postacie zależności na entropię, aproksymujące wartość entropii łącznej dla danego źródła informacji, wynikają z uproszczonych modeli źródeł.

Entropia modelu źródła może być rozumiana jako średnia ilość informacji przypadająca na generowany symbol źródła, jaką należy koniecznie dostarczyć, aby usunąć wszelką nieokreśloność (niepewność) z sekwencji tych symboli. Podstawa logarytmu używanego w definicjach miar określa jednostki używane do wyrażenia ilości informacji. Jeśli ustala się podstawę równą 2, wtedy entropia według (2.8) wyraża w bitach na symbol średnią ilość informacji zawartą w zbiorze danych (tak przyjęto w rozważaniach o entropii).

Dla poszczególnych modeli źródeł informacji można określić ilość informacji generowanej przez te źródła. Ponieważ modele źródeł tylko naśladują (aproksymują) cechy źródeł rzeczywistych (często niedoskonale), obliczanie entropii dla rzeczywistych zbiorów danych za pomocą tych modeli jest często zbyt dużym uproszczeniem. Należy jednak podkreślić, iż obliczona dla konkretnego źródła ilość informacji tym lepiej będzie przybliżać rzeczywistą informację zawartą w zbiorze danych (wyznaczaną asymptotycznie miarą entropii łącznej), im wierniejszy model źródła informacji został skonstruowany.

Entropia modelu źródła bez pamięci Zakładając, że kolejne symbole są emitowane przez DMS niezależnie, wyrażenie na entropię tego modelu źródła można

wyprowadzić z równania (2.8). Entropia modelu źródła bez pamięci, uzyskana przez uśrednienie ilości informacji własnej po wszystkich symbolach alfabetu źródła wynosi :

$$H(S_{\mathrm{DMS}}) = -\sum_{i=1}^{n} P(a_i) \log_2 P(a_i) \tag{2.9}$$

gdzie n oznacza liczbę symboli a_i w alfabecie. Dla $P(a_i) = 0$ wartość $0 \cdot \log_2 1/0 \equiv 0$, gdyż $\lim_{\phi \to 0^+} \phi \log_2 1/\phi = 0$. Entropia źródła bez pamięci nazywana jest entropią bezwarunkową (od użytej formy prawdopodobieństwa). W przypadku, gdy źródło DMS nie najlepiej opisuje kodowany zbiór danych entropia obliczona według (2.9) jest wyraźnie większa od entropii łącznej, czyli nie jest w tym przypadku najlepszą miarą informacji. Rzeczywista informacja zawarta w zbiorze danych jest pomniejszona o nieuwzględnioną informację wzajemną, zawartą w kontekście wystąpienia kolejnych symboli.

Entropia modelu źródła z pamięcią

Zależności pomiędzy danymi w strumieniu zwykle lepiej opisuje model z pamięcią, a wartość entropii tego źródła (tzw. entropii warunkowej) jest bliższa rzeczywistej ilości informacji zawartej w kompresowanym zbiorze danych. Zależność pomiędzy entropią łączną $H(C, S)$ źródła o zdefiniowanym kontekście C, warunkową $H(S|C)$ oraz tzw. entropią brzegową (entropią źródła obliczoną dla rozkładu brzegowego) $H(C)$ jest następująca:

$$H(C, S) = H(S|C) + H(C) \tag{2.10}$$

Przykładem miary ilości informacji źródeł z pamięcią będzie entropia wyznaczona dla źródeł Markowa, znajdujących bardzo częste zastosowanie w praktyce kompresji.

Entropia modelu źródła Markowa Aby za pomocą modelu źródła Markowa rzędu m obliczyć ilość informacji (średnio na symbol źródła) zawartą w kodowanym zbiorze danych, wykorzystuje się zbiór prawdopodobieństw warunkowych i określa tzw. entropię warunkową źródła znajdującego się w pewnym stanie $(a_{j_1}, a_{j_2}, \ldots, a_{j_m})$ jako:

$$H(S|a_{j_1}, a_{j_2}, \ldots, a_{j_m}) = -\sum_{i=1}^{n} P(a_i|a_{j_1}, \ldots, a_{j_m}) \log_2 P(a_i|a_{j_1}, \ldots, a_{j_m}) \tag{2.11}$$

Następnie oblicza się średnią entropię warunkową źródła S jako sumę ważoną entropii warunkowych po kolejnych stanach źródła wynikających ze wszystkich możliwych konfiguracji (stanów) kontekstu $C^{(m)}$: $A^S_{C^{(m)}} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_j, \ldots, \mathbf{b}_k\}$,

gdzie $\mathbf{b}_j = (a_{j_1}, \ldots, a_{j_m})$ oraz $\forall_{l\in\{1,\ldots,m\}} a_{j_l} \in A_S$, przy czym wagami są prawdopodobieństwa przebywania źródła w danym stanie:

$$H(S|C^{(m)}) = \sum_{A^S_{C^{(m)}}} P(a_{j_1}, \ldots, a_{j_m}) H(S|a_{j_1}, \ldots, a_{j_m}) \tag{2.12}$$

czyli

$$H(S|C^{(m)}) = -\sum_{j_1=1}^{n} \cdots \sum_{j_m=1}^{n} \sum_{i=1}^{n} P(a_{j_1}, \ldots, a_{j_m}, a_i) \log_2 P(a_i|a_{j_1}, \ldots, a_{j_m})$$

Tak określona średnia entropia warunkowa modelu źródła Markowa rzędu m jest mniejsza lub równa entropii bezwarunkowej. Jest ona pomniejszona o średnią ilość informacji zawartą w kontekście wystąpienia każdego symbolu strumienia danych. Jednocześnie entropia warunkowa danych przybliżanych źródłem Markowa rzędu m jest niemniejsza niż entropia łączna źródła emitującego tę sekwencję danych (według równania (2.8)). Zależność pomiędzy postaciami entropii związanymi z przedstawionymi modelami źródeł informacji, opisującymi z większym lub mniejszym przybliżeniem informację zawartą w konkretnym zbiorze danych, jest następująca:

$$H(S) \leqslant H(S|C^{(m)}) \leqslant H(S_{\mathrm{DMS}}) \tag{2.13}$$

Zastosowanie modeli CSM wyższych rzędów zazwyczaj lepiej określa rzeczywistą informację zawartą w zbiorze danych, co pozwala zwiększyć potencjalną efektywność algorytmów kompresji wykorzystujących te modele. W zależności od charakteru kompresowanych danych właściwy dobór kontekstu może wtedy zmniejszyć graniczną długość reprezentacji kodowej. Stosowanie rozbudowanych modeli CSM w konkretnych implementacjach napotyka jednak na szereg trudności, wynikających przede wszystkim z faktu, iż ze wzrostem rzędu kontekstu liczba współczynników opisujących model rośnie wykładniczo. Wiarygodne statystycznie określenie modeli zaczyna być wtedy problemem. Trudniej jest także zrealizować algorytmy adaptacyjne, co w efekcie może zmniejszyć skuteczność kompresji w stosunku do rozwiązań wykorzystujących prostsze modele.

2.3.2 Kodowanie, czyli usuwanie nadmiarowości

W zagadnieniu kompresji jeszcze silniej wykorzystuje się model informacji, tak probabilistyczne, jak też różne formy modeli uwzględniających w jakimś stopniu funkcje semantyczne danych źródłowych (znaczenie pojedynczych symboli, ciągów danych, relacji pomiędzy danymi, itd.). Zasadnicze jest też odniesienie do reprezentacji rozumianej jako ciąg bitów o możliwie zredukowanej długości - kompresją nazywamy wyznaczanie możliwie oszczędnej reprezentacji sekwencji danych.

Poniżej przedstawiono bardziej ścisłe próby zdefiniowania pojęcia kompresji - pierwszą w ujęciu bardziej intuicyjnym, drugą - w rozumieniu najnowszych trendów rozumienia i wykorzystywania procesów kompresji danych.

Definicja 2.7 *Kompresja danych w sensie podstawowym*
Kompresja to odwracalny lub nieodwracalny proces redukcji długości reprezentacji danych. Odwrotny proces odtwarzania źródłowej reprezentacji danych lub jej przybliżenia na podstawie reprezentacji skompresowanej (nazywanej reprezentacją kodową) nazywany jest dekompresją.
□

Cele kompresji w zależności od charakteru danych i rodzaju zastosowań mogą być jednak bardziej różnorodne.

Definicja 2.8 *Kompresja danych w sensie rozszerzonym*
Kompresja to wyznaczanie możliwie użytecznej w określonym zastosowaniu reprezentacji danych, czyli reprezentacji informacji, przy dążeniu do redukcji wszelkiej nadmiarowości na poziomie pojedynczych bitów.
□

Można wyróżnić dwie zasadnicze kategorie metod kompresji danych: bezstratne i stratne. W **kompresji bezstratnej** (inaczej odwracalnej, bezstratnej numerycznie) zrekonstruowany po kompresji ciąg danych jest numerycznie identyczny z sekwencją źródłową z dokładnością do pojedynczego bitu. Taki rodzaj kompresji jest wykorzystywany w zastosowaniach wymagających wiernej rekonstrukcji danych oryginalnych, takich jak archiwizacja dokumentów tekstowych, historii operacji finansowych na kontach bankowych, niektórych obrazów medycznych i wielu innych.

Kompresja z selekcją informacji[8] nie pozwala odtworzyć (zrekonstruować) z dokładnością do pojedynczego bitu danych źródłowych. W przypadku tzw. stratnej kompresji obrazów wprowadza się czasami pojęcie wizualnej bezstratności, a w przypadku dźwięku – bezstratności słuchowej (ogólnie chodzi o bezstratność percepcji danych). Uproszczenie strumienia danych prowadzące do efektywniejszej, przede wszystkim krótszej reprezentacji może być niezauważalne dla obserwatora w normalnych warunkach prezentacji. Przykładowo, przy prezentacji obrazów medycznych o dwunastobitowej dynamice za pomocą stacji roboczej z ośmiobitowym przetwornikiem karty graficznej usunięcie (zniekształcenie) treści zapisanej w czterech najmłodszych bitach oryginalnych wartości pikseli nie spowoduje żadnych zmian w obserwowanym obrazie. Definicja percepcyjnej bezstratności jest jednak względna, zależna od zdolności, umiejętności i zamierzeń odbiorcy, co nakazuje zachowanie ostrożności w konkretnych zastosowaniach.

[8] Inaczej kompresja stratna lub nieodwracalna.

Przykładowo, wystarczy zmiana warunków obserwacji obrazu, np. zmiana okna obserwacji kolejnych map bitowych, użycie danych przetworzonych do rejestracji obrazu na filmie, bądź też poddanie ich dalszemu przetwarzaniu (eliminacja szumów, podkreślenia krawędzi, segmentacja itp.), by wystąpiła zauważalna różnica pomiędzy analizowanym obrazem źródłowym i obrazem ze wstępnie wyzerowanymi najmłodszymi bitami.

Zgodnie z klasycznym paradygmatem kompresji stratnej dane wejściowe transformuje się w nową przestrzeń pośrednią, w której redukowana jest nadmiarowość reprezentacji źródłowej. Wykorzystuje się przy tym ograniczenie zbioru możliwych wartości pośrednich poprzez kwantyzację, co jako proces nieodwracalny powoduje stratność całej metody. Drugim ważnym etapem jest kodowanie reprezentacji pośredniej. Odtworzona sekwencja danych jest jedynie przybliżeniem sekwencji źródłowej zachowującym w założeniu istotne jej właściwości.

Uproszczenie charakteru danych (związane z redukcją informacji rozumianej syntaktycznie) w procesie kwantyzacji, przeprowadzanej w dziedzinie efektywnej transformaty, pozwala znacznie zwiększyć stopień kompresji w stosunku do metod bezstratnych. Wymaga to jednak rzetelnej kontroli jakości danych rekonstruowanych za pomocą wiarygodnych miar zachowanej ilości informacji. Kontrola ta pozwoli ustalić wartości dopuszczalnych stopni kompresji w określonych zastosowaniach.

Zasadniczym celem selekcji informacji w kompresowanym zbiorze danych jest usunięcie wszystkiego, co nie jest informacją dla odbiorcy, aby uprościć reprezentację danych i zredukować jej długość. Przykładowo, w obrazie może zostać wydzielony obszar zainteresowania (ROI), którego rekonstrukcja ze skompresowanej reprezentacji pozwoli odtworzyć oryginał z dokładnością do pojedynczego bitu, podczas gdy pozostała część obrazu może zostać maksymalnie uproszczona, bez zachowania nawet elementarnej treści. W innym przypadku selekcja może prowadzić do wiernego zachowania jedynie tych właściwości odtwarzanych danych, które są istotne dla odbiorcy. Opis wybranych metod kompresji przedstawiono w rozdziale 3.

Efektywność kompresji może być rozumiana zależnie od zastosowania, rodzaju danych kodowanych, sprzętowych możliwości implementacji, parametrów środowiska transmisji—gromadzenia informacji, wymagań użytkownika czy sposobu rozpowszechniania informacji itp. Najbardziej powszechnym rozumieniem tego pojęcia jest efekt minimalizacji rozmiaru reprezentacji skompresowanej danych oryginalnych. Do liczbowych miar tak rozumianej efektywności należą przede wszystkim: stopień kompresji CR (*compression ratio*), procent kompresji CP (*compression percentage*) oraz średnia bitowa BR (*bit rate*).

Stopień kompresji wyrażany jest stosunkiem liczby bitów reprezentacji oryginalnej do liczby bitów reprezentacji skompresowanej wyrażanej w postaci n:1, np. 2:1, 100:1. Procent kompresji (stosowany często w ocenie skuteczności archiwize-

rów tekstu) określany jest wyrażeniem $CP = (1 - \frac{1}{CR}) \cdot 100\%$, a średnia bitowa to średnia ilość bitów reprezentacji skompresowanej przypadająca na element źródłowej sekwencji danych. Efektywność (skuteczność, wydajność) kompresji oznacza wtedy uzyskanie możliwie dużych wartości CR i CP, czy też możliwie małej średniej bitowej BR. Miary CR i CP są miarami względnymi, przydatnymi np. w ocenie efektywności koderów w zastosowaniach archiwizacji (ich wartość łatwo przekłada się na poziom oszczędności kosztów nośników). Bezwzględna wartość BR charakteryzuje rozmiar wyjściowych danych kodera i jest użyteczna w zastosowaniach transmisyjnych (łatwo określić przepustowość sieci np. wymaganą przy transmisji w czasie rzeczywistym).

W innych zastosowaniach efektywność może być związana z minimalizacją czasu kompresji (lub dekompresji), np. przy rejestracji danych pomiarowych w czasie rzeczywistym, wielokrotnym odczytywaniu obrazów zgromadzonych w ogólnodostępnej bazie danych. Kryteriami efektywności mogą być także: minimalny iloczyn: czas $\times$ średnia bitowa, wysoka odporność strumienia danych skompresowanych na błędy transmisji, dobra jakość danych po kompresji/dekompresji w zależności od rodzaju wprowadzonych zniekształceń, możliwość elastycznego odtwarzanie danych źródłowych w szerokim zakresie skal, możliwość kodowanie wybranego obszaru (fragmentu) zainteresowań w sposób odmienny od pozostałej części zbioru źródłowego, itp.

Można przyjąć, że kodowanie jest w pierwszym przybliżeniu synonimem kompresji, niekiedy rozumianym w nieco zawężonym znaczeniu. Podstawą kompresji-kodowania są **kody**, tj. ustalone reguły tworzenia użytecznej reprezentacji ze zredukowaną nadmiarowością na poziomie bitów. Na bazie kodów podstawowych opracowywane są efektywne metody kompresji danych o różnej specyfice, czemu służą powszechnie przyjęte, sprawdzone wzorce rozwiązań zweryfikowanych w praktyce, czyli **paradygmaty kompresji**.

Kodowanie danych

Współczesne kodeki bazują na podstawach teorii informacji w zakresie stosowanych metod binarnego kodowania oraz probabilistycznego modelowania źródeł informacji (zobacz punkt ...). Wśród wspomagających zasobów efektywnych rozwiązań warto wymienić teorię aproksymacji, przetwarzania sygnałów, klasyfikacji, percepcji wraz z modelami ludzkiego systemu widzenia czy słyszenia.

Wyróżnić można przede wszystkim metody odwracalne i nieodwarcalne, entropijne lub słownikowe, kody symboli lub strumieniowe, transformacyjnego kodowania z opcją podziału na bloki, skalowaniem, osadzaniem, progresją i hierarchicznością lub bez. Inteligencją tych metod jest modelowanie źródła informacji z pełną adaptacją, przy zadawalajacej wiarygodności modelu probabilistycznego w określonym kontekście lub przy maksymalnym dopasowaniu mechanizmu deterministycznego. Przy selekcji informacji ważne jest porządkowanie, maksy-

malny przyrost ilości kodowanej informacji w początkowej fazie kształtowania reprezentacji kodowej, z zachowaniem monotoniczności tego przyrostu.

Metody proste, sprawdzone, użyteczne niemal w każdym zastosowaniu uzupełniane są przez rozwiązania błyskotliwe, specyficzne, dopasowane do współczesnych wymagań aplikacji multimedialnych oraz rosnących zapotrzebowań w obszarze wymiany informacji. Pomysły sprawdzone, jak przykładowo estymacja i kompensacja ruchu bazująca na blokach i predykcji z ramek sąsiednich, doskonalone są poprzez rozszerzanie obszaru przeszukiwań zlaeżności pomiędzy danymi, zmienną wielkość bloku, elastyczny dobór przekształcenia blokowego itp., wykorzystując rosnącą moc obliczeniową współczesnych procesorów.

Kodowanie odwracalne

Zwykle w procesie kompresji odwracalnej (bezstratnej) występują dwa zasadnicze etapy procesu kodowania, które odnoszą się do całej sekwencji danych lub poszczególnych jej części. W pierwszej fazie **modelowania** tworzony jest opis, charakterystyka źródła informacji, jego podstawowych właściwości. Wierność, wiarygodność i prostota modelu decyduje o efektywności zasadniczego etapu **binarnego kodowania** sekwencji źródłowej. Kodowanie binarne polega na tworzeniu możliwie oszczędnej reprezentacji kodowej w postaci bitowej sekwencji jednoznacznie reprezentującej dane źródłowe.

Modelowanie pełni rolę "inteligencji" sterującej "silnikiem kodowania", czyli koderem binarnym. Utworzenie modelu o w.w. wymaganiach jest niekiedy zbyt trudne ze względu na złożoność źródła informacji i brak stabilności (stacjonarności) jego charakterystyki. Wówczas wykorzystywana jest dodatkowa, wstępna **dekompozycja** danych, czyli proste przekształcenie reprezentacji lub też transformacja do nowej dziedziny. Celem jest stworzenie pośredniej reprezentacji źródła informacji, uproszczonej, o przewidywalnych właściwościach, generalnie o większej podatności na kodowanie. Przykładem takiego przekształcenia może być policzenie różnic pomiędzy kolejnymi danymi ciągu źródłowego lub też zastąpienie serii powtarzających się symboli liczbą ich powtórzeń. Przekształcając dane z przestrzeni oryginalnej w inną przestrzeń reprezentacji pośredniej z wykorzystaniem metrycznych (odległościowych) zależności danych, określonego sposobu porządkowania danych lub zmniejszenia wymiarowości oryginalnej dziedziny danych itp. można uzyskać w niektórych przypadkach znaczące zwiększenie stopnia kompresji.

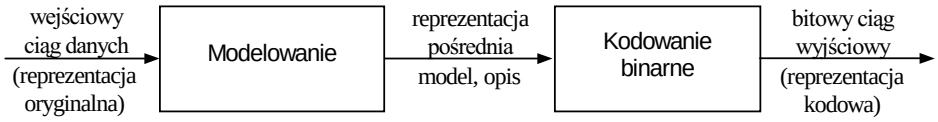
Modelowanie można zrealizować na dwa zasadnicze sposoby:

a) opracowując uogólniony model probabilistyczny (przy założeniu określonej stacjonarności źródła danych) na podstawie przyjętej postaci kontekstu (sąsiedztwa) wystąpienia danych, przy dostępnej statystyce zliczeń – stosowany przede wszystkim w metodach entropijnych;

b) tworząc model deterministyczny opisujący relacje identyczności danych (ciągów danych powtarzających się) w odniesieniu do chwilowych czy lokalnych zależności danych lub też funkcyjne wzory zależności z obliczeniem odstępstw od przyjętego modelu – stosowany przede wszystkim w metodach kodowania długości serii, słownikowych, predykcyjnych;

Możliwe jest również łączenie realizowanych w różnej formie sposobów modelowania kodowanej sekwencji danych w celu uzyskania dokładniejszej charakterystyki źródła, o większej wiarygodości i zwartości opisu (tj. przy możliwie małej liczbie parametrów modelu). Model powinien być dobrze określony, z większym zróżnicowaniem prawdopodobieństw symboli czy też dłuższymi ciągami jednakowych bądź podobnych symboli.

Wstępna dekompozycja, modelowanie oraz kodowanie binarne tworzą odwracalne odwzorowanie (tj. bijekcję) metod odwracalnej kompresji na wiele różnych sposobów. Te rozdzielone etapy kodowania mogą być niekiedy przeplatane, zintegrowane, przenikające się, a w niektórych rozwiązaniach wręcz komplementarne. Ogólny paradygmat kodowania odwracalnego przedstawiono na rys. 2.4, zaś w przykładzie 2.3.2 opisano proste jego realizacje.



Rysunek 2.4: Ogólny paradygmat odwracalnych metod kompresji.

Proste przykłady kodowania

Jedną z najprostszych metod kodowania jest zastąpienie serii identycznych symboli liczbą powtórzeń danego symbolu (zobacz więcej na temat tej metody na stronie 94). Niech sekwencja danych wejściowych będzie następująca:
$\mathbf{s}_{we} = (5, 5, 5, 2, 2, 11, 11, 11, 11, 11, 8)$. Jeśli wyznaczymy model opisujący dane źródłowe seriami jednakowych symboli według schematu: (liczba powtórzeń, symbol), wówczas opis $\mathbf{s}_{we}$ źródła za pomocą takiego modelu jest następujący:
$\mathcal{M}^{\text{serie}}_{\mathbf{s}_{we}} = ((3, 5), (2, 2), (5, 11), (1, 8))$. Ustalmy, że wynikająca z przyjętego modelu postać opisu jest kodowana binarnie w najprostrzy sposób według zasady, bazującej na przyjętych wstępnie ograniczeniach:

- przy założeniu długości serii nie dłuższej niż 8, na pierwszych trzech bitach zapisywana jest dwójkowo liczba (pomniejszona o jeden) powtórzeń symbolu,
- przy założeniu postaci alfabetu $A_S = \{0, \ldots, 15\}$, na kolejnych czterech bitach kodu dwójkowego reprezentowany jest symbol tej serii.

Binarna postać sekwencji kodowej $\mathbf{s}_{wy}$ wygląda wówczas następująco: $\mathbf{s}_{wy} = (\texttt{0100101},\ \texttt{0010010},\ \texttt{1001011}, \texttt{0001000})$. Uzyskano efekt redukcji długości reprezentacji danych z 44 bitów postaci źródłowej (przyjmując 4 bity na pojedynczy symbol) do 28 bitów wyjściowych. Efektywność tej metody rośnie, gdy pojawiają się długie serie powtórzeń symboli.

Inna metoda kodowania wykorzystuje fakt, że niektóre symbole źródłowe pojawiają się częściej, zaś pozostałe rzadziej. Różnicowanie długości słów kodowych poszczególnych symboli polega na przypisaniu krótszych słów symbolom występującym częściej. Model opisuje wówczas częstości wystapień symboli za pomocą wag $w(\cdot)$, równych sumie wystąpień poszczególnych symboli alfabetu. W przypadku rozważanej $\mathbf{s}_{wy}$ mamy więc: $\mathcal{M}^{\text{wagi}}_{\mathbf{s}_{we}} = \{w(5) = 3, w(2) = 2, w(11) = 5, w(8) = 1\}$. Jeśli na podstawie tych wag zróżnicujemy słowa kodowe w sposób następujący: $\varsigma(5) = \texttt{10}, \varsigma(2) = \texttt{110}, \varsigma(11) = \texttt{0}, \varsigma(8) = \texttt{111}$, wówczas otrzymamy ciąg wyjściowy: $\mathbf{s}_{wy} = (\texttt{10}, \texttt{10}, \texttt{10}, \texttt{110}, \texttt{110}, \texttt{0}, \texttt{0}, \texttt{0}, \texttt{0}, \texttt{0}, \texttt{111})$ o długości 20 bitów. Druga forma modelowania okazała się skuteczniejsza, chociaż wymaga jeszcze szereg dookreśleń (w zakresie metody ustalania słów kodowych o zmiennej długości oraz konieczności przekazania w nagłówku pliku skompresowanego parametrów modelu $\mathcal{M}^{\text{wagi}}_{\mathbf{s}_{we}}$.

Przykładem wstępnej dekompozycji danych w celu uproszczenia modelu jest wyskorzystanie prostego mechanizmu liczenia różnic pomiędzy wartością kodowaną i wartością bezpośrednio ją poprzedzającą:
$\mathcal{D}^{\text{różnice}} = \{r_i :\ r_i = s_i - s_{i-1}, i = 1, \ldots, 11,\ s_0 = 0\}$. Dla rozważanej $\mathbf{s}_{we}$ mamy wtedy $\mathcal{D}^{\text{różnice}}_{\mathbf{s}_{we}} = \{5, 0, 0, -3, 0, 9, 0, 0, 0, 0, -3\}$. Dalej stosując modelowanie z wagami ustalamy: $\mathcal{M}^{\text{wagi różnic}}_{\mathbf{s}_{we}} = \{w(5) = 1, w(0) = 7, w(-3) = 2, w(9) = 1\}$. Przypisując tym symbolom zróżnicowane długością słowa kodowe: $\varsigma(5) = \texttt{110}$, $\varsigma(0) = \texttt{0}$, $\varsigma(-3) = \texttt{10}$, $\varsigma(9) = \texttt{111}$, uzyskamy 18 bitów postaci zakodowanej.

Modelowanie to użycie skutecznych modeli statystycznych i predykcyjnych, oszczędnego opisu lokalnych zależności danych, konstrukcja słownika z najczęściej występującymi frazami (ciągami danych wejściowych), wykorzystanie wiedzy dostępnej *a priori* na temat kompresowanego zbioru danych, poprzedzone niekiedy przekształceniem (dekompozycją) danych zwiększającą ich podatność na modelowanie, a w konsekwencji - na kodowanie binarne, itd. Istotne są przy tym kontekstowe zależności danych sąsiednich w sekwencji wejściowej lub też w oryginalnej przestrzeni danych kodowanych, np. pewne metryczne zależności w przestrzeni koloru i w przestrzeni geometrycznej w obrazach.

Na podstawie uzyskanej reprezentacji pośredniej tworzona jest binarna sekwencja wyjściowa (kodowa), poprzez przypisanie ciągów bitów (słów kodowych) poszczególnym danym (pojedynczym symbolom alfabetu źródła informacji), całej sekwencji danych wejściowych lub jej poszczególnym częściom.

Kod dwójkowy stałej długości

Typowa reprezentacja źródłowych danych cyfrowych ma postać kodu dwójkowego stałej długości. Kod dwójkowy B_k jest kodem symboli, nazywanym także kodem naturalnym ponieważ słowa tego kodu są dwójkową reprezentacją kolejnych liczb naturalnych (tj. nieujemnych liczb całkowitych). B_k charakteryzuje stała, k bitowa precyzja słów przypisanych n symbolom alfabetu A_S, gdzie $k \triangleq \lceil \log_2 n \rceil$. Przykładowo, przy $n = 8$ mamy następujące słowa kodowe o precyzji 3 bitów: $A_{B_3} = \{000, 001, 010, 011, 100, 101\}$.

Słowa kodowe B_k oznaczmy jako $\varsigma_i = (i)_{2,k}$ (dwójkowy zapis indeksów $i \in \{0, \ldots, n-1\}$ kolejnych symboli w $A_S = \{a_0, \ldots, a_{n-1}\}$). Prosty algorytm zakodowania symboli A_S metodą przyrostową za pomocą kodu dwójkowego wygląda następująco (algorytm 2.1):

Algorytm 2.1 *Kodowanie symboli źródła z wykorzystaniem kodu dwójkowego stałej długości*

1. Pobierz kolejny symbol do zakodowania $s_i = a_j \in A_S = \{a_0, \ldots, a_{n-1}\}$ lub zakończ;
2. Dołącz do sekwencji wyjściowej bitowy ciąg: $\varsigma_i = (i)_{2,k}$;
3. Kontynuuj krok 1.

□

Algorytm dekodera kodu dwójkowego o k bitowej długości słów, który sprowadza się do kolejnego odczytywania zapisanej na k bitach pozycji dekodowanego symbolu w alfabecie A_S, jest następujący (algorytm 2.2):

Algorytm 2.2 *Dekoder kodu dwójkowego o stałej długości*

1. Czytaj ciąg k bitów i zapisz go jako α; jeśli liczba bitów możliwych do przeczytania jest mniejsza od k to zakończ;
2. Dla $\alpha = (i)_{2,k}$ jeśli $i < n$, to emituj na wyjście symbol $a_i \in A_S$; w przeciwnym razie sygnalizuj błąd dekodera i zakończ;
3. Kontynuuj krok 1.

□

Ponadto, zapis danych źródła w kodzie B_k można wykorzystać w procedurze redukcji nadmiarowości oryginalnej reprezentacji danych o dynamice równej n_o możliwych wartości. Odwzorowanie n różnych wartości danych (symboli) źródła, gdzie $n < n_o$ w n kolejnych słów kodowych B_k pozwoli uzyskać stopień kompresji $CR > 1$. Podobny eksperyment wykonano w przypadku kompresji danych obrazowych o niewykorzystanej pełnej dynamice bajtowych wartości pikseli [22].

Kodowanie długości sekwencji (serii)

Metoda kodowania długości serii RLE (*Run Length Encoding*) jest intuicyjną metodą oszczędnego zapisu ciągów jednakowych symboli; jest wykorzystywana powszechnie w wielu współczesnych standardach i narzędziach kompresji.

Kod RLE należy do grupy kodów strumieniowych i sprowadza się do prostej reguły: seria kolejnych powtórzeń symboli źródłowych opisywana jest słowem kodowym określonej długości składającym się dwóch części – binarnej reprezentacji długości serii l_i oraz symbolu s_i. Zmienna liczba danych wejściowych, równa długości serii, kodowana jest za pomocą ciągu bitów o prawie stałej długości. Długość ta zależy od liczby symboli alfabetu źródła informacji oraz dopuszczalnej długości serii powtórzeń $k_{\max}$ (dobranej np. na podstawie obserwacji źródła, oceny jego właściwości we wstępnej analizie kodowanej sekwencji danych).

Ogólnie, kodowanie długości serii wykorzystuje dwa kody: $\mathcal{K}_l$ do zapisu liczby powtórzeń (długości serii) $k = l_i$ oraz $\mathcal{K}_s$ do zakodowania symbolu $s_i = a_j \in A_S = \{a_1, \ldots, a_n\}$. Zwykle stosuje się kody dwójkowe o różnej precyzji lub też, do zapisu liczby powtórzeń, kody zmiennej długości. Alfabet słów kodowych wygląda wtedy następująco:
$A_{\mathrm{RLE}} = \{(\mathcal{K}_l(k), \mathcal{K}_s(a_j)) : \ k = 1, 2, \ldots, k_{\max}, \ j = 1, \ldots, n\}$

Koncepcję RLE wykorzystano m.in. w znanym formacie zapisu obrazów PCX , w podstawowym algorytmie kodowania binarnego normy JPEG, czy też w algorytmie kodowania obrazów skanowanych według techniki DjVu (http://djvu.org/).

Pojęcie nadmiarowości

Bezstratna redukcja rozmiaru określonej sekwencji danych wejściowych możliwa jest dzięki różnego typu nadmiarowości oryginalnej reprezentacji tej sekwencji. Proces kompresji polega więc na efektywnym zmniejszaniu lub w najlepszym przypadku całkowitej eliminacji nadmiarowości reprezentacji danych źródłowych.

Zwykle informacja ze źródeł pierwotnych podawana jest w postaci, która nie nadaje się do bezpośredniego przetwarzania, archiwizacji czy przesyłania w systemach cyfrowych. Konieczne jest przekształcenie dostarczanej przez źródło informacji, często o charakterze analogowym, w dyskretny ciąg symboli, tj. elementów alfabetu o skwantowanych wartościach. Bitowa reprezentacja symboli powinna się charakteryzować odpowiednim stopniem złożoności, odpowiadającym naturze (znanym właściwościom) rejestrowanej informacji. W tym celu potrzebna jest reguła przyporządkowania symboli tego alfabetu złożonym formom postaci, w jakich występuje informacja danego źródła. Za pomocą tej reguły tworzony jest ciąg symboli źródła informacji, czyli oryginalna reprezentacja danych wejściowych poddawana kompresji.

Proste przykłady takich reguł tworzących reprezentacje danych to przyporządkowanie naturalnym pojęciom opisującym świat, ludzkim wrażeniom, odczuciom ciągów liter, słów, wyrażeń układających się w sensowne zdania określonego

języka zapisane z wykorzystaniem kodów ASCII. Informację o charakterze ziarnistym można opisać za pomocą ciągów liczbowych, np. w systemie dwójkowym. Urządzenia pomiarowe, systemy akwizycji różnego typu danych rejestrują za pomocą czujników sygnały naturalne, a przetworniki analogowo-cyfrowe zapewniają ich konwersję do postaci cyfrowej o odpowiedniej dynamice, opisanej skończonym alfabetem źródła.

W systemach gromadzenia danych stosowany jest zwykle uniwersalny format danych, który uwzględnia charakter rejestrowanych zjawisk: dynamikę rejestrowanego procesu, konieczną dokładność rozróżnienia informacji szczegółowych, zależności czasowe, stabilność, krótkoterminowe i długoterminowe tendencje zmian, itp. oraz zapewnia wygodny odczyt danych, łatwość analizy i przetwarzania, itd. Powoduje to często znaczną redundancję reprezentacji w odniesieniu do wybranej, zarejestrowanej w określonym przedziale czasowym sekwencji danych.

Ponadto, naturalne właściwości rejestrowanego zjawiska przekładające się na cechy informacji wyrażonej za pomocą sekwencji danych o określonej reprezentacji powodują, że pomiędzy danymi tej sekwencji (najczęściej kolejnymi, ale nie tylko) pojawiają się zazwyczaj różnego typu lokalne (a czasami nawet bardziej globalne) zależności, np. wielokrotne kolejne powtórzenie tej samej wartości (symbolu) w ciągu danych.

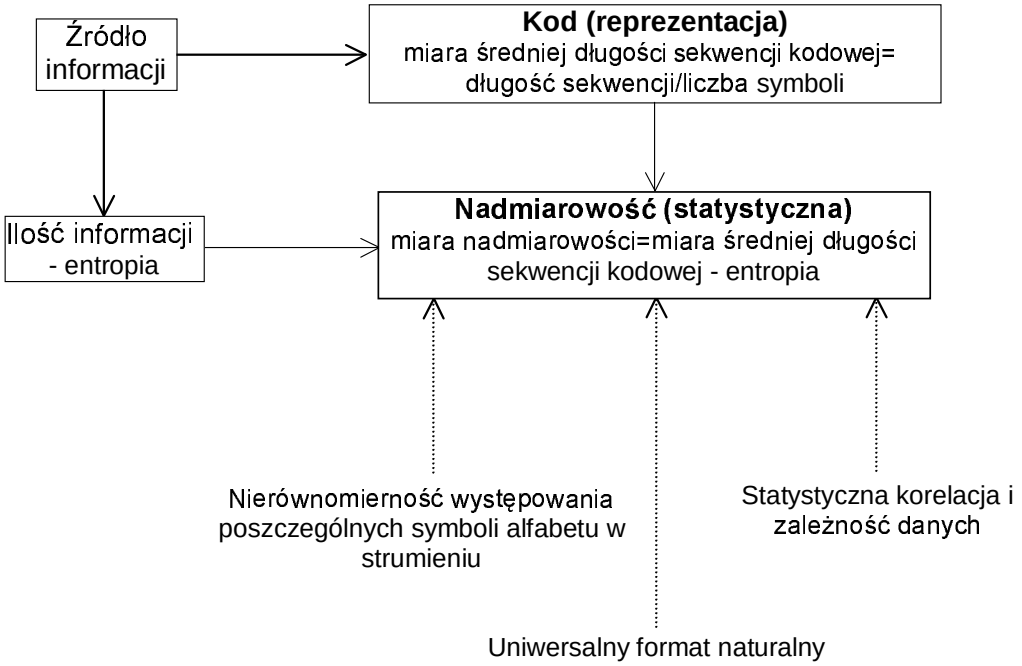
Z reguł języka polskiego wynika, że statystycznie rzecz biorąc znacznie częściej po literce 't' występuje literka 'a' niż 'x', a po 'ź' prawie nigdy nie występuje drugie 'ź' czy 'ż' itd. Natomiast w typowych fragmentach tekstu literka 'a' występuje znacznie częściej niż 'ą' czy 'w'. W obrazach przedstawiających obiekty o rozmiarach większych od pojedynczego piksela wartości sąsiednich pikseli są ze sobą skorelowane [9], a cały obraz można zazwyczaj scharakteryzować poprzez określenie dominującego koloru.

Zależności te można wyznaczać np. za pomocą statystycznego rozkładu wartości danych w zbiorze wejściowym wykorzystując histogram. Rozkład ten jest zwykle nierównomierny w wersji globalnej (dla całego ciągu danych kodowanych), a już na pewno gdy jest liczony lokalnie (dla fragmentu tego ciągu). Wagi poszczególnych wartości (symboli) są wówczas wyznaczane niezależnie i aproksymują niezależne prawdopodobieństwa symboli alfabetu w modelu źródła bez pamięci.

Wpływ wartości występujących w pewnym sąsiedztwie (kontekście) na to jaki będzie kodowany aktualnie symbol można określić za pomocą histogramu wielowymiarowego, warunkowego z kontekstem przyczynowym. Odpowiada mu model prawdopodobieństw warunkowych źródła z pamięcią.

Poniżej zdefiniowano pojęcie nadmiarowości statystycznej w wersji ogólnej i bardziej praktycznej.

[9] Korelacja to szczególny przypadek zależności danych, tj. zależność liniowa. Dekorelacja nie zawsze oznacza więc niezależność. Zaletą często stosowanego opisu informacji za pomocą procesu gaussowskiego jest statystyczna niezależność zdekorelowanych danych gaussowskich.



Rysunek 2.5: Ilościowe szacowanie statystycznej nadmiarowości oryginalnej reprezentacji kompresowanych danych oraz główne źródła tej nadmiarowości.

Definicja 2.9 *Nadmiarowość źródłowa stochastyczna*
Nadmiarowość stochastyczna sekwencji danych źródła informacji określana jest jako różnica pomiędzy entropią tego źródła i średnią bitową reprezentacji danych.
□

Definicja 2.10 *Nadmiarowość kodowania*
Nadmiarowość zakodowanej reprezentacji sekwencji danych, uzyskanej za pomocą kodu wykorzystującego określony model źródła informacji, jest to różnica pomiędzy entropią tego źródła i średnią bitową reprezentacji kodowej.
□

Metodę wyznaczania liczbowej miary nadmiarowości w sensie statystycznym oraz podstawowe przyczyny nadmiarowości reprezentacji danych źródłowych przedstawiono na rys. 2.5.

Dokładniejsza analiza typu nadmiarowości silnie zależy od rodzaju danych i charakteru zawartej tam informacji. W przypadku danych obrazowych można wyróżnić następujące typy nadmiarowości:

- przestrzenna (wewnątrzobrazowa i międzyobrazowa), związana z występowaniem zależności pomiędzy wartościami sąsiednich pikseli, zarówno w obrębie jednego obrazu, jak też serii obrazów kolejnych (zbiory danych trójwymiarowych);

- czasowa, pojawiająca się wskutek korelacji obrazów sekwencji rejestrowanej w kolejnych chwilach czasowych;

- spektralna, wynikająca z korelacji komponentów w obrazach wielokomponentowych (kolorowych, pseudokolorowych, innych);

- percepcyjna, powodowana niedoskonałością narządu wzroku odbierającego informację; część informacji może być nieprzydatna, bo obserwator nie jest w stanie jej zauważyć, a więc można ją usunąć w metodach stratnych;

- semantyczna, powstająca na poziomie interpretacji informacji, która wynika z faktu, że nie cała informacja reprezentowana ciągiem danych jest użyteczna dla odbiorcy; informacja ta podlega redukcji w metodach kompresji stratnej.

Warto podkreślić duże znaczenie nadmiarowości semantycznej i percepcyjnej, choć są one wykorzystywane w konstrukcji metod stratnych. Metody bezstratne mogą jednak stanowić ich uzupełnienie, np. do kodowania wybranych obszarów zainteresowania o dużym znaczeniu diagnostycznym, czy też do archiwizacji wiernych wersji obrazów źródłowych w celach badawczych, porównawczych, aby uczynić zadość rygorom prawnym.

Przykładowo nadmiarowość semantyczna występuje często w obrazach medycznych. Znaczna ilość informacji zawarta w obrazach może nie być istotna diagnostycznie, a więc jej redukcja, zniekształcenie czy całkowita eliminacja nie zmniejsza wiarygodości diagnostycznej obrazu. W niektórych rodzajach badań, np. scyntygraficznych duże obszary obrazu pokryte są jedynie szumem wynikającym z metody pomiarowej, bądź występują tam artefakty bez żadnej wartości diagnostycznej.

Szum i artefakty mogą nawet utrudniać dalszą analizę obrazów w systemach medycznych prowokując błędną interpretację u mniej doświadczonego radiologa, a ich kodowanie jest wyjątkowo mało efektywne (wartość entropii może być znacząca). Semantyczne rozumienie informacji użytkowej i nadmiarowości ma znaczenie nadrzędne w diagnostycznej interpretacji obrazów.

Podstawowe zasady kodowania

Algorytmy kodowania wykorzystujące opisane wyżej modele źródeł informacji pozwalają tworzyć wyjściową *reprezentację kodową*, która jest sekwencją bitową o skończonej długości, utworzoną z bitowych słów kodowych charakterystycznych dla danej metody. Realizację algorytmu kodowania, np. w określonym języku programowania lub sprzętową, nazwiemy *koderem*. Analogicznie realizację algorytmu dekodowania – *dekoderem*. Algorytm kodowania realizuje *kod*, czyli wspomnianą regułę (zasadę, funkcję, przekształcenie) przyporządkowującą ciągowi symboli

wejściowych (opisanych modelem źródła informacji o zdefiniowanym alfabecie A_S) bitową sekwencję kodową (wyjściową).

Kodowanie binarne jest metodą wykorzystania określonego kodu do kompresji sekwencji danych wejściowych źródła informacji A_S. Kod ten bazuje na określonym modelu źródła informacji, który steruje procesem kodowania binarnego. Podstawowy algorytm kodowania składa się więc z dwóch zasadniczych elementów: modelowania oraz binarnego kodowania.

Kody jednoznacznie dekodowalne W metodach kompresji, w przeciwieństwie np. do technik szyfrowania, istnieje warunek konieczny, aby reprezentacja kodowa (tj. bitowa sekwencja wyjściowa powstała w wyniku realizacji reguły kodu na strumieniu wejściowym) była jednoznacznie dekodowalna. Oznacza to, iż na podstawie wyjściowej sekwencji bitowej kodera realizującego ustalony kod opisany funkcją kodowania $\mathcal{K}$ można jednoznacznie odtworzyć oryginalny zbiór symboli wejściowych.

Kodowanie jest więc przekształceniem różnowartościowym 'jeden w jeden', czyli bijekcją. Jeśli dla ciągów symboli wejściowych $s'_1, s'_2, \ldots, s'_t$ oraz $s''_1, s''_2, \ldots, s''_r$ o wartościach z alfabetu $A_S = \{a_1, \ldots, a_n\}$ przyporządkowano bitową sekwencję kodową $z \in Z$ (Z - zbiór wszystkich sekwencji bitowych generowanych przez $\mathcal{K}$):

$$\mathcal{K}(s'_1, \ldots, s'_t) = \mathcal{K}(s''_1, \ldots, s''_r) = z \quad \Rightarrow \quad t = r \quad \text{oraz} \quad s'_i = s''_i, \quad i = 1, \ldots, t \tag{2.14}$$

Oznaczmy przez A_S^+ zbiór wszystkich niepustych i skończonych sekwencji symboli alfabetu A_S o N symbolach. Wtedy ogólna postać funkcji kodowania $\mathcal{K}: A_S^+ \rightarrow A_{\{0,1\}}^+$ z binarnym alfabetem sekwencji wyjściowych $\{0, 1\}$.

Przykładem kodu jednoznacznie dekodowalnego jest prosty kod dwójkowy stałej długości B_4 (zobacz punkt 2.3.2 na stronie 93) z czteroelementowym alfabetem A_S. Każdemu z symboli alfabetu przypisano słowo kodowe w postaci binarnego zapisu liczby naturalnej od 0 do 3, będącej indeksem (wskazującej pozycję) danego symbolu w A_S na $k = \log_2 4 = 2$ bitach. Zbiór słów kodowych jest więc następujący: $A_{B_4} = \{\texttt{00}, \texttt{01}, \texttt{10}, \texttt{11}\}$, a kod $\mathcal{K} = B_4$ jest odwzorowaniem różnowartościowym (co wynika z unikatowej postaci binarnej reprezentacji indeksu kolejnych symboli A_S).

Kodowanie według B_4 polega na przypisaniu kolejnym symbolom sekwencji wejściowej s_i, $s_i \in A_S$, $i = 1, 2 \ldots$ odpowiednich binarnych słów kodowych $B_4(s_i)$, takich że $B_4(s_i = a_1) = \texttt{00}$, $B_4(s_i = a_2) = \texttt{01}$, $B_4(s_i = a_3) = \texttt{10}$ i $B_4(s_i = a_4) = \texttt{11}$, dołączając je do wyjściowej sekwencji bitów (konkatenacja bitów słów kodowych symboli źródła S według porządku ich występowania na wejściu kodera). Praca dekodera tego kodu polega na odtwarzaniu symboli z A_S według czytanych kolejno 2 bitowych indeksów, co pozwala jednoznacznie zdekodować sekwencję kodową i wiernie zrekonstruować postać źródłowej sekwencji $\mathbf{s}^t$.

Kodowanie według B_k jest przykładem tzw. *kodowania przyrostowego*: $\mathcal{K}(s_1, s_2, \ldots, s_t) = \mathcal{K}(s_1)\mathcal{K}(s_2)\cdots\mathcal{K}(s_t)$, gdzie $s_i \in A_S$, a $\mathcal{K}(s_i) \in A_\mathcal{K} = \{\varsigma_1, \varsigma_2, \ldots, \varsigma_n\}$ ($A_\mathcal{K}$ nazwiemy alfabetem kodu $\mathcal{K}$). Ogólniej kodowanie przyrostowe polega na kolejnym dołączaniu sekwencji bitowych, przypisywanych czytanym sekwencjom symboli wejściowych według określonej reguły (tj. kodu), tworząc jedną sekwencję kodową. Kod może przypisywać sekwencje kodowe pojedynczym symbolom (jak wyżej) lub całym wieloelementowym grupom (blokom) symboli (zobacz rozróżnienie kodów symboli i strumieniowych na str. 104).

Dany jest alfabet $A_S = \{a_1, a_2, a_3\}$ oraz zbiór wszystkich niepustych sekwencji nad tym alfabetem A_S^+. Trywialnym przykładem kodowania, które po rozszerzeniu nie jest jednoznacznie dekodowalne, jest funkcja $\mathcal{K}_1(a_i) = \beta$, która dowolnemu symbolowi z A_S przypisuje to samo słowo kodowe $\beta \in A_{\{0,1\}}^+$. Dekoder odczytując β nie jest w stanie podjąć jednoznacznej decyzji: $\mathcal{K}_1^{-1}(\beta) = a_1 \vee \ldots$ $\ldots \vee \mathcal{K}_1^{-1}(\beta) = a_2 \vee \mathcal{K}_1^{-1}(\beta) = a_3$.

Innym przykładem kodu niejednoznacznie dekodowalnego przy dwuelementowym alfabecie źródła jest: $\mathcal{K}_2(a_1) = \texttt{0}$ i $\mathcal{K}_2(a_2) = \texttt{00}$, kiedy to sekwencję $\texttt{00}$ można zdekodować jako ciąg symboli (a_1, a_1) lub też jako pojedynczy symbol a_2. Przy większym alfabecie źródła, np. $A_S = \{a_1, a_2, a_3, a_4\}$ określmy $\mathcal{K}_3$ jako: $\mathcal{K}_3(a_1) = \texttt{0}$, $\mathcal{K}_3(a_2) = \texttt{1}$, $\mathcal{K}_3(a_3) = \texttt{01}$ i $\mathcal{K}_3(a_4) = \texttt{10}$. Dekodowanie sekwencji bitowych $\beta_1 = \texttt{01}$ lub $\beta_2 = \texttt{10}$ nie jest jednoznaczne, gdyż $\mathcal{K}_3^{-1}(\beta_1) = a_3$ lub też $\mathcal{K}_3^{-1}(\beta_1) = (a_1, a_2)$, a w przypadku β_2 mamy $\mathcal{K}_3^{-1}(\beta_2) = a_4 = (a_2, a_1)$.

Koder $\mathcal{K}_4$ przyporządkuje czterem symbolom alfabetu źródła informacji A_S następujące słowa kodowe: $A_{\mathcal{K}_4} = \{\mathcal{K}_4(a_1), \mathcal{K}_4(a_2), \mathcal{K}_4(a_3)\mathcal{K}_4(a_4)\} =$
$= \{\varsigma_1^{(\mathcal{K}_4)}, \varsigma_2^{(\mathcal{K}_4)}, \varsigma_3^{(\mathcal{K}_4)}, \varsigma_4^{(\mathcal{K}_4)}\} = \{\texttt{0}, \texttt{01}, \texttt{001}, \texttt{0011}\}$. Rozpatrzmy krótką sekwencję kodową $\beta = \texttt{001}$ utworzoną ze słów kodowych alfabetu $A_{\mathcal{K}_4}$, która może być interpretowana jako połączenie słów $\texttt{0}$ i $\texttt{01}$ ($\varsigma_1\varsigma_2$) lub też $\texttt{001}$ (czyli ς_3). Kod realizowany przez ten koder również nie jest kodem jednoznacznie dekodowalnym, a można to stwierdzić na podstawie analizy alfabetu postaci (zbioru) słów kodowych danego kodu. Dla koderów wykorzystujących inne zbiory binarnych słów kodowych można w analogiczny sposób stwierdzić, że:

- kod $A_{\mathcal{K}_5} = \{\texttt{1}, \texttt{01}, \texttt{001}, \texttt{0001}\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_6} = \{\texttt{0}, \texttt{10}, \texttt{110}, \texttt{111}\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_7} = \{\texttt{0}, \texttt{10}, \texttt{11}, \texttt{111}\}$ nie jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_8} = \{\texttt{0}, \texttt{01}, \texttt{11}\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_9} = \{\texttt{001}, \texttt{1}, \texttt{100}\}$ nie jest jednoznacznie dekodowalny.

Aby kod był jednoznacznie dekodowalny musi mieć N różnych słów kodowych (w przypadku $\mathcal{K}_1$ warunek ten nie jest spełniony). Jeśli w danym kodzie jedno ze słów kodowych jest konkatenacją innych słów, wówczas kod nie jest jednoznacznie

dekodowalny (zobacz przypadki kodów $\mathcal{K}_2$ i $\mathcal{K}_3$). Uogólniając tę zasadę, jeśli połączenie kilku słów kodowych daje taki sam efekt, jak połączenie innych słów z alfabetu kodu, wówczas także mamy do czynienia z brakiem jednoznaczności dekodowania. Tak jest w przypadku kodu $\mathcal{K}_7$, gdzie połączenie słów $\varsigma_3\varsigma_2 = \varsigma_4\varsigma_1$, jak również w przypadku kodu $\mathcal{K}_9$, ponieważ $\varsigma_2\varsigma_1 = \varsigma_3\varsigma_2$.

Kody przedrostkowe W przypadku kodów o liczniejszych alfabetach źródeł informacji trudno jest nieraz zweryfikować kod pod względem jednoznaczności dekodowania. Można jednak zauważyć, że we wszystkich przypadkach przedstawionych powyżej kodów, które nie są jednoznacznie dekodowalne, jedno ze słów kodowych jest przedrostkiem innego. Słowo $\varsigma_i \in A_\mathcal{K}$ jest *przedrostkiem* słowa $\varsigma_j \in A_\mathcal{K}, i \neq j$, jeśli $\varsigma_j = \varsigma_i\beta$, gdzie $\beta \in A^+_{\{0,1\}}$. Dla kodów $\mathcal{K}_4, \mathcal{K}_5$ oraz $\mathcal{K}_7$ żadne ze słów nie jest przedrostkiem innego. Są to więc kody przedrostkowe, tj. dla dowolnej pary różnych symboli $a_i, a_j \in A_S$ zachodzi relacja $\mathcal{K}(a_i) \not\preceq \mathcal{K}(a_j)$ (gdzie $\preceq$ oznacza relację początku w sekwencji bitowej). Relacja ta przenosi się na sekwencje symboli źródła A_S. Skoro żaden element z $A_\mathcal{K}$ nie jest przedrostkiem innego słowa tego zbioru, to sekwencja kodowa α dowolnego ciągu symboli z A_S także nie może być przedrostkiem sekwencji słów kodowych β innego ciągu symboli alfabetu źródła. Efektem wystąpienia choćby jednego różnego symbolu w sekwencji wejściowej jest pojawienie się słowa, który nie jest przedrostkiem innego, co powoduje utratę cechy przedrostkowości sekwencji: $\alpha = \gamma_1\mathcal{K}(a_i)\gamma_2, \beta = \gamma_1\mathcal{K}(a_j)\gamma_2,\ i \neq j,\ \gamma_1, \gamma_2 \in A^+_\mathcal{K} \cup \emptyset \Rightarrow \alpha \not\preceq \beta$, ponieważ $\mathcal{K}(a_i) \not\preceq \mathcal{K}(a_j)$. Ta właściwość pozwala jednoznacznie zdekodować sekwencję kodową ciagu symboli wejściowych.

Kody przedrostkowe (*prefix codes*) w literaturze angielskojęzycznej nazywane są także *prefix condition codes*, *prefix-free codes* lub też *comma-free code*. Wydaje się, że nazwa *kody bezprzedrostkowe* lepiej oddaje istotę zagadnienia.

Ze względu na możliwość dekodowania od razu kolejnych słów kodowych (od lewej do prawej, co daje prostą budowę kodera), kody przedrostkowe znane są także jako *instantaneous codes*, czyli kody z natychmiastowym dekodowaniem, bez konieczności odczytywania dalszych bitów w celu poprawnej interpretacji sekwencji bitów. Prosty algorytm dekodera kodu przedrostkowego $\mathcal{K}$, który polega na odczytywaniu kolejnych słów kodowych i dekodowaniu odpowiadających im symboli źródła, jest następujący:

Algorytm 2.3 *Dekodowanie sekwencji kodu przedrostkowego*

1. Wyzeruj zmienną α przechowującą wejściową sekwencję bitów;

2. Czytaj do α tyle bitów, ile wynosi minimalna długość słowa kodowego w $A_\mathcal{K}$, a wobec braku danych wejściowych zakończ;

3. Porównanie α ze słowami kodu $\mathcal{K}$: dla $i = 0, \ldots, n-1$ sprawdź, jeśli $\alpha = \varsigma_i$ (gdzie $\varsigma_i \in A_{\mathcal{K}}$), to emituj na wyjście symbol $a_i = \mathcal{K}^{-1}(\varsigma_i)$, $a_i \in A_S$ i kontynuuj krok 1;

4. Czytaj bit z wejścia i dopisz go do α, a wobec braku danych wejściowych zakończ;

5. Przejdź do kroku 3.

□

Kody przedrostkowe mogą być reprezentowane za pomocą struktury binarnych drzew kodowych (są kodami drzew binarnych), gdzie etykietowane gałęzie ustalają kolejne bity słów symboli alfabetu źródła, przypisanych liściom tego drzewa. Przejście od liścia do korzenia ze spisaniem dwójkowych etykiet kolejnych gałęzi daje słowo kodowe danego symbolu (w odwrotnej kolejności, tj. od najmłodszego bitu do nastarszego). Przy dekodowaniu symbolu, odczytywane kolejno bity słowa kodowego (od najstarszego do najmłodszego) prowadzą od korzenia do odpowiedniego liścia. Wykorzystanie struktury drzewa umożliwia wygodną realizację kodera i dekodera danego kodu przedrostkowego.

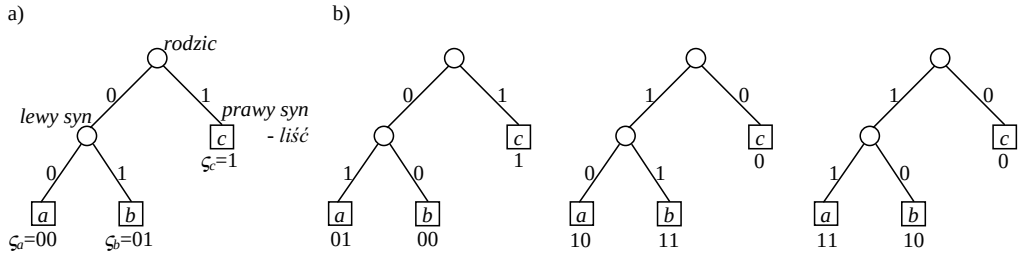
Przedrostkowość kodu jest warunkiem wystarczającym jego jednoznacznej dekodowalności. Nie jest jednak warunkiem koniecznym, co pokazuje przykład kodu $\mathcal{K}_8$. Kody jednoznacznie dekodowalne, nie będące kodami przedrostkowymi, są jednak trudniejsze w dekodowaniu (komplikuje się nieco budowa dekodera, a interpretacja sekwencji bitów odbywa się z opóźnieniem). Po odczytaniu słowa kodowego wymagają niekiedy odczytania kilku dodatkowych bitów, aby dokonać poprawnej jego interpretacji. Weźmy sekwencję kodową utworzoną według kodu $\mathcal{K}_8$ postaci: $\beta = \texttt{001110} = \varsigma_1\varsigma_2\varsigma_3\varsigma_1$. Zdekodowanie pierwszego symbolu $a_1 = \mathcal{K}_8^{-1}(\varsigma_1)$ trójelementowego alfabetu A_S wymaga odczytania dwóch pierwszych bitów (drugi bit sekwencji kodowej równy `0` wskazuje, że pierwsze `0` to ς_1, a nie pierwszy bit słowa ς_2). Aby poprawnie zdekodować bit drugi i trzeci sekwencji β (jako ς_2) koniecznym jest odczytanie wszystkich pozostałych bitów β (tj. czwartego, piątego i szóstego). Bowiem po odczytaniu piątego bitu jeszcze nie wiadomo, czy `0111` (bity od 2 do 5 sekwencji β) to złączone słowa $\varsigma_2\varsigma_3$, czy też konkatenacja słów $\varsigma_1\varsigma_3$ z pierwszym bitem kolejnego słowa ς_3. Dopiero po odczytaniu ostatniego `0` (szóstego bitu z β) wszystko się wyjaśnia (gdyby nie był to ostatni bit sekwencji kodowej, nie byłoby wiadomo czy należy dekodować ς_1, czy też jest to pierwszy bit ς_2).

Kody drzew binarnych Wykorzystanie struktury drzewa w projektowaniu kodu przedrostkowego pozwala w wygodny sposób realizować założenia dotyczące konstrukcji kodu efektywnego w kompresji danych o określonej charakterystyce. Utworzenie drzewa binarnego dla definiowanego kodu przedrostkowego ułatwia

analizę kodu, pozwala wykazać ewentualną jego nadmiarowość, umożliwia prostą jego modyfikację poprzez zmianę sposobu etykietowania drzewa (wpływającego na postać poszczególnych słów) lub dodanie liści (rozszerzenie alfabetu kodu o nowe słowa).

Właściwości kodu drzewa binarnego są określane na etapie konstruowania struktury drzewa dla założonej liczby liści, przy ustalaniu jego głębokości, poziomu zrównoważenia w skali całego drzewa lub wybranych poddrzew, sposobu rozmieszczenia symboli w liściach oraz zasad etykietowania. Głębokość umieszczenia poszczególnych liści decyduje o długości słów kodowych. Poziom zrównoważenia drzewa wpływa na zróżnicowanie długości słów symboli.

Formowanie słów kodowych poszczególnych liści odbywa się poprzez spisywanie etykiet gałęzi przejścia od korzenia do liścia, nigdy w kierunku odwrotnym (wynika to z naturalnego kierunku przejścia od rodzica do dzieci, pozwala rozróżnić potomstwo, choć nieco komplikuje budowę kodera wymuszając konieczność odwrócenia bitów). Idąc od korzenia na kolejnych poziomach formowany jest, w zależności od drogi, przedrostek słów kodowych. Przykładowo, przy ustalaniu słowa liścia z symbolem b w drzewie na rys. 2.6a, w węźle wewnętrznym pierwszego poziomu określony jest przedrostek `0`, a następnie idąc do prawego syna-liścia otrzymujemy słowo $\varsigma_b =$ `01`. W zależności od wybranego sposobu etykietowania gałęzi, binarne drzewo kodowe (tj. służące do realizacji danego kodu) o określonej strukturze wyznacza wiele kodów o takiej samej długości słów. Słowa te w sposób jednoznaczny opisują drogę przejścia od korzenia do symboli alfabetu A_S przypisanych poszczególnym liściom.



Rysunek 2.6: Etykietowanie drzew binanych: a) drzewo o dwóch wierzchołkach wewnętrznych i trzech liściach z symbolami a, b, c, z zaznaczoną relacją rodzic-dzieci oraz ustalonymi słowami kodowymi $\varsigma_a, \varsigma_b, \varsigma_c$; b) trzy inne sposoby etykietowania tego samego drzewa.

Przy etykietowaniu drzewa binarnego gałęzie łączące wierzchołek rodzica z dziećmi muszą mieć różne etykiety. Dwa możliwe sposoby etykietowania gałęzi każdego elementarnego poddrzewa (rodzic-dwójka dzieci) to `0` dla połączenia z lewym synem i `1` - z prawym lub odwrotnie (zobacz rys. 2.6a). Skoro istnieją dwie możliwości ustalania przedrostka dzieci każdego wierzchołka wewnętrznego (rodzica), to przy M wierzchołkach wewnętrznych w drzewie mamy 2^M różnych

sposobów utworzenia słów kodowych. Przykład różnych sposobów etykietowania drzewa o $M = 2$ przedstawiono na rys. 2.6.

W związku z tym, w przypadku budowy drzewa określonego kodu przedrostkowego możliwych jest 2^M różnych postaci drzewa binarnego, w zależności od przyjętego sposobu etykietowania. Długość słów poszczególnych symboli alfabetu tego kodu jest określona, czyli odpowiednie liście muszą zawsze znajdować się na tej samej głębokości, chociaż kształt struktury drzewa może być różny.

Długość słowa symbolu danego liścia drzewa jest równa głębokości (numerowi poziomu, licząc od zerowego poziomu korzenia), na jakiej znajduje się liść. Dla drzew z rysunku 2.6 liście a i b znajdują się na 2 poziomie (na głębokości równej 2), więc mają słowa dwubitowe, zaś liściowi c z pierwszego poziomu przypisano słowo jednobitowe. Zmiana sposobu etykietowania nie wpływa na długość poszczególnych słów.

Efektywna postać kodu drzewa binarnego Podstawowy warunek zapewniający efektywność kodu sprowadza się do konstrukcji drzewa, które nie ma zdegenerowanych elementarnych poddrzew, tj. rodzica tylko z jednym dzieckiem. Stąd każdy wierzchołek takiego drzewa, zwanego *drzewem lokalnie pełnym*, z wyjątkiem korzenia, ma brata. Puste miejsce w zdegenerowanym poddrzewie jest niewykorzystane - brakuje gałęzi z etykietą, dla której zarezerwowano miejsce w drzewie. Zawsze wtedy można zlikwidować wierzchołek rodzica takiego poddrzewa przesuwając na jego miejsce jedyne dziecko (wraz z poddrzewem) i skracając w ten sposób o jeden długość związanego z dzieckiem przedrostka słów liści z jego poddrzewa.

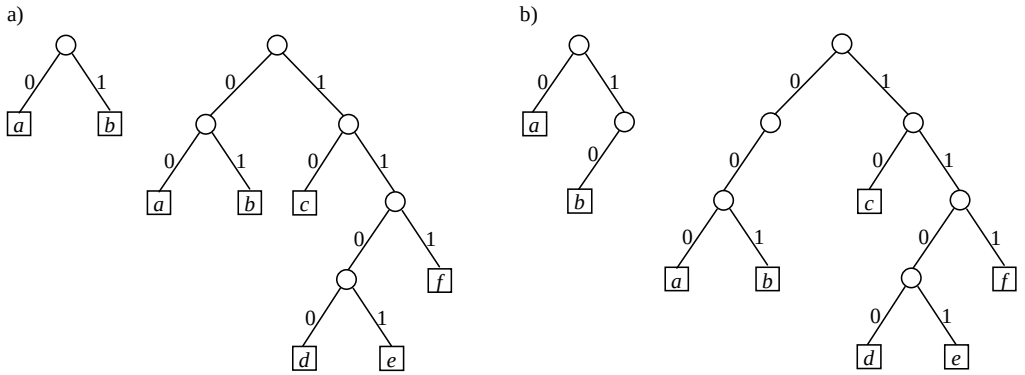
Drzewo lokalnie pełne o zadanej liczbie liści można zbudować, zaczynając np. od struktury z jednym wierzchołkiem, korzeniem, dodając nowe liście zgodnie z poniższą regułą.

Reguła rozbudowy drzewa lokalnie pełnego: wybrany węzeł aktualnej postaci drzewa przenosimy wraz z poddrzewem o jeden poziom niżej, zaś na jego miejsce wstawiamy nowy węzeł wewnętrzny, którego drugim dzieckiem jest nowo dołączany liść.

□

Drzewo o $N = 2$ liściach ma jeden węzeł wewnętrzny, tj. korzeń. Dodanie nowego liścia według tej reguły powoduje, że uzyskujemy drzewo z $N = 3$ liśćmi przy dwóch wierzchołkach wewnętrznych. W każdym kolejnym kroku rozbudowy drzewa dochodzi jeden liść i jeden wierzchołek wewnętrzny, a więc liczba wierzchołków wewnętrznych drzewa lokalnie pełnego jest zawsze równa: $M = N - 1$. Wszystkich wierzchołków drzewa jest $K = 2N - 1$. Przykłady drzew przedstawiono na rys. 2.7.

Aby uzyskać postać drzewa, które utraci cechę lokalnej pełności można zmo-



Rysunek 2.7: Przykłady drzew: a) lokalnie pełnych; b) nie będących lokalnie pełnymi.

dyfikować opisaną wyżej metodę konstrukcji drzewa lokalnie pełnego w sposób następujący. Wybrany liść przesuwamy poziom niżej, zaś w jego poprzednim miejscu tworzymy wierzchołek wewnętrzny (bez dołączania nowego liścia). W ten sposób powstaje zdegenerowane elementarne poddrzewo, rośnie o jeden liczba węzłów wewnętrznych przy niezmiennej liczbie liści równej N. W drzewach nie będących lokalnie pełnymi o N liściach liczba węzłów wewnętrznych jest więc równa N lub większa, a liczba wszystkich wierzchołków $K > 2N - 1$.

W mniejszym drzewie na rysunku 2.7a liczba wierzchołków $N = 2,\ K = 3$. Wersja tego drzewa po utracie cechy lokalnej pełności z rys. 2.7b ma odpowiednio $N = 2,\ K = 4$ (dwa wierzchołki wewnętrzne plus dwa liście), czyli $K > 2N - 1$. Nastąpiło wydłużenie długości słowa kodowego symbolu b o jeden bit. Analogicznie większe drzewo z rys. 2.7a ma $N = 6,\ K = 11$, jego modyfikacja na rys. 2.7b odpowiednio $N = 6,\ K = 13$ oraz wydłużone słowa symboli a i b o jeden bit.

Każdy wewnętrzny wierzchołek T drzewa lokalnie pełnego o lokalizacji jednoznacznie określonej jego przedrostkiem α ma zawsze dwóch synów: lewego i prawego o przedrostkach $\alpha\mathtt{0}$ i $\alpha\mathtt{1}$. Przedrostki te muszą pojawić się jako przedrostki słów wszystkich liści występujących w poddrzewie wierzchołka T. Stąd w kodzie takiego drzewa każdy przedrostek właściwy α (mający przedłużenie w słowie) dowolnego słowa kodowego ς_i ma przedłużenia $\alpha\mathtt{0}$ i $\alpha\mathtt{1}$ będące przedrostkami tego słowa oraz innego $\varsigma_j \neq \varsigma_i$. Kod drzewa lokalnie pełnego nazywany jest *kodem przedrostkowym pełnym*.

W większym drzewie z rys. 2.7a widać, że dowolny przedrostek związany z wierzchołkiem wewnętrznym ma przedłużenia z dołączonym bitem `0` oraz `1` w przedrostkach dzieci tego wierzchołka, a ostatecznie w słowach liści należących do jego poddrzewa. Z kolei słowa $\varsigma_a = \mathtt{0}$ i $\varsigma_b = \mathtt{10}$ mniejszego drzewa z rys. 2.7b świadczą o tym, że nie jest on przedrostkowo pełny: przedrostek właściwy $\alpha = \mathtt{1}$ słowa ς_b nie ma przedłużenia $\alpha\mathtt{1}$, a jedynie przedłużenie $\alpha\mathtt{0}$ w słowie symbolu b.

Kody symboli i kody strumieniowe Kod dwójkowy stałej długości, jak również inne wzmiankowane wyżej przykłady kodów pozwalające lub nie na jednoznaczne dekodowanie tworzonych sekwencji, należą do klasy tzw. kodów symboli (*symbol codes*). W *kodach symboli* sekwencja wyjściowa powstaje poprzez przypisanie kolejnym symbolom pojawiającym się na wejściu odpowiednich słów kodowych w schemacie „jeden symbol - jedno słowo". Aby uzyskać efekt kompresji potrzebne jest zróżnicowanie długości poszczególnych słów (w przeciwieństwie do kodu B_k, który ustala zwykle oryginalną reprezentację kodowanych danych). Bardziej praktyczne są więc kody o zmiennej długości bitowej słów kodowych (*variable-length symbol codes*).

Inną, potencjalnie bardziej efektywną grupę metod kodowania stanowią kody strumieniowe (*stream codes*). W *kodach strumieniowych* pojedyncze słowo kodowe przypisywane jest ciągowi symboli wejściowych. Może to być ciąg symboli o stałej lub zmiennej długości (np. w koderach słownikowych), a w skrajnym przypadku wszystkim symbolom strumienia wejściowego odpowiada jedno słowo kodowe, będące jednoznacznie dekodowalną reprezentacją danych źródłowych. Takie słowo tworzone jest w koderze arytmetycznym, stanowiącym automat skończony, który w każdym takcie (a więc po wczytaniu kolejnego symbolu wejściowego) wyprowadza (nieraz pustą) sekwencję bitów w zależności od czytanego symbolu i aktualnego stanu automatu.

Jednoznaczna dekodowalność kodów strumieniowych wynika z różnowartościowości przekształcenia sekwencji symboli wejściowych w słowo kodowe (sekwencję kodową). Decydujący wpływ ma tutaj metoda tworzenia słów kodowych, przystosowana do znacznie większego alfabetu możliwych postaci danych wejściowych, zwykle adaptacyjna i bardziej złożona niż w przypadku kodów symboli.

W charakterystyce kodów symboli o zmiennej długości istotną rolę odgrywa także nierówność Krafta-MacMillana. Dotyczy ona kodów jednoznacznie dekodowalnych zawierających n słów kodowych o długościach $L_i = |\varsigma_i|$, dla których zachodzi następująca zależność:

$$\sum_{i=1}^{n} 2^{-L_i} \leqslant 1 \tag{2.15}$$

Odwrotnie, mając dany zbiór n dodatnich liczb całkowitych $\{L_1, \ldots, L_n\}$ spełniających (2.15) istnieje jednoznacznie dekodowalny kod symboli o długościach kolejnych słów kodowych równych L_i (formalny dowód można znaleźć np. w [23]).

Nierówność (2.15) jest prawdziwa dla kodów drzew binarnych, co można wykazać na podstawie właściwości struktury drzewa. Dowolną postać drzewa binarnego daje się uzyskać poprzez rozbudowę drzewa najprostszej postaci. Kolejne liście dołączane są w wolnych miejscach istniejącej struktury węzłów lub też poprzez dodanie nowego węzła wewnętrznego.

Najprostsza (i najbardziej efektywna) postać drzewa z jednym liściem to korzeń z podpiętym bezpośrednio liściem. Mamy wtedy jedno słowo kodowe o dłu-

gości $L_1 = 1$, czyli spełniona jest nierówność (2.15). Wstawienie na wolne miejsce (jako drugie dziecko korzenia) drugiego liścia daje dwa słowa kodowe o długościach $L_1 = 1$ i $L_2 = 1$, wobec czego zachodzi równość $\sum_{i=1}^{2} 2^{-L_i} = 1$.

Uzyskaliśmy w ten sposób drzewo lokalnie pełne. Dodanie nowego liścia z zachowaniem lokalnej pełności drzewa może się odbywać według reguły rozbudowy drzewa lokalnie pełnego ze str. 103 stosowanej do dowolnego liścia. Powoduje to następującą zmianę w alfabecie słów kodowych: słowo $\varsigma_i = \alpha$ o długości L_i zastępowane jest dwoma słowami postaci $\varsigma_i' = \alpha 0$ i $\varsigma_i'' = \alpha 1$ o długości L_{i+1}. Odpowiedni składnik sumy z wyrażenia (2.15) się nie zmienia, gdyż $2^{-L_i} = 2^{-L_i+1} + 2^{-L_i+1}$. Według tej zasady można uzyskać dowolne drzewo lokalnie pełne. Pozwala to stwierdzić, że dla drzew lokalnie pełnych, czyli dla kodu przedrostkowego pełnego zawsze zachodzi równość: $\sum_{i=1}^{n} 2^{-L_i} = 1$.

Rozbudowa drzewa bez zachowania cechy lokalnej pełności powoduje zmniejszenie efektywności kodu poprzez wydłużenie słów kodowych. Utratę tej cechy powoduje modyfikacja drzewa polegająca na przesunięciu o jeden poziom w dół wierzchołka wraz z całym poddrzewem i ustanowieniu w zwolnionym miejscu nowego węzła wewnętrznego. Operacja ta powoduje wydłużenie długości słów kodowych wszystkich liści tego poddrzewa. Mamy więc $\sum_{i=1}^{n} 2^{-L_i'} < \sum_{i=1}^{n} 2^{-L_i} \leqslant 1$, gdzie L_i' to długości słów po modyfikacji. Dla kodu przedrostkowego bez cechy pełności mamy więc zawsze zależność: $\sum_{i=1}^{n} 2^{-L_i} < 1$.

Spełnienie nierówności (2.15) jest warunkiem koniecznym jednoznacznej dekodowalności kodu, nie jest jednak warunkiem wystarczającym (dostatecznym). Potwierdzeniem jest np. kod symboli określony słowami $A_{\mathcal{K}} = \{\mathtt{0}, \mathtt{10}, \mathtt{110}, \mathtt{101}\}$ (niewielka modyfikacja kodu $A_{\mathcal{K}_6}$) spełniający (2.15). Przykładowe kody $A_{\mathcal{K}_3}$ i $A_{\mathcal{K}_7}$ nie spełniają nierówności (2.15), co dowodzi, że nie są jednoznacznie dekodowalne. Zaś w przypadku kodów $A_{\mathcal{K}_2}, A_{\mathcal{K}_4}$ i $A_{\mathcal{K}_9}$ zależność ta nie wystarcza do wykazania braku niejednoznacznej dekodowalności.

Nierówność (2.15) można odnieść do pojęcia informacji własnej związanej z wystąpieniem (z prawdopodobieństwem p_i) pojedynczego symbolu kodowanego słowem o długości L_i zapisując $L_i = -\log_2 p_i + \epsilon_i$, gdzie ϵ_i jest nadmiarową—niedomiarową (w stosunku do wartości informacji własnej) długością słowa L_i (wynikającą np. z ograniczenia długości słowa do całkowitej liczby bitów). W wyniku prostych przekształceń $2^{-L_i} = 2^{(\log_2 p_i - \epsilon_i)} = p_i \cdot 2^{-\epsilon_i}$, co pozwala zapisać (2.15) jako $\sum_{i=1}^{n} p_i \cdot 2^{-\epsilon_i} \leqslant 1$. Jeśli założymy stałą wartość $\epsilon_i = \epsilon$ dla wszystkich słów, wtedy $\epsilon \geqslant 0$. W przypadku, gdy kod dobiera krótsze od informacji własnej słowa kodowe dla wybranych symboli alfabetu, to dla innych symboli słowa muszą być wydłużone (średnio) tak samo lub bardziej.

Twierdzenia o kodowaniu źródeł Przy konstruowaniu technik odwracalnej kompresji interesującym jest pytanie o granicę możliwej do uzyskania efektywności kompresji. Intuicyjnie wiadomo, że nie można stworzyć nowej reprezentacji da-

nych o dowolnie małej długości przy zachowaniu pełnej informacji dostarczonej ze źródła. Okazuje się, że entropia łączna $H(S)$, wyznaczona według równań (2.8) dla kodowanego zbioru danych stanowi graniczną (minimalną) wartość średniej bitowej reprezentacji kodowej - jest bowiem miarą ilości informacji pochodzącej ze źródła. Mówią o tym twierdzenia Shannona o kodowaniu źródeł, w tym szczególnie istotne twierdzenie o bezstratnym kodowaniu źródła (tj. z kanałem bezszumnym, z doskonałą rekonstrukcją sekwencji symboli źródła - bez ograniczenia liczby bitów) (*noiseless source coding theorem*) [21, 24], zwane też pierwszym twierdzeniem Shannona. Według tego twierdzenia, aby zakodować dany proces (sekwencję symboli generowanych przez źródło) o entropii $H(S)$ do postaci sekwencji symboli binarnych, na podstawie której możliwe będzie dokładne zdekodowanie (rekonstrukcja) procesu źródłowego, potrzeba co najmniej $H(S)$ symboli binarnych (bitów). Możliwa jest przy tym realizacja schematu kodowania, który pozwoli uzyskać bitową reprezentację informacji z takiego źródła o długości bardzo bliskiej wartości $H(S)$.

Bardziej praktyczna odmiana twierdzenia o bezstratnym kodowaniu źródła dotyczy granicznej efektywności kodowania, osiąganej poprzez kodowanie dużych bloków symboli alfabetu źródła (tj. N-tego rozszerzenia źródła, przy zmianie struktury informacji pojedynczej) za pomocą binarnych słów kodowych. Zamiast przypisywania słów kodowych pojedynczym symbolom alfabetu, jak w koderach symboli, koncepcja skuteczniejszego kodera prowadzi w kierunku realizacji kodu strumieniowego, uwzględniając przy tym w możliwie wydajny sposób zależności pomiędzy poszczególnymi symbolami kodowanej sekwencji.

Twierdzenie 2.2 *O bezstratnym kodowaniu źródła*
Niech S będzie ergodycznym źródłem z alfabetem o t wygenerowanych elementach i entropii $H(S)$. Ponadto, niech bloki po N ($N \leqslant t$) symboli alfabetu źródła S kodowane będą jednocześnie za pomocą binarnych słów kodowych dających kod jednoznacznie dekodowalny. Wówczas dla dowolnego $\delta > 0$ możliwa jest, poprzez dobór odpowiednio dużej wartości N, taka konstrukcja kodu, że średnia liczba bitów reprezentacji kodowej przypadająca na symbol tego źródła $\overline{L}_S$ spełnia równanie:

$$H(S) \leqslant \overline{L}_S < H(S) + \delta \tag{2.16}$$

Ponadto, nierówność $H(S) \leqslant \overline{L}_S$ jest spełniona dla dowolnego jednoznacznie dekodowalnego kodu przypisującego słowa kodowe N- elementowym blokom symboli źródła.
□

Okazuje się, że zawarta w tym twierdzeniu sugestia o zwiększaniu efektywności kompresji poprzez konstruowanie coraz większych rozszerzeń źródła (rosnące N) jest cenną wskazówką, ukazującą kierunek optymalizacji koderów odwracalnych. W dosłownej realizacji tej sugestii występuje jednak problem skutecznego okre-

ślenia prawdopodobieństw łącznego wystąpienia N symboli źródła na podstawie kodowanego strumienia danych.

Z twierdzenia o bezstratnym kodowaniu źródła jasno wynika, że każde źródło danych może być bezstratnie kodowane przy użyciu kodu jednoznacznie dekodowalnego, którego średnia liczba bitów na symbol źródła jest dowolnie bliska, ale nie mniejsza niż entropia źródła (określona na podstawie prawdopodobieństw występowania poszczególnych symboli i grup symboli źródła) wyrażona w bitach. Jest to naturalne ograniczenie wszystkich metod bezstratnej kompresji odnoszące się do założonych modeli źródeł informacji. Oczywiście użyty model źródła winien jak najlepiej charakteryzować (przybliżać) zbiór danych rzeczywistych. Należy więc podkreślić względność wyznaczanych wartości granicznych wobec przyjętego modelu źródła informacji. Przykładowo, przy konstruowaniu kodera na podstawie modelu źródła bez pamięci, graniczną wartością efektywności tego kodera będzie wartość entropii $H(S_{\mathrm{DMS}})$.

Uzyskanie większej skuteczności kompresji metod bezstratnych jest możliwe poprzez doskonalenie modelu źródła informacji przybliżającego coraz wierniej kompresowany zbiór danych (oryginalnych, współczynników falkowych itp.), nawet kosztem rosnącej złożoności modelu. Potrzeba coraz dokładniej modelować wpływ kontekstu, zarówno za pomocą rozbudowanych, dynamicznych (adaptacyjnych) modeli predykcyjnych, jak też coraz szybciej i pełniej określanych probabilistycznych modeli Markowa nawet wyższych rzędów. Modelowanie musi być skojarzone z efektywnymi rozwiązaniami kodów binarnych (powstających na podstawie tych modeli), pozwalających osiągnąć minimalną długość bitowej sekwencji nowej reprezentacji danych.

2.3.3 Semantyczna teoria informacji

W niektórych przypadkach semantyka obrazów odgrywa na tyle znaczącą rolę w interpretacji (użytkowaniu, odczytaniu zawartej informacji) przekazywanych danych, że powinna stanowić ważny element w procesie optymalizacji algorytmów kompresji jako "uzupełnienie" statystycznej teorii informacji. Ważnym obszarem zastosowań jest tutaj obrazowanie medyczne.

Dwa zasadnicze cele teorii Shannona to: a) obliczanie ilości informacji dostarczanej przez źródła możliwie wiernie opisujące przesyłane dane, b) opracowanie efektywnych kodów bazujących na właściwych modelach statystycznych źródeł informacji. Oba te elementy wymagają ustalenia znaczeń pojedynczej danej, kontekstu jej wystąpienia, charakteru reprezentowanej informacji.

Należy tutaj odwołać się do podstaw ogólnej teorii informacji, w której występuje pojęcie semantycznej teorii informacji. Początki tej teorii stanowią prace Bar-Hillela i Carnapa [16] również z lat 50 zeszłego wieku, a istotą jest określanie dodatkowo znaczenia poszczególnych symboli alfabetu źródła informacji. Wykorzystanie ontologicznych i aksjologicznych aspektów rozumienia informacji jest

trudne do przełożenia na formalny i algorytmiczny opis modeli źródeł informacji, jednak wykorzystanie elementów tych teorii wydaje się konieczne w takich zastosowaniach jak np. analiza, rekonstrukcja i indeksowanie treści obrazów medycznych w sposób wiarygodny diagnostycznie.

Algorytmiczne wykorzystanie zasad semantycznej teorii informacji może być nieco łatwiejsze poprzez skorzystanie z tych metod matematycznych, które ułatwią semantyczną selekcję w algorytmach rozpoznawania. Chodzi tu o wykorzystanie analizy funkcjonalnej, które pozwala w większym uwolnić się od nierealistycznych założeń statystycznych dobierając bazy przekształceń przybliżających efektywnie lokalne, chwilowe właściwości sygnału. Ułatwia to także dostosowanie metody analizy do semantyki źródła informacji poprzez większe uporządkowanie informacji (zhierarchizowanie jej opisu) i dokładniejszy jej opis (czasową charakterystykę wielu skal ułatwiającą rozdzielenie sygnału od szumu) w nowej dziedzinie przekształcenia.

W latach 50 zeszłego stulecia powstała "moskiewska szkoła teorii informacji" z jej najznakomitszym przedstawicielem A.N. Kołmogorowem. Obok probabilistycznych sposobów modelowania źródeł informacji wykorzystano tam także teorię przybliżania (aproksymacji) źródeł z wykorzystaniem metod analizy funkcjonalnej. Stochastyczny proces opisujący źródło informacji zastąpiony jest przez klasę funkcji (sygnałów) f określonych w dziedzinie T. Dowolna funkcja jest aproksymowana i dyskretyzowana przez koder za pomocą sieci aproksymacji. Sieć aproksymacji według koncepcji Kołmogorowa [17, 18] jest zbiorem funkcji możliwie zupełnym i mało-licznym (kontrolowanym przez inne pojęcie entropii), aproksymującym istotne cechy sygnałów (funkcji) źródłowych. Kryteria doboru i optymalizacji postaci sieci formułowane jako minimalizacja błędu przybliżenia definiują proces reprezentacji (modelowania) danych z dopuszczeniem strat.

Na poziomie ogólności proponowanym przez teorię Kołmogorowa bardzo niewiele można powiedzieć o strukturze optymalnej dla danego problemu sieci aproksymacji. Chociaż alternatywna teoria opisu informacji nie prowadzi to dokładnych oszacowań relacji złożoności modelu źródła do ilości reprezentowanej informacji, to jest istotna ze względu na sugestię funkcjonalnego modelowania źródeł. Chodzi o wyznaczenie efektywnych baz przekształceń obrazów w celu lepszego opisu informacji.

Wykorzystanie doświadczeń semantycznej teorii informacji, dodatkowo coraz doskonalszej wiedzy medycznej na temat zasad percepcji psychowizualnej danych obrazowych i obiektywizacji metod ich interpretacji pozwala konstruować doskonalsze sposoby selekcji i porządkowania analizowanej informacji na podstawie optymalizowanych metod opisu danych (duża rola analizy funkcjonalnej, przede wszystkim analizy harmonicznej, czyli analizy funkcji z wykorzystaniem transformacji o bazie wielu skal dobieranej w przestrzeni czas-częstotliwość).

Rozumienie obrazów

Rozumienie bazuje na wykrytych wcześniej w obrazie $\mathcal{I}$ obiektach $o \in \mathcal{O}_\mathcal{I}$, którym przypisywana jest niezerowa funkcja semantyczna Σ_o, definiująca znaczenie obiektu w kontekście określonych uwarunkowań jego występowania w obrazie (np. tło, postrzegane właściwości), a także mającej zastosowanie procedury interpretacji. Wykrycie wszystkich obiektów istotnych dla zrozumienia treści obrazu, poprawne przypisanie znaczeń, a także trafne określenie charakteru kontekstu ich wystąpienia, elementów tła itp. warunkuje poprawność kompleksowego procesu rozumienia treści.

Wykrycie obiektów istotnych nie jest jednak zwykle zadaniem prostym. Przykładem może być rozpoznanie niewielkich, lekko zarysowanych zmian, będących ostrzegawczym symptomem anormalności, czy nawet patologii w obrazach medycznych. Obiekt definiuje mniej lub bardziej dostrzegalna odmienność określonej cechy/zespołu cech względem najbliższego otoczenia, przy jednoczesnym zachowaniu podobieństwa całego zbioru ogólnych i szczególnych właściwości kontekstu występowania. Może to być bardzo subtelna różnica tekstury czy średniej barwy, wręcz niedostrzegalna przy rutynowej obserwacji czy też w typowej analizie numerycznej. Obiektem może być więc nieznacznie wyróżniający się obszar, bez wyraźnych konturów, którego trafne rozpoznanie wymaga generalnie odpowiedniej wiedzy $\mathcal{W}$ – ogólnej lub specyficznej, własnego doświadczenia, niekiedy szczególnych zdolności kojarzenia i wnioskowania, intuicji itp.. Funkcja semantyczna, określająca znaczenie obiektu zależy więc od wiedzy: $\Sigma_o(\mathcal{W})$.

Komputerowe przetwarzanie obrazów może obejmować doskonalenie etapu wykrycia i określenia charakteru obiektów, w tym

- poprawa jakości redukująca ograniczenia systemu akwizycji;
- przetwarzanie zmierzające do poprawy percepcji lokalnych cech obrazów;
- wyznaczenie dodatkowych cech obliczeniowych różnicujących obszary czy dzielących je na podobszary;
- zastosowanie czułych deskryptorów specyficznych cech obiektów;
- wskazanie potencjalnych obszarów zainteresowania, zgodnych z modelami wzorców, ale też m.in. anormalności, obszarów czy struktur nietypowych, podejrzanych;
- interaktywne dopasowanie warunków prezentacji (wizualizacji) wybranych obszarów, itp.

Aby ustalić treść przekazu obrazowego, należy dodatkowo określić znaczenie wzajemnych relacji pomiędzy obiektami występującymi w obrazie: $\Sigma_{\mathcal{R}(o_1,o_2,\ldots)}(\mathcal{W}) = \Sigma_{\mathcal{R}(\mathcal{O}_\mathcal{I})}(\mathcal{W})$, odwołując się do odpowiednich zasobów wiedzy,

zarówno *a priori* jak i *a posteriori*. Semantyczna funkcja relacji wzajemnych obejmuje wszystkie istniejące związki znaczeniowe, wynikające z łącznego wystąpienia dowolnego podzbioru obiektów $\mathcal{O}_\mathcal{I}$, przy uwzględnieniu wszelkich liczących się, realnych uwarunkowań zobrazowania $\mathcal{I}$.

Ze znaczenia wzajemnych relacji obiektów wynika, często na zasadach synergii, całościowa treść zawarta w danych obrazowych:

$$\Sigma_\mathcal{I}(\mathcal{O}, \mathcal{R}, \mathcal{W}) = \Sigma(\mathcal{I}, \mathcal{O}_\mathcal{I}, \Sigma_{\mathcal{O}_\mathcal{I}}, \Sigma_{\mathcal{R}(\mathcal{O}_\mathcal{I})}, \mathcal{W}) \tag{2.17}$$

gdzie wektor znaczeń obiektów $\Sigma_{\mathcal{O}_\mathcal{I}} = [\Sigma_{o_1}, \Sigma_{o_2}, \ldots]$. Ważne jest przy tym uwzględnienie wszelkiej dostępnej wiedzy dodatkowej, mającej znaczenie przy odczytywaniu całościowej treści obrazu. Wiedza ta wynika m.in. z okoliczności przeprowadzenia procesu akwizycji obrazu, ze sposobów wykorzystywania informacji obrazowej określonego typu, aktualnego celu odczytywania treści i stanowi naturalne uzupełnienie treści dostrzegalnej bezpośrednio z obrazu, wpływając na sumaryczną wymowę przekazywanej treści.

Poprawne rozumienie treści przekazu informacji obrazowej, czyli trafne, możliwie jednoznaczne odczytanie całej użytecznej zawartości przekazu pozwala realizować zasadnicze etapy użytkowania treści obrazowej, takie jak

a) ocena odczytanej treści w kontekście określonych celów użytkowania, rozpoznania klasy problemu lub rodzaju przypadku, np. detekcję (rozpoznanie) zmian (obszarów) podejrzanych w obrazie medycznym;

b) interpretacja rozpoznanej rzeczywistości na wyższym poziomie abstrakcji, zależnie od uwarunkowań zastosowania, np. przypisanie określonej kategorii diagnostycznej według stosowanej skali;

c) podjęcie określonych działań, wynikających z przyjętej interpretacji informacji obrazowej, ostatecznie potwierdzających przydatność odczytanej informacji.

Bezpośrednim owocem rozumienia (operator rozumienia oznaczmy przez $\mathcal{U}$) treści obrazowej $\Sigma_\mathcal{I}(\mathcal{O}, \mathcal{R}, \mathcal{W})$ jest ocena odczytanej treści, czyli wskazanie przedmiotu zainteresowania (SOI, czyli *subject of interests*) lub też stwierdzenie jego braku (brak np. poszukiwanego obiektu w określonym zobrazowaniu może mieć bardzo dużą wartość użytkową, np. brak poszukiwanego przestępcy w grupie przebywającej na monitorowanym dworcu). Mamy więc:

$$\mathrm{SOI} = \mathcal{U}\{\Sigma_\mathcal{I}(\mathcal{O}, \mathcal{R}, \mathcal{W})\} \tag{2.18}$$

Skuteczna segmentacja i rozpoznanie obiektów, określenie relacji pomiędzy nimi tak w zakresie porównania ich właściwości za pomocą numerycznych deskryptorów, jak też ich formalnego, ontologicznego opisu i wynikających z tego

zależności, pozwala scharakteryzować treść całego przekazu obrazowego, którego znaczenie może zostać określone za pomocą odwołań do specyficznych modeli wiedzy, mających źródła ontologiczne, referencyjne (porównanie z wzorcami) lub bazujące na różnorakich modelach numerycznych odwołujących się do semantyki (lingwistyczne, stochastyczne, funkcjonalne, obiektowe itp.). Interpretacja i podejmowanie działań czy formułowanie decyzji odwołuje się do głębszych zasobów wiedzy abstrakcyjnej, specjalistycznej, wieloletnich doświadczeń, własnych heurystyk, do ludzkiej świadomości, intuicji, zdolności kojarzenia i wnioskowania, których nie sposób skutecznie zamodelować w przypadku bardziej złożonych zastosowań [20].

Komputerowe wspomaganie procesu rozumienia obrazów zakłada podmiotowość osoby użytkownika i jedynie pomocniczą rolę komputerowej asystencji. Zakłada optymalizację procesu rozumienia obrazów, by zweryfikowana i klarowniej reprezentowana informacja obrazowa stanowiła przedmiot właściwych ocen, możliwie jednoznacznej interpretacji i trafnych decyzji specjalistów. Efektem automatycznego rozumienia obrazów może być więc uproszczona forma danych źródłowych typu SOI, po redukcji nadmiarowości semantycznej, z możliwie przejrzystą identyfikacją semantycznie rozpoznawalnych składników tej informacji, czyli struktury informacji.

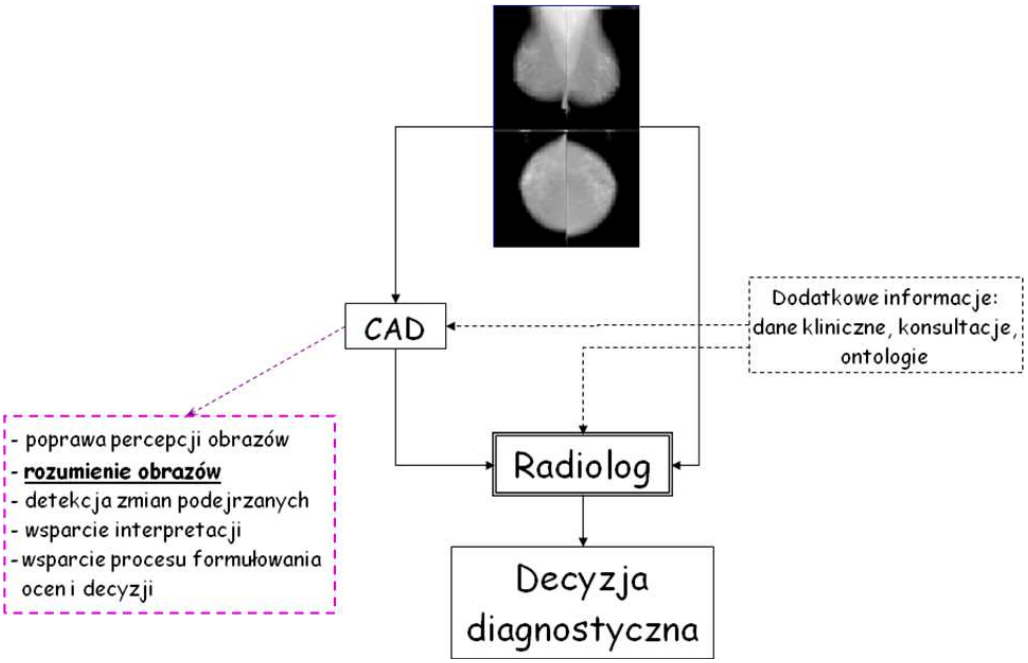
Jako przykład zastosowań, istotną rolę metod rozumienia obrazów medycznych wykorzystywanych w koncepcji komputerowego wspomagania obrazowej diagnostyki medycznej pokazano na rys. 2.8.

2.3.4 Indeksowanie, czyli znakowanie treści

Gwałtownie rosną cyfrowe zasoby danych multimedialnych, pęcznieją przepełnione archiwa, palące stają się potrzeby sprawnego dostępu do stale rozbudowywanych hurtowni danych. Według najnowszego raportu IDC Digital Universe Study[10] światowe zasoby danych cyfrowych ulegają podwojeniu co dwa lata, zaś przypuszczalna liczba danych wytworzonych i powielonych w 2011 roku szacowana jest na poziomie 1,8 zeta bajtów (czyli 10^{21} bajtów). Przeciążone sieci, trudności z szybkim wyszukaniem niezbędnych informacji, problem wydobycia właściwej treści z potoku nadmiarowych strumieni danych i tym podobne, coraz powszechniej występujące zjawiska wymagają coraz doskonalszych mechanizmów zarządzania zasobami danych cyfrowych.

Szczególnego znaczenia nabierają systemy selektywnego i możliwie szybkiego wyszukiwania pożądanej treści. Przeszukiwanie baz danych o rozległych, lawinowo rosnących zasobach wymaga obok efektywnych obliczeniowo struktur danych, wygodnych mechanizmów formułowania zapytań i prezentacji odpowiedzi, osadzonych w rzetelnie realizowanych systemach teleinformatycznych, także semantycznych technologii opisu treści. Szczególnie istotnym wymaganiem jest tutaj

[10]IDC Digital Universe Study *Extracting Value from Chaos*, 28.06.2011, sponsored by EMC.



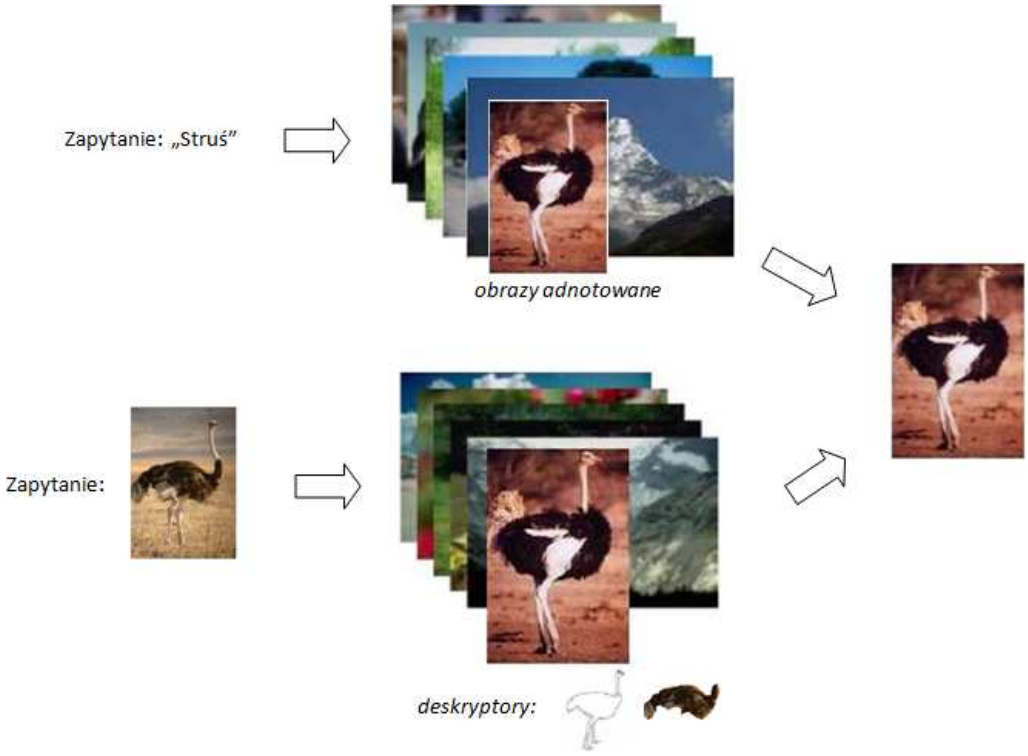
Rysunek 2.8: Przykład roli, jaką pełnią metody rozumienia obrazów w realizacji koncepcji komputerowego wspomagania obrazowej diagnostyki medycznej – CAD.

obliczeniowe **rozumienie** treści danych na sposób zbliżony do intencji użytkownika.

By wyszukać dane odpowiedniej treści, należy je opisać w sposób reprezentatywny, czyli tak, by uwzględnić wszystkie najistotniejsze właściwości, by różnicować obiekty zgodnie z oczekiwaniem użytkownika (klasy, kategorie), przy jednoczesnej, możliwie upakowanej formie opisu. Metoda wyszukiwania obok reprezentatywnego opisu wykorzystuje także w niektórych przypadkach funkcję podobieństwa obiektów określonego typu, dopasowaną do ich charakteru i właściwości. Znajduje ona zastosowanie przy wyszukaniu obiektów najbardziej podobnych do przykładu zapytania.

Zasadniczo stosowany jest opis dwojakiego rodzaju: a) tekstowy (*text-based*), bazujący na słowach kluczowych i określonej syntaktyce, wymagający zaangażowania osoby interpretującej treść danych; b) po zawartości (*content-based*), bazujący na automatycznej analizie treści oraz obliczeniowej jej charakterystyce za pomocą numerycznych deskryptorów sygnałowych. Wprzypadku obrazów konsekwencją są dwie metody wyszukiwania: TBIR (*text-based image retrieval*) oraz CBIR (*content-based image retrieval*), przedstawione schematycznie na rys. 2.9.

Solidny opis tekstowy wymaga dużych nakładów ludzkiej pracy, a na efekty końcowe duży wpływ ma czynnik subiektywny. Często też nie sposób za pomocą



Rysunek 2.9: Zestawienie koncepcji wyszukiwania obrazów z wykorzystaniem koncepcji opisu tekstowego TBIR (u góry) oraz obliczeniowej charakterystyki zawartości CBIR (na podstawie rysunku zaczerpniętego z [28].

ustalonych *a priori* reguł tworzenia tekstu wyrazić bogactwa treści zawartej w obrazie, filmie, czy nawet zapisie dźwiękowym. Często interpretacja treści ma charakter względny, zależny od kontekstu, okoliczności przywołania danych obiektów multimedialnych.

Drugi sposób opisu, bazujący na automatycznej analizie zawartości danych, w przypadku obiektów multimedialnych o złożonej treści jest często mało skuteczny – uzyskiwane efekty różnicującego opisu obiektów są ograniczone. Cechy numeryczne wyznaczane na podstawie obrazu czy zapisu dźwięku powinny potencjalnie obejmować szeroki, niemal nieograniczony zakres semantyczny możliwej treści. Teoria cyfrowego przetwarzania i analizy sygnałów dostarcza bogatego zestawu narzędzi do konstrukcji deskryptorów wizualnych opisujących obrazy, zgodnie z założonym zestawem argumentów, czy też deskryptorów audio do opisów dźwięku i mowy. Nierzadko problemem okazuje się jednak ustalenie zbioru cech reprezentatywnych oraz dobór takich kryteriów podobieństwa, które by odpowiadały oczekiwaniom użytkownika. Nieraz okazuje się, że drobny element treści obrazu, występujący na obszarze zajmującym mniej niż 1% jego powierzchni, decyduje o jego charakterze, nadaje zasadniczy sens wyrażonej treści – specyficzny

deskryptor tej cechy powinien zostać uwzględniony w pierwszej kolejności przy formułowaniu kryterium podobieństwa.

Podstawowe pojęcia

Indeksowanie multimediów to w pierwszym przybliżeniu tworzenie formalnego w sensie reprezentacji cyfrowej opisu treści multimedialnej, zawartej w różnego typu zbiorach i bazach danych, zbiorach obiektów lub ogólniej – kolekcjach multimedialnych, czy nawet strumieniach multimedialnych (indeksowanie i wyszukiwanie *on-line*)[29].

Atrybutem obiektu multimedialnego nazywana jest ustalona, istotna jego właściwość, służąca specyficznej (różnicującej) charakterystyce obiektu w odniesieniu do przeglądanych zasobów (kolekcji obiektów). **Cecha** to wartość danego atrybutu przypisana obiektowi. Dana cecha c obiektu o uwzględnia więc określony aspekt obiektu definiowany atrybutem a. Atrybut $a : \mathcal{O} \rightarrow \mathcal{C}_a$ przypisuje cechę $c \in \mathcal{C}_a$ obiektowi $o \in \mathcal{O}$, gdzie $\mathcal{O}$ jest kolekcją obiektów multimedialnych.

Dobór atrybutu efektywnie różnicującego obiekty może być niekiedy zadaniem niełatwym, mało intuicyjnym. Często rozważa się w takich przypadkach cały zestaw możliwych właściwości, przy czym dopiero ich połączenie pozwala uzyskać zadawalający opis kolekcji. Atrybut przyjmuje wtedy postać atrybutu złożonego, tj. wieloelementowego. Może to być atrybut listowy, zawierający zestaw cech tego samego typu - np. kolory dominujące, opisujący kilka kolorów przeważających lokalnie (przykładowo w blokach sztywnego podziału obrazu) lub statystycznie w skali całego obrazka (np. na podstawie histogramu kolorów). Lista cech kolorów dominujących skonstruowana jest wtedy na podstawie atrybutu podstawowego - kolor obrazu. Ten sam alfabet wartości koloru przypisywany jest poszczególnym elementom atrybutu listowego.

Każdy obiekt $o \in \mathcal{O}$ charakteryzowany jest wtedy za pomocą zestawu atrybutów $\{a_k\}_{k=1,\ldots,K}$, tj. właściwości istotnych np. ze względu na sposób użytkowania. W przypadku danego obiektu właściwości przyjmują określoną postać tworząc wektor cech obiektu: $\mathcal{C}_{\mathbf{a}} = [a_1(o), \ldots, a_K(o)] = [c_1, \ldots, c_K]$, gdzie zestaw cech pojedynczych $c_k = a_k(o)$ stanowi numeryczny deskryptor obiektu danej kolekcji, względem ustalonego zestawu atrybutów.

Na podstawie opisów właściwości poszczególnych elementów bazy tworzona jest struktura indeksu. **Indeks** konstruowany jest wokół ustalonego atrybutu (atrybut jest argumentem indeksu) jako zestaw (lista) cech przypisanych obiektom opisywanej kolekcji. Każda z cech odnosi się z kolei do listy wskaźników obiektów posiadających daną cechę. Dwa obiekty są podobne, jeśli posiadają daną cechę lub cechę "zbliżoną" (w przybliżeniu podobną). Przy opisie obiektów za pomocą zestawu atrybutów, indeks danej kolekcji budowany jest dla każdego rozpatrywanego atrybutu z osobna [27].

Przykładowo, jeśli atrybutem jest procentowa (np. z dokładnością do 5%)

zawartość scen z przemocą w filmach danej kolekcji, to kolejne pozycje indeksu obok precyzyjnych zakresów procentowych wartości zawierają bazodanowe identyfikatory filmów, zgodnie z określoną przez atrybut scen przemocy ich charakterystyką. Idąc dalej, indeks zasobów przykładowej wypożyczalni filmów może zostać zaprojektowany według zestawu atrybutów (atrybutu złożonego), składającego się dodatkowo z kategorii opisującej charakter filmu, jego treść, dozwolony przedział wiekowy potencjalnych widzów, gwiazdkową skalę atrakcyjności według ocen ekspertów oraz wskaźnik popularności wypożyczeń czy orientacyjny czas trwania filmu (z listą cech: krótki - czas trwania do godziny, średni - czas trwania od 1h do 2h, długi - 2h $<$ czas $\leqslant$ 3h , bardzo długi - czas $>$ 3h).

Zależnie od rodzaju atrybutu, a także oczekiwanej zdolności różnicowania obiektów opisanych indeksem można opracować zestaw mniej lub bardziej szczegółowych cech, które stają się cechami reprezentatywnymi. Przykładowo, lista cech atrybutu dominujący kolor obrazu może zawierać jedynie pozycję *zielony* lub też bardziej szczegółowe: *jasnozielony*, *zielony*, *ciemnozielony* czy *turkusowy*. W każdym z tych przypadków należy precyzyjnie określić przedziały wartości np. trzech składowych RGB, które odpowiadają poszczególnym barwom. Dwa obiekty są podobne w sensie dominującego koloru, jeśli ich barwa opisana liczbowo w przestrzeni RGB przyporządkowana zostaje tej samej cesze z listy atrybutu indeksu. Mamy w tym przypadku określenie subiektywnej cechy pojęciowej (opisanej tekstowo) za pomocą obliczeniowo obiektywnej cechy numerycznej, inaczej deskryptora numerycznego. Taka definicja nie musi być jednak jednoznaczna (niekoniecznie ustalone przedziały liczbowe zostaną uznane przez wszystkich użytkowników; subiektywne wrażenie barwy zależy niekiedy od kontekstu, więc odczytanie lokalnej barwy pikseli może być odmienne przy tych samych wartościach pikseli w przestrzeni RGB itp.).

Podobieństwo cech obiektów jest pojęciem bardzo istotnym przy konstrukcji indeksów oraz organizacji całej procedury wyszukiwania treści multimedialnej. Podobieństwo obiektów ze względu na określony atrybut dotyczy bliskości ich cech lub też jest funkcją odwrotną ich odległości (metryki). Podobieństwo w sensie metrycznym określane jest za pomocą znormalizowanej **funkcji podobieństwa** cech atrybutu a jako $\rho_a \colon \mathcal{C}_a \times \mathcal{C}_a \leftarrow [0,1]$. Dwa obiekty są więc bardziej podobne ze względu na określoną ich właściwość, jeśli $\rho_a(c1,c2) = \rho_a(a(o_1), a(o_2))$ jest bliższe wartości 1. Ustawiając sztywną wartość progową, np. $t = 0,9$ definiujemy jako podobne względem siebie te obiekty, dla których $\rho_a(c1,c2) \geqslant t$. Z kolei jeśli ze zbioru obiektów $\{o\}_{1,2\ldots}$ chcemy wybrać najbardziej podobny do o (w sensie określonego a), wtedy

$$o_{\text{naj\_pod}} = \arg \max_{\{o_i; i=1,2\ldots\}} \{\rho_a(a(o), a(o_i))\}$$

Przykładowo, jeśli cechami są liczby różniące się maksymalnie o M, wówczas

funkcję podobieństwa można ogólnie zdefiniować jako:

$$\rho(c1, c2) \triangleq 1 - \frac{|c1 - c2|}{M}$$

Ukonkretniając, jeśli cechami są dowolne punkty w kwadracie o boku 1, wtedy $\rho(c1, c2) \triangleq 1 - \frac{\|c1-c2\|}{\sqrt{2}}$, gdzie metryka $\|\cdot\|$ rozumiana jest w sensie euklidesowej odległości na płaszczyźnie dwóch wektorów wskazujących cechy obiektów.

Ogólniej, podobieństwo opisywane za pomocą odległości cech $\delta(c1, c2)$ można zapisać jako

$$\rho(c1, c2) \triangleq 1 - \frac{\delta(c1, c2)}{\max_c |c|} \tag{2.19}$$

Przy określaniu podobieństwa wyrazów czy ogólniej danych typu tekstowego użyteczne jest podobieństwo typu edycyjnego czy też rangowego (dotyczącego pozycji cechy w ustalonym porządku). Funkcję podobieństwa edycyjnego można wyznaczyć za pomocą (2.19), przy czym odległość $\delta(c1, c2)$ pomiędzy wyrazami (słowami, terminami) określana jest jako najmniejsza liczba operacji zmiany, usuwania i dołączania pojedynczego symbolu (litery) w dowolnym miejscu, dzięki którym sekwencja $c1$ przekształcana jest w sekwencję $c2$. Normalizująca wartość $\max_c |c|$ uwzględnia największą możliwą długość słowa w danym zbiorze cech.

Przykładowo, obliczając podobieństwo słów $c1 = Mama$ oraz $c2 = Matka$ minimalna liczba operacji przekształcenia $c1$ i $c2$ wynosi 3, bo aby przekształcić *Mama* w *Matka* wystarczy wykonać kolejno: USUŃ m,3 (usuń m na pozycji 3, licząc od 1); DODAJ t,3; DODAJ k,4 (symetrycznie trzy odwrotne operacje przekształcają *Matka* w *Mama*). Tak więc podobieństwo wynosi

$$\rho(Mama, Matka) = 1 - \frac{3}{25} = 0,88$$

Określając podobieństwo rangowe wykorzystuje się określenie pozycji cechy $r(c)$:

$$\rho(c1, c2) \triangleq 1 - \frac{|r(c1) - r(c2)|}{\max_c r(c) - \min_c r(c)} \tag{2.20}$$

Przykładem może być obliczenie rangowego podobieństwa planet naszego układu słonecznego. Przyjmując, że mamy 8 planet układu – od Merkurego po Neptuna, na podstawie odległości trzeciej Ziemi od piątego Jowisza możemy policzyć

$$\rho(Ziemia, Jowisz) = 1 - \frac{|3 - 5|}{8 - 1} \approxeq 0,714$$

Przy opisie terminów cyklicznych, jak np. miesięcy roku, podobieństwo rangowe należałoby zmodyfikować do postaci

$$\rho(c1, c2) \triangleq 1 - \frac{\min\Big(|r(c1) - r(c2)|, \max_c r(c) - |r(c1) - r(c2)|\Big)}{\max_c r(c) - \min_c r(c)} \tag{2.21}$$

Wtedy podobieństwo *czerwca* do *października* wynosi

$$\rho(\textit{czerwiec,październik}) = 1 - \frac{\min(|6-10|, 12-|6-10|)}{12-1} =\approx 0,636$$

zaś *stycznia* do *grudnia*

$$\rho(\textit{styczeń,grudzień}) = 1 - \frac{\min(|1-12|, 12-|1-12|)}{12-1} = 1 - \frac{\min(12,1|)}{11} \approx 0,91$$

W procesie indeksowania konieczne jest uwzględnienie zarówno manualnego, pojęciowego, jak i automatycznego mechanizmu wyznaczania wartości atrybutu dla danego obiektu z kolekcji. Proces **ekstrakcji cech** atrybutu a sprowadza się do wyznaczenia podzbioru wszystkich możliwych wartości a w postaci tzw. **cech reprezentatywnych**: $\mathcal{C}_a^\tau \subset \mathcal{C}_a$. Reprezentatywność cech oznacza istnienie funkcji reprezentacji $\tau_a : \mathcal{C}_a \to \mathcal{C}_a^\tau$, przypisującej dowolnej wartości atrybutu jego cechę reprezentatywną. Mają tutaj zastosowanie m.in. efektywne metody kwantyzacji danych. Przybliżenie danej cechy obiektu $c = a(o)$ za pomocą cechy reprezentatywnej $c^\tau = \tau_a(c) = \tau_a(a(o))$ powinno przynieść następujące efekty:

a) uprościć strukturę indeksu (zredukować zajętość pamięci przechowującej indeks, przyspieszyć operacje na indeksie),

b) uprościć i uczynić bardziej przejrzystym kryterium podobieństwa cech poprzez możliwe jednoznaczne odniesienie wartości liczbowych do reprezentowanej treści, zgodnie z oczekiwaniem użytkownika,

c) zachować, a nawet zwiększyć selektywność wyszukiwania poprzez precyzyjniejszy opis semantyczny.

Przykładowo, kształt guza w obrazach mammograficznych można opisać całym zestawem parametrów badających kolistość, relację długości obwodu do pola powierzchni, pole powierzchni odniesione do pola wpisanego prostokąta, gładkość konturu i jego symetryczność, itd. Znormalizowane wartości liczbowe tych parametrów można następnie zredukować usuwając nadmiarowość takiej reprezentacji (np. metodą analizy składowych głównych PCA). Zredukowaną liczbę tak uzyskanych parametrów można ustawić w wektor liczbowy, kwantowany następnie do kilkunastu możliwych postaci wektora reprezentującego guzy. Schemat kwantyzacji można zaprojektować odnosząc się do analogicznej charakterystyki wzorców guzów złośliwych i łagodnych zapewniając, by dobrane poziomy kwantyzacji dawały największe zróżnicowanie przypadków zdrowych i chorobowych.

Cechami reprezentatywnymi atrybutów o wartościach typu tekstowego mogą być np. rdzenie wyrazów, np. *matur*, będący rdzeniem słów *przedmaturalny*, *pomaturalny*, *maturą*, *maturze*, *matury* itp.

Deskryptorem jest słowo, fraza, znaki alfanumeryczne, zestawy liczb lub też metoda czy algorytm służące charakterystyce czy wręcz identyfikacji obiektów (składników treści) sygnałów naturalnych w systemach gromadzenia i przeszukiwania informacji. Deskryptor dotyczy określonego atrybutu, czyli wybranej właściwości obiektów. W konwencji standardu MPEG-7 jest numerycznym sposobem opisu atrybutów pojęciowych, czy też realizatorem numerycznego opisu danego atrybutu. Są to meta-dane wyznaczane automatycznie na bazie sygnału cyfrowego przenoszącego treść multimedialną.

W przypadku deskryptorów numerycznych możliwe jest dokładniejsze różnicowanie podobieństwa obiektów z listy danej cechy pojęciowej. Przykładowo, euklidesowa odległość [11] wektorów opisujących obiekty w przestrzeni RGB i definiujących *de facto* ich kolor pozwala precyzyjnie ustalić podobieństwo cech - mniejsza odległość wskazuje na większe podobieństwo koloru obiektów. Możliwe jest wtedy zbudowanie indeksu wartości numerycznych deskryptorów koloru dominującego, z przypisaną dodatkowo kategorią cechy pojęciowej. Algorytm pozwala ustalić precyzyjnie kolor dominujący danego obrazu czy regionu, który zostaje skwantowany zgodnie z przedziałami pojęciowego opisu barw (tj. *zielony*, *niebieski* itd.). Mamy wtedy do czynienia z indeksem tekstowo-numerycznym, który może być przeszukiwany z kryterium bliskości cech numerycznych, jak też identyczności cech pojęciowych (np. *zielony* = *zielony*).

Proces znakowania treści określonych zasobów danych z wykorzystaniem struktury indeksu nazywamy **indeksowaniem**. Indeksowanie multimediów dotyczy metod konstrukcji indeksów kolekcji (zbiorów) obiektów multimedialnych. Ze względu na bogactwo zawartej informacji, szczególnie interesującym zagadnieniem jest indeksowanie obrazów. Stosowną od kilkudziesięciu lat praktyką syntetycznej charakterystyki treści obrazów jest manualne tworzenie opisów alfanumerycznych. Do ich przeszukiwania wykorzystywano zwykle silnik bazodanowy DBMS (*Database Management System*) [25]. Opracowano wiele technik związanych w oceną efektywności zapytań, strukturami danych, metodami przeszukiwania i przechowywania indeksów.

Opis schematu znakowania i wyszukania treści

Dla każdego rozważanego atrybutu a możemy mówić o indeksie danej kolekcji obiektów $\mathcal{O}$. Indeks zawiera listę kolejnych cech reprezentatywnych c^τ, przy czym każdej z nich przypisana jest lista $\iota_1, \iota_2, \ldots$ (w uproszczonej postaci) identyfikatorów $\iota(o)$ wszystkich obiektów $o \in \mathcal{O}$ o atrybucie reprezentowanych przez $\tau_a(a(o)) = c^\tau$. Lista l_{c^τ} obiektów podobnych (skrótowo: lista obiektowa) w sensie właściwości a i przyjętej ρ_a wygląda następująco: $\mathcal{L}_{c^\tau} = \{l_{c^\tau}; \iota_1, \iota_2, \ldots, \iota_{l_{c^\tau}}\}$.

[11] Odległość definiowana zgodnie z metryką euklidesową:
$||x - y|| = \sqrt{(R_x - R_y)^2 + (G_x - G_y)^2 + (B_x - B_y)^2}$ w przestrzeni kolorów RGB.

Liczba obiektów l_{c^τ} na liście zależy od przyjętego kryterium podobieństwa, przy czym zwykle przyjmowana jest progowa definicja zbioru obiektów podobnych jako $\mathcal{O}_{c^\tau} = \{o \in \mathcal{O}\colon \rho(a(o), c^\tau) \geqslant \rho_{min}\}$, z minimalnym progiem podobieństwa $t = \rho_{min} \in [0, 1]$. Wtedy $l_{c^\tau} = |\mathcal{O}_{c^\tau}|$. Dodatkowo, liczba obiektów podobnych identyfikowanych na liście ograniczana jest za pomocą zadanych granic jej liczności, tj. $L_{min} \leqslant l_{c^\tau} \leqslant L_{max}$. Przy większej liczbie obiektów podobnych względem ρ_{min} wskazywanych jest jedynie L_{max} najbardziej podobnych. Natomiast przy braku wystarczającej liczby obiektów podobnych względem ρ_{min}, dopisywane są kolejne najbardziej podobne obiekty kosztem obniżenia minimalnego progu podobieństwa. Założona liczność list obiektowych może wynikać z przewidywanego mechanizmu wyszukiwania, kiedy to sztywna liczba zwracanych obiektów podobnych wynosi L_{max}, natomiast nie może być ona mniejsza niż L_{min}. Często jednak realizowane procedury wyszukiwania wykorzystują bardziej różnorodne, adaptacyjne formy odpowiedzi na zapytania.

Wyszukiwanie bazuje na odpowiednio przygotowanym indeksie. Typowy, możliwie ogólny scenariusz wyszukiwania jest następujący:

- założenia wstępne: wyszukiwanie bazuje na indeksie atrybutu a kolekcji multimediów $\mathcal{O}$ i polega na sformułowaniu odpowiedzi na zapytanie przy ustalonych uwarunkowaniach – np. zwracając co najmniej K_{min} i co najwyżej K_{max} obiektów najbardziej podobnych, tj. posortowanych według wartości funkcji podobieństwa do cechy zapytania c_{query}, przekraczających próg ρ_{min}; kolejne działania procedury wyszukiwania bazują na przygotowanej strukturze indeksu, odwołując się do listy cech, list obiektowych z identyfikatorami obiektów przypisanych danej cesze reprezentatywnej czy też do algorytmów liczących cechy obiektów i porównujących je;

- sformułowanie zapytania w postaci cechy $c_{query} \in C_a$, określonej przez użytkownika za pomocą odpowiednio przygotowanego interfejsu lub poprzez algorytm ekstrakcji cech atrybutu a z obiektu o_{query} stanowiącego zapytanie przez przykład;

- wyszukanie najbardziej podobnych obiektów poprzez znalezienie w zbiorze cech reprezentatywnych atrybutu a cech reprezentatywnych c^τ spełniających warunek podobieństwa w stosunku do c_{query}; w przypadku definicji progowej mamy

$$\rho(c_{query}, c^\tau) \geqslant \rho_{min} \tag{2.22}$$

uzyskując ciąg $c_1^\tau, \cdots, c_L^\tau$ cech podobnych ze wspólną listą obiektową

$$\mathcal{L}_{c_{query}} = \mathcal{L}_{c_1^\tau} \cup \mathcal{L}_{c_2^\tau} \cup \cdots \cup \mathcal{L}_{c_L^\tau}$$

dobór wielkości ρ_{min} powinien być podyktowany względami merytorycznymi, zależnie od rodzaju danych oraz celów wyszukiwania, przy czym możli-

wy jest interaktywny dobór tego parametru przez użytkownika; w przypadku kryterium jedynie ilościowego realizowany jest schemat z pożądaną liczbą obiektów wskazywanych przez $\mathcal{L}_{c_{query}}$, przykładowo poprzez posortowanie cech reprezentatywnych względem ich malejącego podobieństwa do c_{query}, a następnie dołączanie do tworzonej wspólnej listy obiektowej w pierwszej kolejności obiektów z list cech reprezentatywnych najbliższych c_{query}, aż do uzyskania założonej liczby $L_{max}^{c_{query}}$ obiektów;

- sformułowanie odpowiedzi na zapytanie w postaci zestawu obiektów najbardziej podobnych, tj. obiektów identyfikowanych przez połączoną listę $\mathcal{L}_{c_{query}}$; w przypadku ograniczenia odpowiedzi do K_{max} obiektów podobnych, możliwe jest dodatkowe posortowanie obiektów z $\mathcal{L}_{c_{query}}$ według obliczonej dokładnie wartości podobieństwa $\rho(c_{query}, a(o))$ ($c(o)$ mogą być przechowywane w słowniku); spełnienie warunku K_{min} dla trudnych zapytań można uzyskać poprzez obniżenie wartości progu ρ_{min}.

W praktyce realizowanych jest kilka typowych uwarunkowań wyszukiwania, dotyczących:

- ustalonej liczby K obiektów najbardziej podobnych do zapytania, gdzie $K_{min} = K_{max} = K > 0$, niezależnie od stopnia podobieństwa cech obiektów ($\rho_{min} = 0$);
- wszystkich obiektów istotnie podobnych, tj. spełniających ustalone kryterium podobieństwa – np. z progiem $\rho_{min} > 0$; wtedy $K_{min} = 0, K_{max} = \infty$;
- ograniczonej liczby obiektów istotnie podobnych, co jest logicznym połączeniem dwóch powyższych kategorii; warunkiem koniecznym odpowiedzi jest spełnione kryterium podobieństwa (określone np. progiem $\rho_{min} > 0$), przy ograniczeniu liczby obiektów stanowiących odpowiedź jedynie do K najistotniejszych, czyli $K_{min} = 0, K_{max} = K$;
- przynajmniej K obiektów podobnych, gdzie $K_{min} = K > 0$, zaś $K_{max} = \infty$ i $\rho_{min} > 0$; zapewnienie minimalnej liczby obiektów podobnych odbywa się niekiedy kosztem złagodzenia kryterium podobieństwa.

Możliwe są też inne kombinacje podstawowych parametrów definiujących warunki wyszukiwania, możliwa jest adaptacja reguł zapytania do potrzeb użytkownika, np. poprzez wprowadzenia mechanizmu doboru relacji pomiędzy K i ρ_{min} na podstawie liczności ustalonej $\mathcal{L}_{c_{query}}$.

W realizacji powyższego scenariusza użyteczne są następujące struktury danych realizujące indeks [27]:

- słownikowa *DictionaryOfFeatures* cech reprezentatywnych atrybutu, zwykle w pamięci operacyjnej; w kolejnych pozycjach słownika obok c^τ umieszczane są adresy przypisanych im list obiektowych $\iota_{\mathcal{L}_{c^\tau}}$;

- tablicowa *CollectionOfLists* listy obiektowych, w pamięci operacyjnej lub dyskowej, zawierająca poszczególnych cech zapisane w reprezentacji wygodnej do szybkich odwołań, niekiedy kodowanej; każda lista wskazywana jest przez swój adres $\iota_{\mathcal{L}_{c^\tau}}$ liczony względem początku tablicy;
- słownikowa *DictionaryOfObjects* identyfikatorów wszystkich obiektów przeszukiwanej kolekcji $\mathcal{O}$, w pamięci operacyjnej lub dyskowej; kolejne frazy słownika identyfikowane przez $\iota(o)$ z list obiektowych zawierają referencje zapewniające dostęp do zawartości obiektu multimedialnego; opcjonalnie element słownika zawiera dodatkowo cechę obiektu $c = a(o)$ lub referencję dającą do niej dostęp.

Dobór atrybutów zawartości

Ekstrakcja i reprezentacja cech obiektów jest pierwszym etapem projektowania systemu wyszukiwania treści. Skuteczność indeksowania zawartości, służącego realizacji różnych schematów wyszukiwania treści istotnej z bazy obiektów multimedialnych zależy w pierwszej kolejności od doboru zestawu atrybutów opisujących treść w sposób specyficzny, możliwie kompletny (wieloaspektowy, hierarchiczny do poziomu istotnych szczegółów), a przy tym różnicujący ze względu na odmienne kategorie opisywanej treści. Metoda konstrukcji efektywnych w danym zastosowaniu atrybutów powinna wykorzystywać przede wszystkim: a) całą dostępną *a priori* wiedzę dziedzinową, b) rzetelną charakterystykę jakościową opisywanych danych, c) wiarygodne profile użytkownika, w tym możliwie zupełny zbiór przewidywanych celów wyszukiwania (można na tej podstawie zróżnicować także formę zapytań). Zwykle sposób rozwiązania problemu doboru atrybutów, czyli sposobu skutecznego opisu obiektów kolekcji ma zasadniczy wpływ na ostateczny kształt mechanizmu indeksującego oraz schemat wyszukiwania.

Najczęściej poszukiwane cechy atrybutów i deskryptorów realizujących numeryczny opis danego atrybutu to:

- precyzyjne i możliwie kompletne różnicowanie treści obiektów;
- pozwalające na proste i jednoznaczne określenie podobieństwa (odległości) cech obiektów;
- dające możliwie zwarty opis (upakowane w sensie stosunku zakresu opisywanych cech do wymiaru deskryptora);
- postać znormalizowana oraz niezmienniczość względem warunków akwizycji, przekształceń afinicznych itp.;
- pozwalające na oszczędne obliczeniowo implementacje,
- możliwie duża zgodność z intencjami użytkownika (dające semantycznie poprawną charakterystykę obiektów).

Wyszukując treść podobną rodzi się zasadnicze pytanie o sposób opisu, a następnie określenia stopnia podobieństwa treści zawartej w zbiorach danych. Jest to pytanie o semantykę danych pojedynczych, grup danych łączonych według określonego kryterium przynależności do obiektu, czy też zbioru obiektów o określonych cechach i wzajemnych relacjach. Jak opisując dane uzyskać wiarygodny wykładnik treści? Jak takie semantyczne deskryptory porównywać ze sobą, by wskazać obiekty podobne w rozumieniu treści zgodnym z intencjami użytkownika? Odpowiedzi na te pytania nie są proste, chociaż najlepsze odpowiedzi wcale nie muszą być bardzo złożone i skomplikowane.

Ponieważ atrybuty zawartości obiektów i przyporządkowane ich deskryptory służą realizacji zapytań formułowanych przez użytkownika, przy ich wyborze czy projektowaniu warto zdać sobie sprawę w przypuszczalnych intencji pytającego. W przypadku obrazów mogą one dotyczyć m.in.:

- obiektów prostych, o określonej kombinacji cech podstawowych, takich jak kolor, tekstura czy kształt – np. znajdź obrazy zawierające prostokątne, białe tablice z napisami;
- specyficznych typów obiektów lub grupy obiektów w obrazie, np. samochodu danej marki, logo stacji telewizyjnej, czy też zestawu kanapy z fotelem;
- identyfikacji specyficznego obiektu, np. określeniu tożsamości osoby na zdjęciu czy rozpoznaniu cech szczególnych danej osoby, kategorii przynależności do określonej grupy;
- określonego zdarzenia, np. koncertu na molo, meczu futbolu amerykańskiego czy zdjęć ukazujących lwy polujące na bawoły;
- szczegółów określonego zdarzenia, dotyczących obecności danej osoby, zwierzęcia czy przedmiotu, np. meczu piłkarskiego drużyny polskiej, spotkania Jarosława Kaczyńskiego z wyborcami czy galerii wystawiających obraz *Mona Lisa*;
- subiektywnych emocji towarzyszących jakiemuś wydarzeniu czy ogólnie rejestrowanej scenie, np. agresji na spotkaniach z czytelnikami, szczęśliwego wyrazu twarzy; przykład – znajdź obrazy wyrażające ludzkie cierpienie;
- właściwej interpretacji zapytania, np. w przypadku radiologicznego opisu przypadku zobrazowania trudnego w ocenie diagnostycznej;
- metadanych związanych z danym obrazem, np. dotyczących autora zdjęcia, daty powstania, miejsca, nazwisk osób zobrazowanych itp.; w tym przypadku intencją użytkownika jest nałożenie określonych ograniczeń zapytaniu o zawartość opisaną deskryptorem numerycznym.

Zarówno spodziewane zapytania, jak też dobierane cechy opisu obiektów kolekcji mogą być konstruowane na różnym poziomie abstrakcji [30]. Zasadniczo można wyróżnić poziom:

- wizualnych cech podstawowych (*primitive*), gdzie rozważa się np. takie atrybuty obrazów jak kolor, kształt, tekstura, lokalizacja, a podobieństwo dotyczy tej samej kolorystyki czy tego samego kształtu, zwykle prostego prymitywu geometrycznego, bez odwołań do specjalistycznej wiedzy dziedzinowej; stosowane jest niekiedy porównanie obrazów na podstawie takiego podobieństwa użytkownik może jakby przy okazji odkryć pewne relacje treściowe – gdy np. kolor jest decydującym wyróżnikiem szukanej treści; większą skuteczność tego rodzaju opisu uzyskuje się zwykle poprzez konstruowanie coraz dokładniejszych, adaptacyjnych i lokalnych w opisie deskryptorów oraz poprzez łączenie deskryptorów kilku atrybutów cech podstawowych w jeden złożony opis obiektów;

- identyfikacji obiektów złożonych, gdzie podobieństwo oznacza logiczną przynależność do określonej klasy, kategorii czy rodzaju danych (trzeba więc uwzględnić wiedzę specjalistyczną); zależności pomiędzy numerycznymi deskryptorami atrybutów bardziej złożonych a semantyką opisu obiektów poszukiwane są tutaj już na poziomie konstrukcji indeksu; przykładem może być deskryptor twarzy służący wyszukaniu obrazów będących zdjęciami ludzkich twarzy; im bardziej abstrakcyjny obiekt, tym konstrukcja efektywnego deskryptora trudniejsza; użytkownik może się zadowolić stopniem identyfikacji w właściwej danemu zastosowaniu hierarchii treści, może też stwierdzić nieskuteczność wyszukania pożądanej, stojącej na wyższym poziomie abstrakcji klasy obiektów; wykorzystywane są tutaj niekiedy różne formy interakcji z użytkownikiem, który weryfikując poprawność odpowiedzi pozwala doprecyzować adaptacyjne algorytmy deskryptorów identyfikujących obiekty;

- rozpoznania specyficznych obiektów abstrakcyjnych w kontekście ich pojawienia się, określonych zdarzeń czy stanów emocjonalnych; w opisie uwzględnia się specjalistyczną treść danych, znaczenie obiektów i wzajemnych relacji; występuje tutaj podobieństwo w sensie wysublimowanej semantyki, wynikającej z kontekstu wiedzy dziedzinowej; stosowane są tutaj tzw. deskryptory semantyczne, konstruowane pod kątem określonej semantyki atrybutów opisu obiektów – opisywane cechy te mają często bardzo odmienny charakter od cech wizualnych; w takich zastosowaniach niekiedy użytkownik nawet nie jest w stanie od razu zweryfikować poprawności odpowiedzi wyszukiwarki – potrzebny jest do tego dodatkowa weryfikacja semantycznego podobieństwa obiektów odpowiedzi do zapytania; przykładem może być wspomaganie obrazowej diagnostyki medycznej za pomocą CBIR – za-

pytaniem jest wtedy trudny w opisie diagnostycznym obraz z podejrzeniem patologii, zaś referencyjne obrazy potwierdzonych klinicznie przypadków stanowiące odpowiedź stanowią sugestię interpretacji obrazu zapytania.

Zdecydowana większość komercyjnych systemów CBIR wykorzystuje jedynie podstawowy poziom, konstruując indeksy atrybutów prostych, ze zwykle łatwą weryfikacją poprawności odpowiedzi (przykładowo Blobworld [36], AltaVista Photofinder, Amor, Berkeley Digital Library Project, Blobworld i in. [37]). Pfund i Marchand-Maillet [30] wykorzystali dodatkowo metadane alfanumeryczne pochodzących z ręcznego opisu przez operatora. Warte podkreślenia są jednak liczne prace badawcze zmierzające do opracowania CBIR na poziomie identyfikacji, a nawet rozpoznania specyficznych obiektów abstrakcyjnych [31, 32, 33, 34].

Konieczność doskonalenia deskryptorów semantycznych wynika z trzech zasadniczych problemów: luki semantycznej, polisemii i bariery sensorycznej. Pierwszym problemem ograniczającym skuteczność opisu obiektów za pomocą projektowanych sposobów liczenia cech jest **luka semantyczna** *(semantic gap)* [38]. Polega to na braku zgodności cech numerycznych, automatycznie ekstrahowanych z obrazu z cechami, które użytkownik uznaje za znaczące w opisywanych obrazach, zależnie od kontekstu ich wykorzystania. Nie znając intencji użytkownika bardzo trudno jest przewidzieć jego oczekiwania, im bardziej specjalistyczne jest zastosowanie, tym kontekst znaczenia wyszukiwanych obiektów jest łatwiejszy do przewidzenia i deskryptory mogą być skuteczniejsze.

Luka semantyczna występuje zwykle pomiędzy podstawowym a wyższymi poziomami abstrakcji opisu obiektów multimedialnych, w tym przypadku obrazów. Stosowane deskryptory nie odzwierciedlają właściwie treści obrazowej powodując formułowanie niesatysfakcjonujących użytkownika odpowiedzi – zobacz przykład na rys. 2.10. Na rys. 2.11 pokazano sytuacje odwrotną, kiedy to podobne treściowo obrazy nie są do siebie mało podobne lub wręcz niepodobne w sensie podstawowych cech obrazowych (kolorystyka, tekstury, nawet kształt).

Ważnym powodem luki semantycznej jest **polisemia**, czyli wieloznaczność wyrazu treści obrazowej. Na rys. 2.12 ukazano trzy różne poziomy wieloznaczności obrazów. W przypadku wieloznacznej treści obrazowej trudno jest ustalić, jakie są intencje pytającego, na jakiej płaszczyźnie znaczeniowej spodziewana jest odpowiedź. Sposób formułowania zapytania, np. w formie interaktywnej, może zawierać mechanizmy precyzujące sposób interpretacji treści przez użytkownika, z określeniem rodzaju istotnej semantyki opisu obrazów. Zwykle jesteśmy w stanie dokonać tego jedynie w ograniczonym zakresie [40].

Dobrym przykładem problemu polisemii jest ustalenie podobieństwa do zapytania w zastosowaniach medycznej diagnostyki obrazowej. Niekiedy może być ono rozumiane jedynie w kategorii znaczeniowej tej samej modalności zobrazowania, np. każdy obraz ultrasonograficzny (USG) jest podobny do innego obrazu USG. Zwykle jednak problem jest definiowany bardziej precyzyjnie – chodzi o ten sam

Rysunek 2.10: Luka semantyczna procedur wyszukiwania – przykład zaczerpnięty z [39]; dwa obrazy pomimo dzielących je, oczywistych różnic treściowych zostały mylnie określone przez CBIR jako podobne; podobieństwo koloru i do pewnego stopnia kształtu oraz rozmiaru wielu drobnych obiektów nie przekłada się w tym przypadku na wspólną semantykę.

Rysunek 2.11: Ograniczenia skuteczności CBIR powodowane trudnym do opisu numerycznego podobieństwem prezentowanych obrazów – przykład zaczerpnięty z [39]; dwa podobne znaczeniowo obrazy mają wyraźnie odmienne wizualne cechy podstawowe.

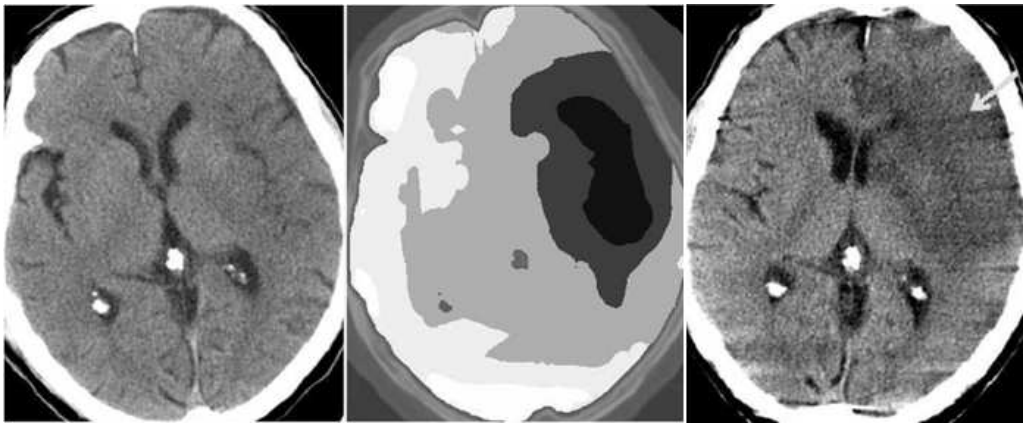
Rysunek 2.12: Wieloznaczność treści obrazów – przykład zaczerpnięty z [39]); od lewej – obraz wieloznaczny (ludzie, różne rasy, biegnący ludzie, przyglądający się ludzie, zawody sportowe, barwy różnych krajów, olimpiada, doping, wysiłek), obraz o ograniczonej wieloznaczności (maszerujący ludzie, wspinaczka górska, krajobraz), obraz dość jednoznaczny (kwiaty).

rodzaj badania, np. badanie echokardiografii, czyli USG serca. We wspomaganiu diagnostyki obrazowej istotny jest "bardziej wymagający rodzaj podobieństwa", definiowany przez obecność patologii, patologii o podobnych cechach lub nawet ten sam rodzaj patologii. Przykładowo, na rys. 2.13 pokazano podobne obrazy rentgenowskie płuc w sensie modalności i rodzaju badania, ale różne ze względu na obecność patologii. Im subtelniejszy, bardziej szczegółowy zakres podobieństwa, tym trudniej opracować semantyczne deskryptory pożądanych atrybutów, które pozwolą właściwie określić podobieństwo treści.

Rysunek 2.13: Wieloznaczność obrazów medycznych – obrazy podobne w sensie modalności i rodzaju zobrazowania (rentgen płuc), jednak różne w interpretacji diagnostycznej, po lewej obraz bez patologii, po prawej z guzem nowotworowym (zaczerpnięty z bazy JSRT - *Japanese Society of Radiological Technology database* - utworzonej przez Japońskie Towarzystwo Radiologiczne w 1998: http://www.jsrt.or.jp/web_data/english03.php).

W kontekście zastosowań medycznych pojawia się dodatkowy problem tzw. *bariery postrzegania* (*sensory gap*). Jest to związane z występowaniem informacji ukrytej, kiedy to cechy wizualne określonej struktury czy ogólniej obiektu są niewystarczająco wyraźne, różnicujące w stosunku do kontekstu występowania, ogólnie tła, by obiekt mógł być dostrzeżony przez obserwatora (w przypadku diagnostyki medycznej, zwykle radiologa). Na rys. 2.14 przedstawiono efekt ekstrakcji cech ukrytego symptomu choroby udaru niedokrwiennego (symptom uwidoczniono dzięki obróbce numerycznej) na podstawie wczesnego badania tomografii komputerowej mózgu. Zdolność numerycznej identyfikacji cech ukrytych, bez szans percepcji przez radiologa, ma w tym przypadku szczególnie istotne znaczenie, gdyż jedynie we wczesnej fazie udaru możliwe jest przeprowadzenie skutecznej terapii trombolitycznej, ratującej życie lub chroniącej przed trwałym inwalidztwem. Efektywne wyszukanie referencyjnych przypadków podobnych me-

todą CBIR pozwola zwiększyć szansę trafnej diagnozy [41].



Rysunek 2.14: Bariera postrzegania w obrazach medycznych na przykładzie obrazowania tomografii komputerowej (TK) wczesnego udaru mózgu; od lewej kolejno a) badanie wczesne, tzn. wykonane w czasie do 3 godzin od wystąpienia incydentu udarowego, z niewidocznym obszarem hipodensyjnym, czyli bezpośrednim symptomem udaru niedokrwiennego [41]; b) ten sam obraz w formie przetworzonej, z ekstrakcją ukrytego obszaru hipodensyjnego (ciemna plama); c) rezultat późniejszego badania TK, wykonanego temu samemu pacjentowi po kilkunastu godzinach, z widocznym już obszarem hipodensyjnym (ciemniejszym, wskazanym strzałką), potwierdzającym wcześniejsze wskazanie specjalizowanego deskryptora semantycznego obszarów hipodensyjnych.

Powodem ograniczeń zdolności postrzegania mogą być realne uwarunkowania procesu akwizycji obrazów – zbyt mała czułość metody rejestracji w stosunku do specyfiki występującej zmiany (np. wczesnej fazy nowotworu czy incydentu udarowego). Poprawna interpretacji obrazu, a przez to stanu zdrowia pacjenta jest wtedy zagrożona, a przy braku innych symptomów, błędna. Zdarza się też tak, że cechy wizualne są powyżej bariery postrzegania, ale mają charakter niejednoznaczny ze względu na niską jakość zobrazowania zmiany. Cechy wizualne symptomów patologii są bowiem często z natury bardzo subtelne, względne, bez ustalonego, stabilnego wzorca, a ich błędna interpretacja może być tragiczna w skutkach.

Receptą jest wstępna poprawa jakości obrazów zmierzająca ku obniżeniu bariery postrzegania – w wielu przypadkach przynosi ona pozytywne skutki poprawy skuteczności diagnozy.

Istnienie bariery postrzegania sugeruje opis obrazu za pomocą cech nie tylko wizualnych. Zdolności obliczeniowe komputera w zakresie ekstrakcji cech natury statystycznej, lokalnej transformacji redukującej nadmiarowość źródłowej przestrzeni obrazowej, określania nieliniowych zależności pomiędzy pikselami czy grupami pikseli, aproksymacji istoty sygnału, reprezentacji za pomocą dobra-

nych atomów przestrzeni czas-częstotliwość itd., dają możliwość projektowania deskryptorów cech uwzględniających inne właściwości obiektów, niedostrzegalne przez człowieka. Wyznaczenie takich form opisu wymaga jednak często złożonych procedur optymalizacyjnych, uczenia za pomocą reprezentatywnych zbiorów treningowych, rozwiązań z zakresu sztucznej inteligencji itd. Okupione dużym kosztem obliczeniowym, szczególnie na etapie wyznaczania efektywnych form deskryptorów, przyczyniają się jednak do istotnej redukcji nie tylko bariery postrzegania, ale też luki semantycznej.

Przykłady prostych deskryptorów wybranych atrybutów

Poniżej przedstawiono kilka prostych zasad tworzenia podstawowych atrybutów obrazowych.

Opis globalny i lokalny Poszczególne deskryptory mogą być wyznaczane na podstawie całego obrazu (opis globalny), bądź też lokalnie, na podstawie regionu wybranego w sposób arbitralny, np. poprzez narzuconą siatkę blokowego podziału obrazu [38], adaptacyjne – np. poprzez automatyczną segmentację obiektów zainteresowania [44] lub też interaktywnie, poprzez wybór użytkownika [42, 43].

Opis koloru Charakterystyka kolorów treści obrazowej związana jest w pierwszej kolejności z wyborem przestrzeni barw. Chociaż obrazy opisywane są zwykle w przestrzeni RGB (składowe: czerwony, zielony, niebieski), przede wszystkim ze względu na budowę ludzkiego oka (trzy rodzaje czopków o odmiennej charakterystyce widmowej, gdzie maksima przypadają na te trzy kolory) oraz działanie typowych urządzeń służących rejestracji i prezentacji treści obrazowej (kamery, aparaty fotograficzne, monitory, telewizory, skanery), to model percepcji obrazów naturalnych opisujący podobieństwo kolorów wygodniej jest tworzyć w innych przestrzeniach barw.

Przestrzeń RGB nie pozwala na równomierny opis zdolności percepcji zmian barwowych – równym odległościom w RGB nie odpowiadają równomierne zmiany obserwowanych kolorów. Ponadto składowe są od siebie silnie zależne, wpływ na ich wartość ma natężenie oświetlenia oraz inne, oprócz barwowych czynniki. Reprezentacja RGB we współrzędnych biegunowych z uwzględnieniem korekcji gamma to przestrzeń barwna (odcień, nasycenie, jasność). Dobre wyniki w wykorzystaniu RGB lub częściej HSV można osiągnąć jedynie przy opisie obrazów grafiki komputerowej, gdy obrazy są definiowane czy projektowane w tej przestrzeni (np. przy indeksowaniu znaków handlowych) lub też jeśli akwizycja opisywanych obrazów jest dokonywana w ustalonych, stabilnych warunkach (np. w bazie danych obrazowych dotyczących malarskich dzieł sztuki).

Przy określaniu podobieństwa kolorów użyteczna stała się konwersja RGB do innych, lepiej opisujących ludzkie zdolności percepcji przestrzeni barw. Ko-

rzystne – szczególnie takich, w których jedna składowa opisuje poziom jasności (sygnał luminancji), a dwie pozostałe dotyczą cech koloru. Warto tu wspomnieć o przestrzeniach takich jak historyczne XYZ (z 1931 roku, niezależna, pozwalające określać kolory w sposób bezwzględny), YIQ (wykorzystywany w telewizji systemu NTSC), YCrCb (częsty w systemach kompresji obrazów i wideo rodzin JPEG oraz MPEG), YUV (wykorzystywany w telewizji systemu PAL), CIE Luv i CIE Lab (z nieliniowymi modelami dającymi równomierny rozkład percepcji barw, bazując na modelu Munsella [46]) i inne. Różnice barw liczone w tych przestrzeniach lepiej wyrażają obserwowane zróżnicowanie barw, w sposób bardziej niezależny od uwarunkowań procesu akwizycji [47, 48, 49, 50]. Możliwa jest optymalizacja przestrzeni kolorów wiernie odzwierciedlającej warunki oświetlenia czy też niezmienniczej względem cieni [51, 52]. Efektem jest identyfikacja barwy teoretycznie zupełnie niezależnie od warunków obserwacji, jednak kosztem utraty precyzji w szacowaniu wartości barwy, co może mieć wpływ na efektywność wyszukiwania.

Warto też nadmienić, że istniej wiele definicji określających poszczególne przestrzenie barw, np. CIE RGB, Adobe RGB, NTSC RGB, itd., przy czym szczególną pozycję zajmują normy międzynarodowego komitetu CIE[12].

Przykładowo, wzajemne konwersje wybranych przestrzeni barw opisują proste zależności:

- $RGB \rightarrow YCrCb$, według CIE, zakładając $R, G, B, Y \in [0, 1]$, $Cr, Cb \in [-0.5, 0.5]$

$$\begin{aligned} Y &= 0,2989 \cdot R + 0,5866 \cdot G + 0,1145 \cdot B \\ Cr &= 0,5 \cdot R - 0,4183 \cdot G - 0,0816 \cdot B \\ Cb &= -0,1687 \cdot R - 0,3312 \cdot G + 0,5 \cdot B \end{aligned} \tag{2.23}$$

- $RGB \rightarrow XYZ$, według Adobe RGB, zakładając $R, G, B, Y, X, Z \in [0, 1]$

$$\begin{aligned} Y &= 0,2974 \cdot R + 0,6273 \cdot G + 0,0753 \cdot B \\ X &= 0,5767 \cdot R + 0,1856 \cdot G + 0,1882 \cdot B \\ Z &= 0,0270 \cdot R + 0,0707 \cdot G + 0,9911 \cdot B \end{aligned} \tag{2.24}$$

Cały zestaw możliwych konwersji przestrzeni kolorów jest dostępny na stronie http://brucelindbloom.com/index.html?Math.html.

Podstawowym deskryptorem koloru jest **histogram** skwantowanych wartości składowych danej przestrzeni barw, np. HSV. W przypadku obrazów monochromatycznych, deskryptor koloru analogicznie opisuje rozkład poziomów jasności. Liczba przedziałów kwantyzacji histogramu powinna być kompromisem pomiędzy złożonością deskryptora, a jego reprezentatywnością, zależy też silnie od rodzaju

[12] International Commission on Illumination (CIE).

obrazów. Jeśli gros informacji zawarta jest w przedziale barw jasnych, wówczas sensowne wydaje się zwiększenie liczby przedziałów kwantyzacji je reprezentujących, kosztem mniej istotnych barw ciemnych.

Niestety, skwantowany histogram nie ma cechy niezmienniczości względem niewielkich przesunięć średniego poziomu jasności obrazu. Takie przesunięcie może zmienić przypisanie wartości pikseli leżących na granicy przedziałów kwantyzacji, wpływając niekiedy znacząco na rozkład histogramu. Wadę tę można ograniczyć poprzez zastosowanie tzw. histogramu rozmytego zaproponowanego w [54]. Koncepcja histogramu rozmytego usuwa nieciągłość przypisania wartości pikseli do przedziałów histogramu. Pozwala uprościć histogram za pomocą zachodzących na siebie przedziałów kwantyzacji, rozumianych tutaj jako nośniki zbiorów rozmytych z określoną funkcją przynależności (twórcą teorii zbiorów rozmytych jest Lotfi A. Zadeh [55]).

Rozważmy histogram (więcej o histogramie w p. 3.1.3) oraz jego skwantowaną, uproszczoną postać, pozwalającą konstruować deskryptory koloru. Przyjmijmy, że każda z wartości pikseli $f(k)$ (dla uproszczenia ustawionych w ciąg jednowymiarowy) obrazu źródłowego $\mathbf{f}$ należy do uporządkowanego rosnąco zbioru (alfabetu) wartości możliwych: $f(k) \in A_{\mathbf{f}} = \{a_0, \ldots, a_{M-1}\}$. Histogram, czyli rozkład koloru to zbiór $\{h(m)\}_{m=0}^{M-1}$, gdzie $h(m) = h(a_m)$ oznacza liczbę wystąpień (w punktach k dziedziny obrazu) kolejnych poziomów jasności

$$h(m) = \#\{k | f(k) = a_m\} \tag{2.25}$$

Pojęcie histogramu rozmytego nawiązuje do koncepcji kolorów reprezentatywnych, typowych dla obiektów czy tła występujących w danej klasie obrazów. Można go traktować jako alternatywny sposób opisu klasycznego schematu kwantyzacji, definiowanego za pomocą zbioru przedziałów kwantyzacji oraz wartości reprezentujące te przedziały.

Kwantyzacja histogramu obrazów cyfrowych jest *de facto* kwantyzacją wtórną, będącą skutkiem redukcji dużego zbioru dyskretnych wartości źródłowych koloru f do mniejszego zbioru możliwie najlepiej dobranych wartości reprezentatywnych $\tilde{f}$. Kwantyzacja rozkładu kolorów jest skutkiem działania operatora $Q_f : \mathbb{Z} \to \mathbb{Z}$ przekształcającego $Q_f\Big(f = a(m)\Big) = \Big(\tilde{f} = b(n)\Big)$ z alfabetem kolorów reprezentatywnych $\tilde{f}(k) \in A_{\tilde{\mathbf{f}}} = \{b_0, \ldots, b_n, \ldots, b_{N-1}\}$, $N << M$ oraz rozkładem kolorów reprezentatywnych

$$\tilde{h}(n) = \#\{k | \tilde{f}(k) = b_n\} \tag{2.26}$$

Tak uproszczony histogram spróbujmy przedstawić w konwencji histogramu rozmytego. Po normalizacji skwantowanego histogramu $p(n) = \tilde{h}(n)/H_{\mathbf{f}}$, gdzie $H_{\mathbf{f}} = \sum_{m=0}^{M-1} h(m) = = \sum_{n=0}^{N-1} \tilde{h}(n)$, histogram możemy zinterpretować w kategoriach prawdopodobieństwa [56], rozumiejąc $p(n) = p(b_n) = Pr(b_n) = Pr(\tilde{f} = b_n)$

jako

$$p(n) = \sum_{k}^{H_{\mathbf{f}}} p(n|k)p(k) = \frac{1}{H_{\mathbf{f}}} \sum_{k=1}^{H_{\mathbf{f}}} p(n|k), \quad \forall b_n \in A_{\tilde{\mathbf{f}}} \tag{2.27}$$

gdzie $p(k) = 1/N$ jest prawdopodobieństwem, że dowolny piksel wybrany z $\mathbf{f}$ jest pikselem k-tym. $p(n|k)$ to prawdopodobieństwo warunkowe, że wartość piksela z indeksem k przypisana jest przedziałowi skwantowanego koloru b_n; w klasycznym histogramie definiowane jest ono binarnie: $p(n|k) = 1$ jeśli k-ty piksel należy do przedziału koloru b_n, zaś w innych przypadkach $p(n|k) = 0$.

W przypadku histogramu rozmytego, nawiązującego do teorii zbiorów rozmytych [55], warunkowe prawdopodobieństwo przynależności piksela (ze względu na jego wartość) do przedziału danego koloru zastępowane jest funkcją przynależności: $\mu : \mathbb{Z} \to [0, 1]$ określającą relację wartości każdego piksela do wszystkich przedziałów skwantowanych kolorów b_n w sposób mniej jednoznaczny, z możliwą niezerową przynależnością f_k do więcej niż jednego przedziału. Ustalenie wartości $\mu(n, k) = \mu(b_n, f_k) = \mu(b_n, f_k = a_m) = \mu(b_n, a_m)$ może się odbywać na podstawie znormalizowanej odległości realnego koloru piksela $f_k = a_m$ od b_n, reprezentującego określony przedział wartości koloru źródłowego $\{a_m | a_m \in [b_{n-1}, b_{n+1}]\}$, przy czym najlepiej, jak odległość ta uwzględnia zróżnicowanie ich percepcji, najlepiej w przestrzeni CIE Lab z percepcyjnie równomiernymi przedziałami wartości kolorów. Pozwala to na realizację prostszego obliczeniowo równomiernego rozkładu wartości kolorów reprezentatywnych, tak że $b_n - b_{n-1} = constant$. Głównym zamysłem jest bowiem przypisanie każdemu pikseli percepcyjnego znaczenia (istotności) jego koloru poprzez odniesienie do wartości reprezentatywnych. Spodziewanym efektem jest rozkład kolorów, który odzwierciedla zróżnicowanie cech percepcji kolorów obrazu, niezależnie od innych czynników (zmian oświetlenia, szum, itd.).

Przez analogię do zapisu histogramu klasycznego (2.27), otrzymujemy

$$r(n) = \sum_{k}^{H_{\mathbf{f}}} \mu(n, k)p(k) = \frac{1}{H_{\mathbf{f}}} \sum_{k=1}^{H_{\mathbf{f}}} \mu(n, k), \quad \forall b_n \in A_{\tilde{\mathbf{f}}} \tag{2.28}$$

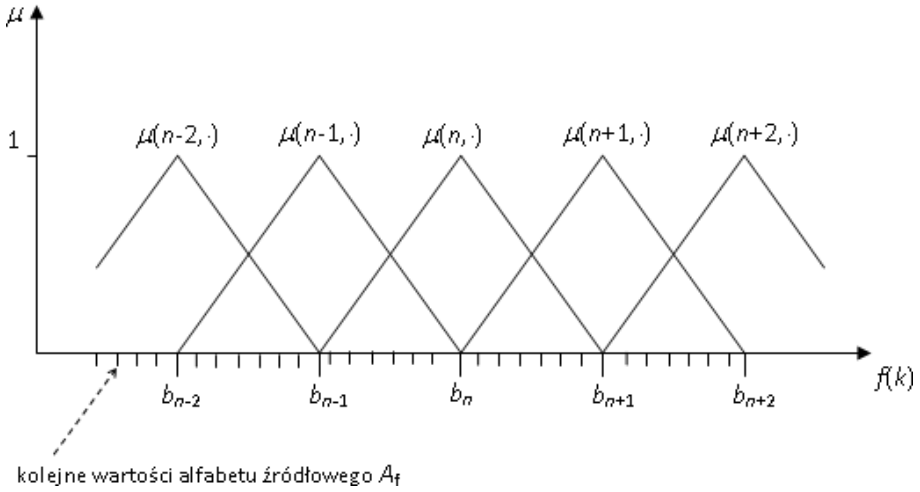
$r(n)$ jest zbiorem rozmytym na przestrzeni wartości pikseli obrazu $\mathbf{f}$, z funkcją przynależności $\mu(n, k)$ do przedziałów kolorów reprezentatywnych b_n, o następujących cechach:

- nośnikiem zbioru rozmytego $r(n)$, czyli $\{f(k) | \mu(n, k) > 0\}$, jest zbiór $\{f(k) | f(k) \in [b_{n-1}, b_{n+1}]\}$;
- $\mu(n, k) = 1$ dla pikseli, których wartość jest dokładnie równa wartości koloru reprezentatywnego $f(k) = b_n$ (piksele te należą do rdzenia $r(n)$);
- $mu(n, k)$ jest ciągła, monotonicznie rosnąca na przedziale $[b_{n-1}, b_n]$ i malejąca na przedziale $[b_n, b_{n+1}]$, symetryczna względem b_n

- spełniona jest zależność

$$\sum_n \mu(n,k) = 1, \quad \forall k = 1, \ldots, H_{\mathbf{f}} \tag{2.29}$$

Na rysunku 2.15 przedstawiono typową funkcję przynależności w postaci dwóch przesuniętych o połowę okresu funkcji trójkątnych. Analogicznie można skonstruować μ dla kolejnych przedziałów kolorów reprezentatywnych za pomocą funkcji trygonometrycznych (dwóch kosinusów przesuniętych o π).

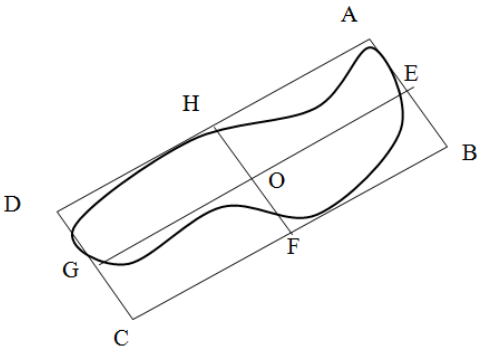


Rysunek 2.15: Ilustracja trójkątnej funkcji przynależności wykorzystywanej przy wyznaczaniu rozmytego histogramu obrazu.

Opis kształtu Charakterystyka kształtu obiektów warunkowana jest wstępną ich segmentacją. Automatyczna segmentacja obiektów istotnych znaczeniowo jest w obrazach naturalnych zagadnieniem trudnym, często mało efektywnym lub wręcz nierozwiązywalnym [58, 36, 44]. Podobnie jest w wielu aplikacja specjalistycznych, np. przy opisie kształtu patologii w obrazach medycznych.

Wśród podstawowych cech opisu kształtu można wyróżnić długość obwodu w odniesieniu do powierzchni obiektu, wklęsłość, oś szkieletu obiektu, relacja pomiędzy parametrami prymitywów geometrycznych (np. prostokątów, sześciokątów foremnych) opisanych i wpisanych w obiekt, gładkość konturu obiektu, ale też cechy samych krawędzi (średni gradient, szerokość), itp.

Na rys. 2.16 podano przykład liczenia cech kształtu określonego obiektu. Poszczególne cechy mogą stanowić wektorowy deskryptor różnicujący kształty struktur przede wszystkim ze względu na ich kolistość i upakowanie, ale też wydłużenie czy relację powierzchni (obwodu) obiektu do powierzchni (obwodu) opisującego go prostokąta.



- najdłuższa oś GE
- najkrótsza oś HF
- obwód i powierzchnia opisującego obiekt prostokąta ABCD
- wydłużenie: GE/HF
- obwód p i powierzchnia S obiektu
- kolistość $C = \frac{4\pi S}{p^2}$
- upakowanie $C_p = \frac{p^2}{S}$

Rysunek 2.16: Przykładowe cechy kształtu wydzielonego (wysegmentowanego) obiektu.

Opis tekstury Tekstura to najogólniej specyficzne cechy obiektu, różnicujące go w stosunku do otoczenia, oddzielające od tła lub innych obiektów. Niekiedy są to powtarzalne wzory, pewna, dająca się dostrzec lub policzyć regularność, nawet odcień barwy. Zależnie od charakteru obiektu, inaczej należy dobierać atrybuty tekstury, inaczej liczyć teksturowe cechy.

Wśród najistotniejszych właściwości tekstury wymienić należy jej regularność (stacjonarność, powtarzalność wzoru, samopodobieństwo, homogeniczność), kierunkowość (zróżnicowanie orientacji, wyrazistość orientacji) oraz skalowalność (rozdzielczość wzorów, zmienność w funkcji skali). Istnieje bardzo wiele metod i koncepcji analizy teksturowych cech obrazów.

W przypadku bardziej stacjonarnych tekstur wykorzystywane są przede wszystkim metody bazujące na binaryzowanych obrazach tekstur (np. według map bitowych), operatorach morfologii matematycznej (domknięcie, otwarcie, gradienty), fraktalach (wymiar fraktalny), probabilistyczne modele losowych pół Markowa, funkcji autokorelacji sygnału czy cechach częstotliwościowych (fourierowskich) (np. w [63]). Często stosowane są też cechy wyznaczane na podstawie macierzy powinowactwa [64, 65, 66], takie jak kontrastowość, zmienność, jednorodność, energia, korelacja.

Niestacjonarne tekstury (zawierające przynajmniej dwa dające się wyróżnić wzory) opisywane są najlepiej za pomocą falek [59, 60] i ich dwuwymiarowych uogólnień (*curvelety*, falki zespolone), innych transformacji typu czas częstotliwość, dwuwymiarowych sygnałów analitycznych, funkcji i kierunkowych filtrów Gabora [47, 61, 62].

Selektywność wyszukiwania

Selektywność wyszukiwania według zadanego scenariusza możemy oceniać, jeżeli znana jest semantyczna relacja równoważności pomiędzy obiektami w kolekcji [27]. To, co nazywamy semantyczną równoważnością w danym przypadku

zależne jest od kontekstu wyszukiwania. W szczególności może ona być bardzo różna dla tych samych obiektów tej samej kolekcji, zależnie od semantyki treści oraz celu użytkowania (wspomniane problemy luki semantycznej).

Będziemy dalej mówić, że zwrócony przez wyszukiwarkę obiekt multimedialny jest poprawny, jeśli jest on semantycznie równoważny z zapytaniem (obiektem wejściowym), tj. treściowa zawartość pytania i odpowiedzi jest równoważna. Konieczne jest więc zdefiniowanie semantycznej (znaczeniowej) relacji równoważności między obiektami w konkretnym zastosowaniu. Przykładowo wyszukane zdjęcia są semantycznie równoważne, jeśli przedstawiają tę samą osobę , a w innym przypadku – jeśli przedstawiają kobietę (poruszany wcześniej problem wielu znaczeń opisywanej treści).

Powszechnie stosowanych jest kilka miar selektywności wyszukiwania, przede wszystkim precyzja (*precision*), przywołania (*recall*), odniesienie precyzji do przywołania, stopa sukcesu (*success rate*) czy średnia ranga (*average rank*) [67, 27]. Miary te są zwykle uśredniane po wielu zapytaniach, by wyniki miały bardziej reprezentatywny charakter.

Precyzja charakteryzuje czułość wyszukiwania pozwalając oszacować jaka liczba (procent) odpowiedzi na zapytanie jest poprawna. Przez Q oznaczmy zbiór wszystkich testowych, możliwie licznych zapytań $q \in Q$. Jeśli wśród wszystkich $K(q)$ odpowiedzi na każde q znajdzie się dokładnie $T(q) \geqslant$ odpowiedzi poprawnych (co oznacza również $K(q) - T(q)$ odpowiedzi niepoprawnych, wtedy precyzję obliczamy jako

$$pr = \frac{1}{|Q|} \sum_{q} \frac{T(q)}{K(q)} \tag{2.30}$$

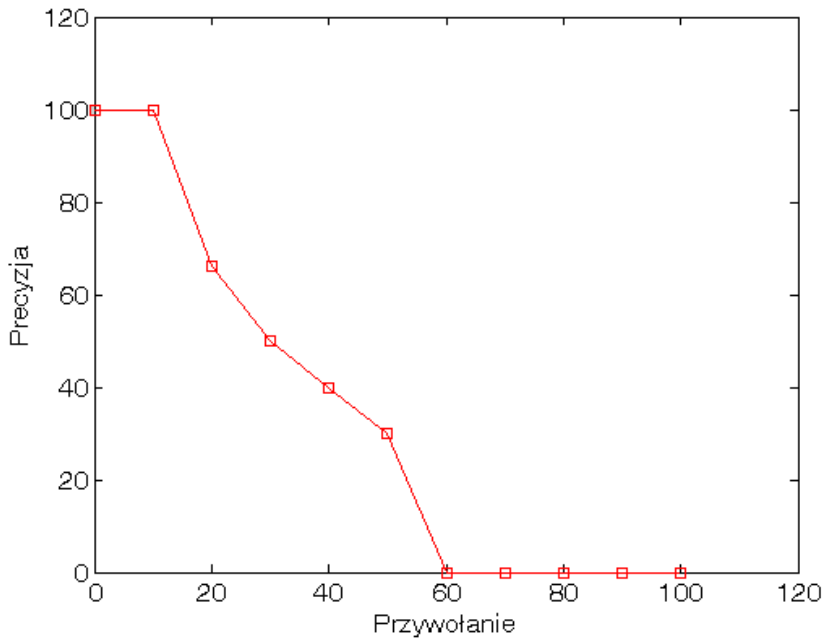
Wygodniej jest ustalić stały scenariusz wyszukiwania odpowiedzi kładąc stałe $K(q) = K$, niezależnie od zapytania. Często jest to koniecznością przy testowaniu bazy o nieznanej zawartości (nie wiemy, ile obrazów danej treści jest w bazie). Jeśli jednak znana i zróżnicowana jest reprezentacja odmiennych treściowo obiektów bazy, rzetelniejsze wyniki można uzyskać ustalając odpowiednio duże $K(q)$. W szczególności, by nie ograniczać liczby możliwych odpowiedzi poprawnych warto przyjąć przynajmniej $K(q) \geqslant N(q)$, gdzie $N(q)$ jest liczbą wszystkich obiektów bazy semantycznie równoważnych q.

Przywołanie wskazuje, ile z wszystkich potencjalnych odpowiedzi poprawnych na zapytanie q (ich liczba jest równa liczbie wszystkich obiektów semantycznie równoważnych q, które znajdują się w bazie) znalazło się wśród obiektów zwróconych przez wyszukiwarkę. Mamy więc przywołanie zdefiniowane jako

$$rec = \frac{1}{|Q|} \sum_{q} \frac{T(q)}{N(q)} \tag{2.31}$$

Miara ta zależy więc jeszcze silniej od scenariusza wyszukiwań testowych (tj. sposobu ustalania wartości $K(q)$). Z (2.30) oraz (2.31) wynika własność $pr, rec \in [0, 1]$.

Precyzja odniesiona do wartości przywołania jest częstym miernikiem efektywności procesu indeksowania – zobacz rys. 2.17. Niekiedy są takie sytuacje, gdy chcielibyśmy wyznaczyć efektywność indeksowania dla pojedynczych zapytań. Można podać przynajmniej dwa przykłady takiej sytuacji. Po pierwsze, uśrednianie po wielu zapytaniach może ukryć pewne niepożądane cechy algorytmu indeksującego, występujące dla specyficznych zapytań. Po drugie, przy porównywaniu algorytmów indeksujących, może być istotna informacja, czy dany algorytm jest lepszy od innego dla każdego zapytania w określonej grupy zapytań testowych. W takich sytuacjach można zastosować pojedynczą wartość precyzji (dla każdego zapytania), traktując jako uśrednioną dla danego poziomu przywołania, zgodnie w podejściem przedstawionym powyżej. Możliwe jest także zaproponowanie innych podejść do wyznaczania precyzji, które mogą być bardziej użyteczne [67], takie jak średnia precyzja dla widzianych dokumentów adekwatnych czy R–precyzja [33]. Średnia precyzja dla widzianych dokumentów adekwatnych polega na wy-



Rysunek 2.17: Przykładowy wykres precyzji w funkcji przywołania. Standardowo precyzja jest wyznaczana dla 11 poziomów przywołania (źródło: [33]).

znaczaniu precyzji poprzez uśrednianie precyzji wyznaczanej po zaobserwowaniu w zbiorze wynikowym adekwatnego dokumentu. Przykładowo, dla danych z ryzunku 2.17 wartość precyzji po pojawieniu się w zbiorze wynikowym kolejnych

dokumentów odpowiadających zapytaniu będzie następująca: 1, 0,66, 0,5, 0,4 oraz 0,3. Na tej podstawie możemy policzyć $pr = 0,57$.

Duże wartości $K(q)$) nie są zbyt użyteczne, gdyż zwykle oczekiwana jest raczej odpowiedź kilku najbardziej podobnych treściowo obiektów. Bardziej użyteczne mogą się więc okazać miary uwzględniające pozycje zwracanych odpowiedzi poprawnych.

Stopa sukcesu wskazuje na procentowy udział takich odpowiedzi na Q, w których na pierwszej pozycji znalazły się obiekty semantycznie równoważne zapytaniu, a mianowicie

$$sr = \frac{1}{|Q|} \sum_{q \in Q} \sigma(q) \tag{2.32}$$

gdzie $\sigma(q) = 1$ gdy na pierwszym miejscu zwrócony został obiekt równoważny zapytaniu, zaś w przeciwnym przypadku $\sigma(q) = 0$. Miara ta zakłada porządkowanie według nierosnącego podobieństwa obiektów zwracanych do zapytania, według przyjętej funkcji podobieństwa obiektów.

Średnia ranga jest miarą bardziej złożoną, odnoszącą się do średniej pozycji odpowiedzi poprawnych w stosunku do testowych zapytań. Podobnie jak poprzednio, lista odpowiedzi jest porządkowana ze względu na nierosnące podobieństwo względem q. Obraz najbardziej podobny zajmuje pozycję 1, a najmniej – pozycję $K(q)$. Średnia ranga ara jest uśrednioną po wszystkich Q wartością średniej pozycji odpowiedzi poprawnych na pojedyncze q. Mamy więc

$$ara = \frac{1}{|Q|} \sum_{q \in Q} ra(q) \tag{2.33}$$

gdzie $ra(q)$ jest średnią pozycją wszystkich obiektów równoważnych q znajdujących się w bazie, w odniesieniu do konkretnej realizacji zapytania q. Tak więc

$$ra = \frac{1}{N(q)} \sum_{n=1}^{N(q)} pos(n, q) \tag{2.34}$$

$pos(n, q)$ określa tę pozycję jako

$$pos(n, q) = \begin{cases} k & \text{jeżeli obiekt } n \text{ zajmuje w odpowiedzi pozycję } k \leqslant K(q) \\ K(q) + 1 & \text{w p. p.} \end{cases}$$

2.4 Podsumowanie

Do najistotniejszych zagadnień poruszonych w tym rozdziale należą podstawy teorii informacji. Sposób definiowania pojęcia informacji kształtuje wykorzystywane modele przekazu wpływając w stopniu decydującym na konstrukcję i doskonalenie aplikacji multimedialnych. Konsekwencją jest ustalenie właściwej reprezentacji danych, określenie potrzeb odbiorcy, zapewnienie korzystnej adaptacji warunków przekazu oraz możliwych form interakcji. Tworzenie efektywnej reprezentacji bazuje na doborze właściwej klasy sygnałów stanowiących syntaktyczną formę przenoszonych informacji ułatwiając semantyczny opis i indeksowanie treści przekazu. Konstrukcja efektywnych deskryptorów warunkuje selektywność wyszukiwania danych według ich zawartości. Generalnie problem semantyki przekazu, komputerowego opisu treści w kontekście mechanizmów rozumienia danych przez użytkownika, a także uwzględnienie przewidywanych sposobów ich interpretacji czy oceny stanowią niewątpliwie główny problem badawczy prac nad doskonaleniem technologii multimedialnych.

Zadania do tego rozdziału podano na stronie 362.

Rozdział 3

Komputerowe przetwarzanie informacji – metody

W rozdziale tym scharakteryzowano wybrane metody komputerowego przetwarzania (w skrócie – obróbki) multimediów, przekształcające jej reprezentację pod kątem określonych zastosowań, ze szczególnym uwzględnieniem charakteru przekazywanej informacji. Wielość i różnorodność stosowanych rozwiązań, sprawdzonych metodologii konstrukcji algorytmów i całych narzędzi, stanowiących o sile współczesnych multimediów, zmusza do prezentacji jedynie wybranych metod ulepszania danych, analizy i syntezy, kompresji czy indeksowania.

Wyróżnikiem są metody służące realizacji zasadniczych koncepcji przetwarzania danych, które są rozumiane przede wszystkim jako metody przetwarzania informacji, w sposób charakterystyczny dla wybranego zagadnienia. Maksymalna użyteczność przekazu, mierzona konkretnymi efektami wykorzystania multimediów stanowi główne kryterium przydatności metod.

3.1 Komputerowe przetwarzanie danych (KPD) multimedialnych

Zasadniczym celem komputerowego przetwarzania danych jest doskonalenie przekazu informacji multimedialnej od etapu pozyskiwania danych źródłowych (rejestracji źródła przekazu) po etap prezentacji danych odbiorcy (grupie odbiorców, według scenariusza). W zależności od zastosowań, to doskonalenie zakłada różne formy wejściowe danych oraz przybiera różne formy wyjściowe dostosowane do modelu (schematu) użytkownika czy użytkowania.

Tak szeroka definicja zagadnienia obejmuje także metody kodowania (kompresji), które służą niewątpliwie doskonaleniu przekazu, podobnie jak metody indeksowania pozwalające opisać, a przez to uporządkować przekaz złożony.

Wśród zagadnień komputerowego przetwarzania danych można wyróżnić przede wszystkim:

- rejestrację danych
- kodowanie danych cyfrowych
- ulepszanie danych
- analizę danych, w tym:
  - rozpoznanie wzorców
  - rozumienie danych
  - opis danych za pomocą numerycznych deskryptorów
  - interpretację danych
- syntezę danych na podstawie:
  - modeli elementarnych (strukturalnych)
  - modeli złożonych (obiektowych)
  - modeli probabilistycznych
  - modeli fizycznych i pseudofizycnych (empirycznych)
- wyszukiwanie poindeksowanych danych podobnych

Zagadnienia te są ograniczone z jednej strony procesem rejestracji sygnałów (danych) naturalnych bądź specjalistycznych, z drugiej zaś strony – charakterystyką użytkowania zawartych w sygnale (w danych) informacji. Pomiędzy nimi znajdują się obszary koncepcji bardziej uniwersalnych, odnoszące się do trzech kluczowych aspektów: reprezentowania danych cyfrowych (kodowanie-kompresja), rozpoznania znaczenia danych (inteligentna analiza abstrakcyjna), porządkującego opisu danych (indeksowanie z kryterium podobieństwa).

W kontekście przetwarzania danych multimedialnych dochodzi jeszcze istotny aspekt integracji strumieni informacji oraz kształtowanie synergii przekazu, obecny w jakimś stopniu w każdym z wymienionych zagadnień.

Arsenał metod KPD jest bardzo bogaty, a próba ich syntetycznego zestawienia nie jest prosta. Poniższy podział służy przede wszystkim wskazaniu najbardziej przydatnych, elementarnych algorytmów przetwarzania danych, ukazujących jednocześnie różnorodność możliwych działań na danych w celu uzyskania zamierzonych efektów aplikacyjnych.

Wśród metod komputerowego przetwarzania danych można wyróżnić

- służące przede wszystkim ulepszeniu danych
  - operacje histogramowe: a) adaptacyjne - na bazie relacji histogramu źródłowego do docelowego w skali globalnej bądź lokalnej; b) według ustalonego przyporządkowania punkt źródłowy - punkt docelowy;
  - filtracje splotowe: a) kontekstowe w przestrzeni źródłowej; b) skalowalne (połączone ze zmianą skali sygnału źródłowego);
  - filtracje częstotliwościowe, wykorzystujące transformacje Fouriera sygnałów źródłowych oraz częstotliwościowe charakterystyki filtrów (mnożone przez widmo sygnału);
  - operacje morfologii matematycznej, wykorzystujące oddziaływanie określonego elementu strukturującego (inaczej strukturalnego) na geometryczne właściwości obiektów;
  - przekształcenia geometryczne źródłowych przestrzeni dyskretnych, przede wszystkim afiniczne (obrót, skalowanie, przesunięcie) w rzeczywistych przestrzeniach euklidesowych (celem np. korekty źle ustawionego obiektywu kamery czy dopasowania obrazów tej samej rzeczywistości wykonanych różnymi technikami);
  - aproksymacje z wykorzystaniem liniowych rozwinięć sygnałów: a) interpolacja (np. w celu zwiększania rozdzielczości danych źródłowych - *superresolution*); b) ekstrapolacja (np. w celu wypełniania dziur, czyli ogólniej - brakujących fragmentów w sygnale źródłowym - *inpainting*);
- służące analizie danych
  - wydzielanie jednorodnych, mających określone znaczenie (semantykę) fragmentów sygnału – zasadniczym celem jest ułatwienie ich percepcji, uproszczenie reprezentacji danych oraz ułatwienie analizy treści; przykładem są metody segmentacji obrazów, które pozwalają wydzielić obiekty obrazowanej przestrzeni;
  - wydzielanie komponentów, czyli składowych sygnału nierozróżnialnych percepcyjnie, celem usunięcia nadmiarowości reprezentacji źródłowej,

a więc jej uproszczenia oraz wydzielenia charakterystycznych, bardziej niezmienniczych składników ułatwiających trafną analizę;

- wyznaczanie szeregu cech opisujących właściwości istotnych fragmentów (obiektów) czy też komponentów sygnału w postaci deskryptorów numerycznych (liczbowych operacji przybliżonych) o możliwie dużych walorach semantycznych (mających znaczenie dla użytkownika);
- selekcja cech i klasyfikacja w celu automatycznego rozpoznania treści; celem jest uformowanie takiej przestrzeni cech opisujących interesujące właściwości sygnału, która pozwoli różnicować wzorce poszczególnych klas obiektów lub ich wzajemnych relacji, aby rozpoznać treść przekazu multimedialnego na ustalonym poziomie abstrakcji;
- formalizacja wiedzy dziedzinowej w celu stworzenia w miarę kompletnego, hierarchicznego i relacyjnego opisu wiedzy w danym obszarze; celem jest stworzenie mechanizmów opisu danych źródłowych (sygnału) w kategoriach semantycznych odpowiadających właściwym dla użytkownika poziomom abstrakcji; należy tu wymienić przede wszystkim a) ontologie z mechanizmami wnioskowania i możliwością integracji z deskryptorami numerycznymi; b) gramatyki formalne i języki; c) encyklopedie wykorzystujące całe *continuum* metod formalizacji danych, od tekstów zapisanych w edytorach i zarejestrowanych obrazów, poprzez leksykony, semantyczne opisy, referencyjne przypadki o ustalonym znaczeniu, drzewa decyzyjne, reguły logiczne, po zaawansowane modele funkcjonalne, reguły decyzyjne czy modele błędów [153]; d) narzędzia integracji wiedzy dziedzinowej z semantycznymi deskryptorami, mechanizmami rozpoznawania i interpretacji treści (w tym interaktywnymi) oraz wiarygodnymi modelami obliczeniowymi aproksymującymi pojęcia abstrakcyjne;

- służące syntezie danych
  - konstrukcja modeli obiektów lub komponentów pozwalających na syntezę treści przekazu multimedialnego: a) na bazie danych z zaplanowanych eksperymentów (np. rejestracja sygnałów z zaprojektowanego zestawu czujników); b) na bazie reprezentatywnych danych referencyjnych (analizowanych dobranymi algorytmami); c) z wykorzystaniem wirtualnych narzędzi na bazie zestawów prostych elementów konstrukcyjnych (prymitywów), określonego typu obiektów czy procedur stochastycznych;
  - projektowanie modeli odbiorcy, jego zdolności percepcji oraz preferencji użytkowych, a także okoliczności przekazu (uwarunkowań, takich jak np. charakterystyka pomieszczeń i zestawów odsłuchowych, czy perspektywy i dynamiki ruchu kamery);

– tworzenie (generację) syntetycznej postaci sygnału z kryterium możliwe efektywnej prezentacji (odsłuchu, wizualizacji), z wykorzystaniem procedur: a) zwiększających realizm (wiarygodność, np. poprzez nałożenie wiarygodnych tekstur w obrazach grafiki komputerowej); b) redukujących złożoność obliczeniową (np. poprzez redukcję liczby obliczeń rzeczywistoliczbowych, uproszczenia widoku czy formy dźwiękowej do granic zdolności percepcji); c) zapewniających niezmienniczość formy prezentacji względem przekształceń (typu przesunięcie czy obrót) w dyskretnych dziedzinach czasu i przestrzeni (np. uwzględniając problem rasteryzacji w obrazach),

– doskonalenie postaci prezentowanego sygnału poprzez miksowanie naturalnych sygnałów źródłowych z syntetycznie odtwarzanymi – problem dopasowania czy wpasowania (*registration*), lokalnego uciągania sygnału na granicach fragmentów łączonych itp.; przykładem zastosowań jest produkcja filmowa;

3.1.1 Komputerowe przetwarzanie obrazów

Wśród zagadnień KPD szczególną rolę odgrywają zagadnienia dotyczące komputerowego przetwarzania obrazów. Są tego dwa podstawowe powody – dominująca w zastosowaniach multimedialnych waga przekazu wizyjnego, wynikająca z potencjału i szeroko wykorzystywanych zalet informacji obrazowej, a także bogactwo stosowanych metod, w których znajdują analogię metody obróbki innych rodzajów danych multimedialnych.

Podstawowe operacje wykonywane na obrazie to przede wszystkim:

a) akwizycja, czyli pozyskiwanie obrazów cyfrowych według schematu:
źródło → obraz

b) ulepszanie (wstępne, poprawa jakości obrazu, poprawa percepcji treści) według schematu:
obraz → obraz

c) analiza (segmentacja-obiekty, dekompozycja-komponenty, rozpoznawanie, komputerowe widzenie) według schematu:
obraz → opis[1]
[1](wyróżnienie i charakterystyka obiektów czy regionów zainteresowań, ich wzajemnych relacji)

d) grafika komputerowa według schematu:
opis (model) → **obraz**

e) interpretacja według schematu:

obraz, opis → opis semantyczny[2]

[2] (na poziomie abstrakcji użytkownika).

3.1.2 Ograniczenia procesu rejestracji danych

Praktyczne cele rejestracji sygnałów mogą być różnorakie – przykładowo:

- utrwalenie chwili (fotografia, film, historyczna kopia zdarzeń) o takich cechach jak
  - wierny, specyficzny, wysokiej jakości,
  - dogodny w dalszej obróbce,
  - o istotnych walorach treściowych,
  - podatny na długoczasowe przechowanie,
  - dostosowany do przewidywanych form odtwarzania (percepcji przez użytkownika);
- zdobycie informacji (niedogodności zapisu, duże koszty organizacyjne, wyjątkowość sytuacji, szybki przekaz), gdzie ważne jest
  - ukształtowanie przekazu informacji (selekcja),
  - zapewnienie wystarczającej jakości zapisu,
  - uzupełnienie metodami obróbki ograniczeń uwarunkowań rejestracji;
- obserwacja natury, rzeczywistości (badania i eksperymenty, zdobycie wiedzy o świecie, obiektywizm, kompleksowość zapisu określonego stanu czy zbioru faktów, rozumienie i odzwierciedlenie praw natury) wymagające takich zasad rejestracji jak
  - rzetelne i wiarygodne dostosowanie do realiów,
  - wtopienie w istotę zapisywanych zjawisk, cierpliwe naśladowanie procesów i zachowań, podpatrywanie bez ingerencji czy zmiany naturalnych uwarunkowań (obojętność względem natury zjawisk),
  - konsekwencja w tropieniu prawdy;
- wykorzystanie natury (symulowanie świata, modelowanie) dające
  - wiarygodny świat wirtualny, tworzący nowe możliwości i szanse na bazie analizowanych, wyczerpujących zapisów natury, pozwalający dostosować realia do ograniczeń percepcji użytkownika (powodowanych np. brakiem możliwości przebywania, chorobą, upośledzeniami),

  - formalny model zjawisk, który pozwala rozszerzać rzeczywistość z zachowaniem kluczowych cech natury, udostępniać na szeroką skalę, uczyć zrozumienia istoty zjawisk;

- rozwój świata technologii cyfrowych, a więc

  - koncepcji internetu, cyfrowego odbicia rzeczywistości bez szeregu naturalnych ograniczeń powodowanych ciągłością czasu i miejsca, globalnego kontaktu i dostępu,
  - komunikacji bez granic, bezprzewodowego, szerokopasmowego przepływu danych na niewyobrażalną skalę,
  - personalnych urządzeń komputerowych ze zdolnością obróbki każdego cyfrowego odbicia realiów dowolnego miejsca i chwili.

Różnorodne metody, urządzenia czy systemy rejestracji danych definiowane są poprzez:

- fizyczne podstawy różnych koncepcji rejestracji sygnałów, czyli

  - wykorzystanie właściwych zjawisk fizycznych, umożliwiających pomiar istotnych cech rejestrowanej rzeczywistości;
  - ewentualne zastosowanie konwersji sygnału jako nośnika informacji do innej formy przenoszenia energii, umożliwiającej wyższej jakości pomiar informacji (np. konwersja promieniowania rentgenowskiego na świetlne za pomocą scyntylatorów w cyfrowych detektorach radiografii rentgenowskiej – dopiero fotony o zdecydowanie niższej energii mogą być skutecznie wyłapywane przez macierze CCD czy też fotodiody sprzężone z przestrzennie dyskretnymi macierzami tranzystorów TFT (*thin film transistors* z amorficznego krzemu);
  - projektowanie czujników/detektorów do konwersji rozkładów mierzonych wielkości fizycznych na energię elektryczną (przepływ ładunków) z zachowaniem geometrii, przestrzennych relacji i możliwej zupełności rejestrowanej treści;
  - konstrukcja urządzeń i systemów zapewniających odpowiednie warunki pomiarów: a) w założonych zakresach dynamiki zmian (odpowiednio szerokich, lecz nie nadmiarowych); b) dotyczące relacji czasowych (zarówno co do czasu trwania pojedynczego pomiaru, jak i niezbędnej liczby pomiarów danej wielkości) i innych okoliczności wynikających z zastosowań;
  - zapewnienie stabilności warunków rejestracji z kontrolą i śledzeniem kluczowych parametrów tego procesu (przede wszystkim stosunku sygnał/szum, zdolności rozdzielczych, zakłóceń, wiarygodności);

- zasady rejestracji sygnałów cyfrowych, uwzględniające:
  - reguły próbkowania, kwantyzacji i kodowania stosowane w uniwersalnych przetwornikach analogowo-cyfrowych A/C, bądź też w określonym systemie (układzie) akwizycji sygnałów – czego efektem jest postać cyfrowej reprezentacji sygnałów;
  - procedury wstępnego przetwarzania danych, służące ulepszeniu sygnału poprzez redukcję znanych z góry ograniczeń danego systemu rejestracji (np. redukcji szumów poprzez uśrednienie sygnałów z kilku kanałów pomiarowych, wycięcia składowych pasożytniczych czy też wzmocnienia składowych użytecznych, wprowadzenia dynamicznego wzmocnienia sygnału rejestrowanego zależnie od czasu propagacji przez obszar mierzony);
  - zasady gromadzenia danych, dotyczące sposobu formowania/rekonstrukcji cyfrowego sygnału zapisu oraz wyznaczania postaci reprezentacji dostosowanej do uwarunkowań systemu cyfrowego danej aplikacji; chodzi tutaj m.in. o uwzględnienie wymagań dotyczących: a) czasowej przepustowości strumienia przechwytywanych danych oraz wynikającej z tego koniecznej wydajności czasowej zapisu danych; b) standaryzacji formatu gromadzonych danych, co może pociągać za sobą konieczną konwersję (na bieżąco) reprezentacji danych zgodnie z wymogami określonej normy; c) dostosowania do uwarunkowań ewentualnej transmisji na bieżąco (*on-line*) rejestrowanych danych, w tym standaryzacji formatu danych, wymogów wprowadzenie wydajnych mechanizmów kolejkowania i buforowania, dostosowujących czasowe uwarunkowania rejestracji do zmiennych czasowo parametrów kanału transmisyjnego.

Ograniczenia metod rejestracji sygnałów dotyczą m.in. takich czynników jak:

- ogólna wiarygodność pomierzonego odbicia rzeczywistości względem zaistniałych realiów (przede wszystkim w zakresie podstawowych właściwości występujących obiektów i ich wzajemnych relacji, a także kompletności zbieranej informacji, uwzględnienia występujących efektów maskowania istotnych cech sygnału, selektywnego wzmacniania itp.),
- poziom i charakter występujących szumów,
- rodzaj i intensywność artefaktów, ogólniej zakłóceń,
- zdolność rozdzielcza zestawu pomiarów (możliwa skala dokładności) akwizowanych danych, jej charakter przestrzenny (kierunkowy, czy też geometryczny) oraz czasowy,

- zakres dynamiki pozwalający z wystarczającą czułością, ale i specyficznością różnicować zapisywaną treść.

Ograniczenia te stanowią istotną przeszkodą w percepcji, analizie czy interpretacji gromadzonej informacji, w kontekście określonych form jej użytkowania wielu różnorodnych zastosowań. Wymagana jest wtedy obróbka danych zmierzająca do poprawy ich użyteczności w ramach dostępnych środków sprzętowych, przy istniejących ograniczeniach realizacyjnych (dotyczących złożoności obliczeniowej, zależności czasowych, natury treści przekazu, zakresu możliwych metod i koncepcji itp.). Przykładowo w zastosowaniach medycznych stosowanie metod obróbki, które nie dają pewności zachowania wszystkich, nawet najdrobniejszych szczegółów, mogących mieć znaczenie w rozpoznaniu patologii, budzi zasadniczy sprzeciw. Dochodzą w tym przypadku także uwarunkowania rekonstrukcji jako rozwiązania problemu odwrotnego. Z kolei dostępna szybkość przetwarzania danych multimedialnych, zestawiona z rozmiarami strumienia danych wymagających obróbki, może okazać się niewystarczająca.

3.1.3 Metody ulepszania danych

Celem wstępnego przetwarzania danych jest zwiększenie ich użyteczności w kontekście określonych zastosowań, wobec ograniczeń procesu rejestracji danych źródłowych.

Ze względu na cel można wyróżnić metody ulepszania danych multimedialnych służące:

- poprawie czy też unormowaniu ogólnej jakości sygnału (obrazu czy dźwięku) w zakresie korekty kontrastu lokalnego i globalnego obrazu, kształtowania odpowiedniej barwy dźwięku, czyli jego charakterystyki częstotliwościowej (np. wzmocnienia tonów istotnych), redukcji szumów i artefaktów, dostosowania rozdzielczości czasowej i przestrzennej (np. do wymagań systemów odsłuchowych); celem jest uzyskanie wyższej skuteczności komputerowych metod analizy przetworzonych wstępnie danych w stosunku do danych źródłowych, a niezbędne w optymalizacji tych metod jest wykorzystanie wiarygodnych miar jakości sygnałów;

- poprawie percepcji (widoczności, odsłuchu) treści przekazu poprzez dodatkowe uwzględnienie modelu odbiorcy w zakresie zdolności percepcji sygnałów dźwiękowych i obrazowych (właściwości ludzkiego systemu słyszenia i widzenia), a także wykonawczej charakterystyki pracy z sygnałem (użytkowego działania odbiorcy informacji, przykładowo za pomocą charakterystyki ROC [6] wykorzystywanej m.in. w psychologii, telekomunikacji i medycynie); istotnym warunkiem ich skuteczności jest zapewnienie możliwie optymalnych warunków prezentacji przetworzonych danych – kluczowa

może się okazać integracja metod przetwarzania i wizualizacji (odsłuchu), uwzględniająca uwarunkowania sprzętowe i środowiskowe (czyli okoliczności prezentacji sygnału);

- uwydatnieniu walorów użytkowych, określonych dostępną wiedzą dziedzinową danego zastosowania, w tym przede wszystkim
  - ocenie-kontroli jakościowej danych źródłowych, by określić ich przydatność według wymagań aplikacyjnych (np. zastosowanie dobranych metod przetwarzania może wykazać brak istotnych walorów jakościowych zapobiegając błędom interpretacji)
  - ekstrakcji, wzmocnieniu czy wyróżnieniu informacji ukrytej, słabo postrzeganej, maskowanej przez niekorzystne czynniki, wydobyciu czy podkreśleniu kluczowych cech semantycznych przekazywanego sygnału;
  - zastosowaniu wstępnej analizy danych (np. segmentacji regionów zainteresowania) w celu selektywnego przetworzenia jedynie wybranych fragmentów czy też komponentów sygnału źródłowego; ułatwi to interpretację zasadniczej treści zmieniając niekiedy w sposób znaczący charakter prezentowanego czy dalej analizowanego sygnału;
  - innym, bardziej specjalistycznym zastosowaniom.

Typowe zastosowanie metod przetwarzania obrazów to ich poprawa, często poprzez ekstrakcję informacji, która pozwoli lepiej zrozumieć treść przekazu obrazowego. Są to ogólnie metody przetwarzania wstępnego, zaś celem zasadniczym jest zapewnienie pełnego, możliwie czytelnego przekazu informacji źródłowej.

Niekiedy jednak w aplikacjach stosowane jest podejście odwrotne, czyli aby ulepszyć obraz, należy go najpierw dobrze zrozumieć. Ten ulepszony obraz jest *de facto* efektem końcowym, ikoną zasadniczego przekazu informacji. W tym przypadku chodzi przede wszystkim o selekcję treści użytecznej i uproszczenie możliwie jednoznacznej formy jej reprezentacji (w konsekwencji wizualizacji).

Metody ulepszania obrazów

Wśród podstawowych operacji wykonywanych na obrazie (lub wybranym regionie) należy wyróżnić:

- przetwarzanie punktowe jak
  - regulacja kontrastu i jasności według ustalonej funkcji – liniowej, nieliniowej, nieciągłej, kawałkami gładkiej itp.;
  - operacje histogramowe, przede wszystkim korekcja histogramu (wyrównywanie, rozciąganie);

- przetwarzanie kontekstowe, różne formy filtracji splotowej, odszumiającej, wykrywającej lub podkreślającej krawędzie, operacje nieliniowe;
- przetwarzanie globalne całego obrazu, w tym:
  - aproksymacje wykorzystujące transformacje w bazach fourierowskich, wielorozdzielczych, itp., z podziałem na bloki oraz lokalizacją funkcji bazowych, z progowaniem, modyfikowaniem rozkładu współczynników dziedziny przekształcenia;
  - przekształcenia geometryczne lub graficzne (tj. metodami grafiki komputerowej - np. zmiana parametrów oświetlenia sceny, doskonalenie algorytmu globalnej iluminacji i inne);

Dwa zasadnicze kierunki ulepszania to poprawa percepcji oraz zwiększenie skuteczności metod analizy i syntezy obrazów.

Regulacja kontrastu i jasności

Najprostszą formą regulacji kontrastu i jasności w obrazie jest zastosowanie punktowego przekształcenia pikseli obrazu źródłowego $f(k,l),\ k,l \in \mathbb{Z}$ za pomocą regulatora kontrastu - stała κ oraz regulatora jasności - stała β według zależności

$$g(k,l) = \kappa \cdot f(k,l) + \beta \tag{3.1}$$

przy czym należy uwzględnić dopuszczalną dynamikę wartości poziomów jasności tak dla obrazu źródłowego, jak i dla docelowego $\mathbf{g}$. Ogólniej metody poprawy kontrastu według ustalonej funkcji zależności wyjściowych poziomów jasności od wejściowych, zadanej zwykle analitycznie lub za pomocą tablicy przypisań LUT (*look-up table*)) opisane są regułą

$$g(k,l) = F\{f(k,l)\} \tag{3.2}$$

gdzie wartości jasności są znormalizowane $f, g \in [0,1]$. Najczęściej stosowane postacie punktowych przekształceń analitycznych to

- korekcja gamma z operacją potęgowania $g(k,l) = f(k,l)^{\gamma}$, gdzie np. do korekcji zdjęć cyfrowych stosuje się przyciemniające $\gamma = 2,5$ lub rozjaśniające $\gamma = 0,5$, a do korekcji wyświetlania na monitorach CRT $\gamma = 1/2,2$;
- negatyw $g(k,l) = 1 - f(k,l)$ lub tez stosowana w fotografice solaryzacja

$$g(k,l) = \begin{cases} 2 \cdot f(k,l) & \text{dla } 0 \leqslant f(k,l) \leqslant 0,5 \\ 2(1 - f(k,l)) & \text{dla } 0,5 < f(k,l) \leqslant 1 \end{cases} \tag{3.3}$$

- rozjaśniająca funkcja logarytmiczna $g(k,l) = \frac{1}{\ln 2} \ln(f(k,l) + 1)$

- funkcja okna, wykorzystywana np. do wizualizacji jedynie określonego podzakresu dostępnej dynamiki danych źródłowych $[d, g] \subset [0, 1]$, postaci

$$g(k,l) = \begin{cases} 0 & \text{dla } 0 \leqslant f(k,l) \leqslant d \\ \frac{1}{g-d}(f(k,l) - d) & \text{dla } d < f(k,l) < g \\ 1 & \text{dla } g \leqslant f(k,l) \leqslant 1 \end{cases} \tag{3.4}$$

Przykładowe efekty punktowego przetwarzania obrazów według wybranych funkcji regulacji kontrastu i jasności pokazano na rys. 3.1. Dobór funkcji okna nabiera szczególnego znaczenia w obrazowaniu medycznym, np. przy ocenie obrazów warstwowych tomografii komputerowej. Radiolog obserwuje wówczas jedynie wybrany zakres dynamiki 12. bitowych danych, obejmujący interesujący rodzaj tkanki - rys. 3.2. Przedstawione obrazy dotyczą badań diagnostycznych wczesnego udaru mózgu, zaś podane wartości okna wyrażono w jednostkach Hunsfielda[1].

Operacje histogramowe

Histogram obrazu jest graficzną reprezentacją rozkładu wartości pikseli (inaczej wartości funkcji jasności lub poziomów jasności) w obrazie. Każda z wartości pikseli $f(k, l)$ obrazu $\mathbf{f}$ należy do uporządkowanego rosnąco zbioru (alfabetu) wartości możliwych: $f(k, l) \in A_{\mathbf{f}} = \{a_0, \ldots, a_{M-1}\}$. Liczba wystąpień kolejnych poziomów jasności (inaczej, liczba pikseli dziedziny obrazu Ω_f z określonym poziomem jasności) $h(m)$ dla $m = 0, \ldots, M - 1$, takich że

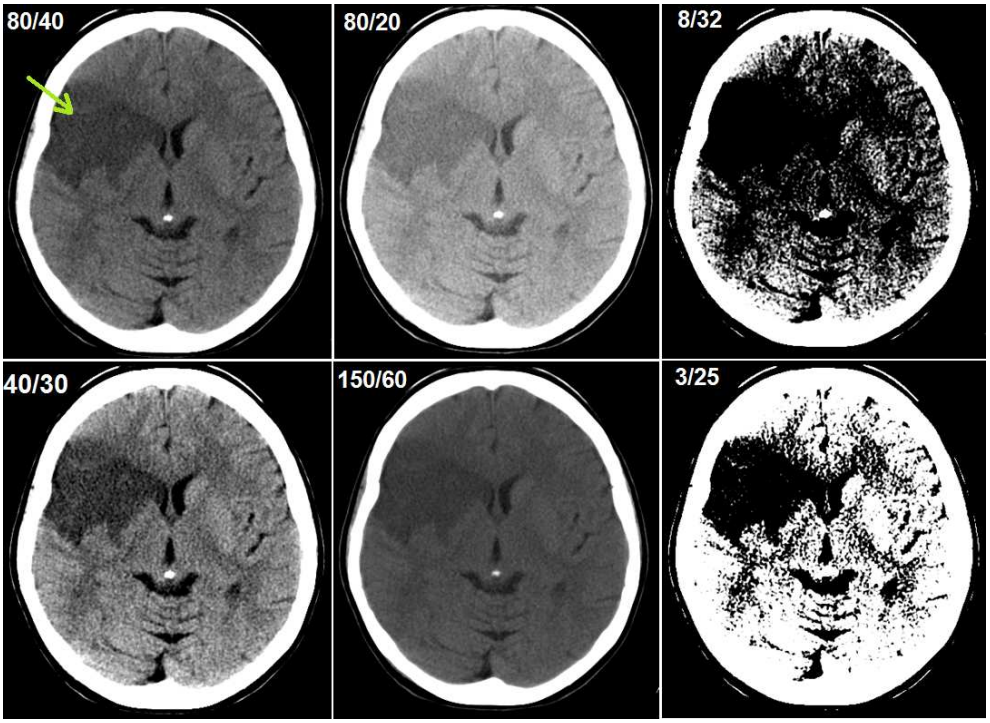
$$h(m) = \#\{(k,l) \in \Omega_f | f(k,l) = a_m\} \tag{3.5}$$

stanowi histogram prezentowany w określonej formie graficznej (zwykle wykresu słupkowego, niekiedy w postaci tablicy par wartości $(a_m, h(m))$.

Histogram charakteryzuje globalne skontrastowanie obrazu, przy czym rozkład zbliżony do równomiernego świadczy zasadniczo o "dobrym" kontraście i zachowanej równowadze pomiędzy obszarami o różnej jasności. Zachwianie tej równowagi może w klasycznych zobrazowaniach oznaczać deficyt obszarów jasnych czy ciemnych, lub też dowodzić zbytniej polaryzacji (np. czarno-białej) rozkładu wartości pikseli. Wyjątkiem są obrazy specjalistyczne, kiedy to same zasady pomiaru i przebieg procesu akwizycji narzucają niekiedy silne nierównomierności histogramów. W takich przypadkach rozważa się niekiedy histogramy lokalne, liczone jedynie w obszarach zainteresowań jako wskaźniki ich podatności na percepcję określonych szczegółów obrazu. Zachowanie kryterium równomierności rozkładów pozwala wtedy kontrolować kontrast w przekazie informacji obrazowej.

[1] Są to jednostki ilościowej skali opisującej względną gęstość radiologiczną tkanki prześwietlanej promieniami rentgenowskimi; jest to wyrażona w promilach różnica liniowych współczynników osłabiania tkanki względem wody, odniesiona do różnicy współczynników osłabiania wody i powietrza

Rysunek 3.1: Przykładowe efekty regulacji kontrastu i jasności obrazów z wykorzystaniem przekształceń punktowych; w porządku od lewej do prawej, góra-dół mamy kolejno obraz źródłowy, korekcję gamma $\gamma = 2,2$, korekcję gamma $\gamma = 1/2,2$, negatyw, solaryzację oraz funkcję okna ustawioną na zakres $d = 0,4$, $g = 0,9$ znormalizowanej dynamiki (okno tak dobrano, by uwidocznić szczegóły twarzy poprzez poprawę kontrastu).

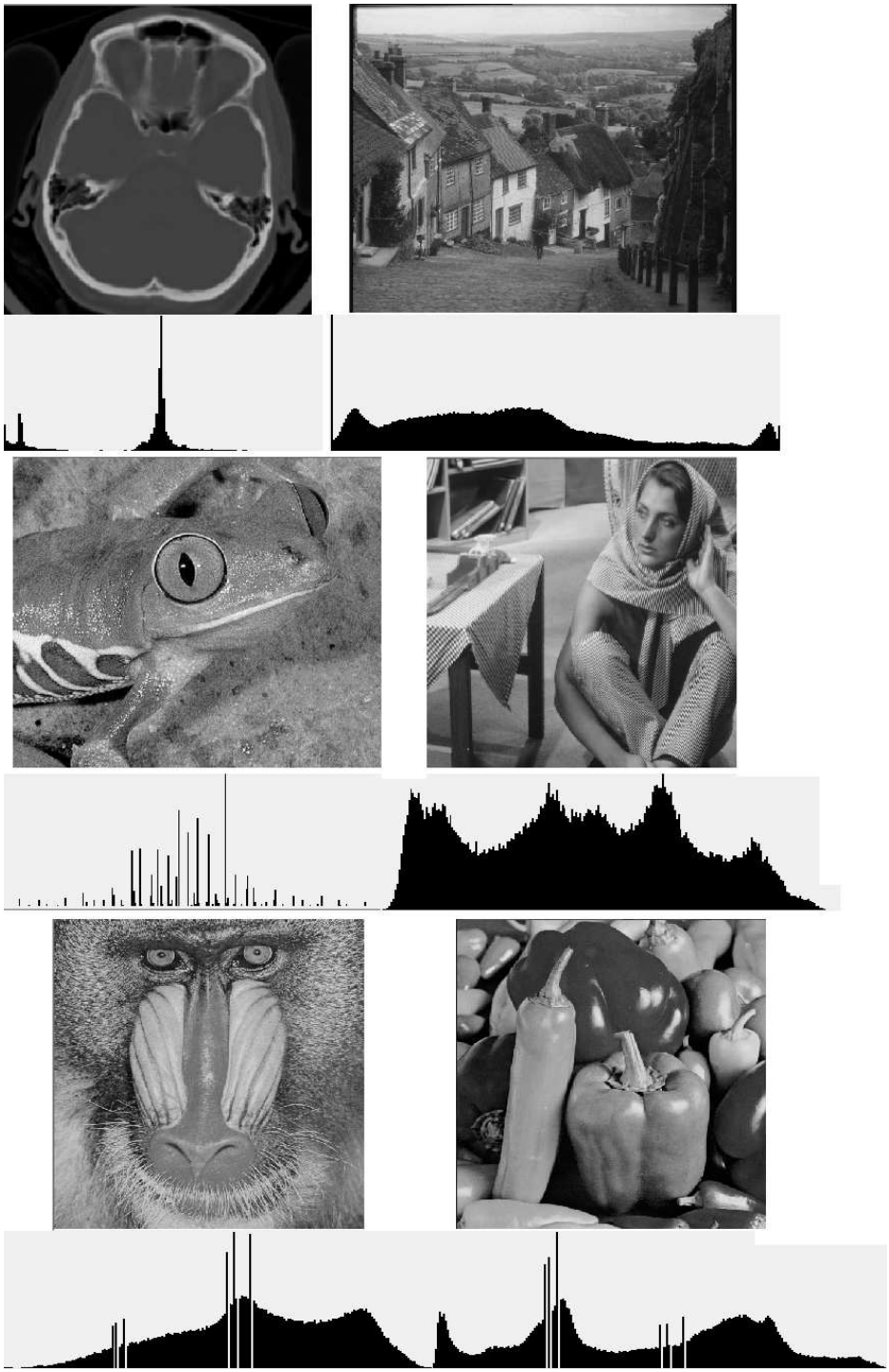


Rysunek 3.2: Dobór funkcji okna w przypadku obrazowej diagnostyki medycznej, którego celem jest lepsze uwidocznienie zmiany hipodensyjnej (ciemniejsza plama wskazana strzałką); w lewym górnym rogu zaznaczono parametry okna według stosowanej zazwyczaj konwencji - szerokość okna/środek okna, przy czym wartości te podane są w jednostkach Hounsfielda (HU); interesujący w diagnostyce wczesnych udarów niedokrwiennych zakres tkanki miękkiej mózgowia wynosi zwykle 10-50 HU.

Przykładowe histogramy naturalnych obrazów testowych, świadczące o ich globalnym kontraście przedstawiono na rys. 3.3.

Podstawowymi operacjami wykonywanymi na histogramie, które służą poprawie kontrastu w obrazie są: a) rozciąganie, czyli rozszerzenie histogramu na cały zakres możliwych wartości alfabetu – od głębokiej czerni do wysyconej bieli, b) wyrównywanie, czyli przekształcanie histogramu do postaci rozkładu możliwie równomiernego, c) przekształcanie histogramu do założonej *a priori* postaci rozkładu nierównomiernego. Dyskretna – ziarnista postać histogramu stanowi realne ograniczenie przy próbach korekcji histogramu w celu poprawy poziomu skontrastowania obrazów. Poprawa oznacza tutaj przybliżenie kształtu histogramu obrazu źródłowego do rozkładu zamierzonego, by uzyskać efekt większego zróżnicowania odczytywanej treści.

Rozciąganie histogramu do zamierzonego przedziału wartości pikseli $[0, a_{M-1}]$

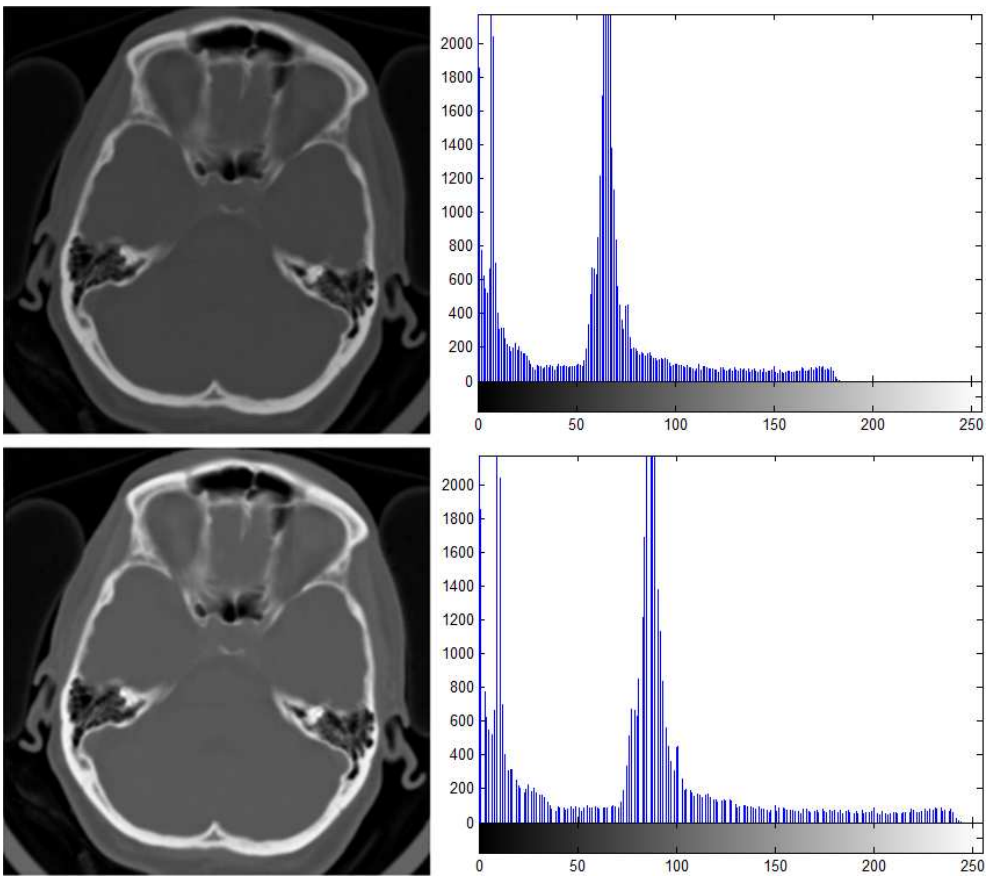


Rysunek 3.3: Przykładowe histogramy wybranych obrazów naturalnych, o zróżnicowanych walorach kontrastu globalnego (pod każdym z obrazów umieszczono jego histogram w zakresie wartości 0–255).

oznacza przeliczenie wartości pikseli według prostej formuły:

$$\bar{f}(k,l) = \frac{f(k,l) - \min_{\mathbf{f}}\{f(k,l)\}}{\max_{\mathbf{f}}\{f(k,l)\} - \min_{\mathbf{f}}\{f(k,l)\}} \cdot a_{M-1} \tag{3.6}$$

powodując pełne wykorzystanie dynamiki dopuszczonej formatem źródłowym, a opisanej alfabetem $A_{\mathbf{f}}$. Przykładowy efekt rozjaśnienia obrazu poprzez rozciąganie histogramu ukazano na rys. 3.4.



Rysunek 3.4: Efekt rozciągania histogramu medycznego obrazu testowego (wybranej warstwy badania tomografii komputerowej głowy) – u góry obraz źródłowy z histogramem wykazującym brak wartości pikseli w zakresie poziomów najjaśniejszych, u dołu - obraz ze zwiększonym kontrastem wskutek rozciągnięcia histogramu na pełen zakres dopuszczalnych wartości.

Wyrównywanie tudzież przekształcenie histogramu do innej, zamierzonej z góry postaci bazuje na przybliżonej znormalizowanym histogramem funkcji gęstości prawdopodobieństwa i wykorzystuje zasadę równoważenia prawdopodobieństw skumulowanych dla dwóch odmiennych rozkładów: wartości źródłowych oraz docelowego.

W znormalizowanym histogramie liczba zliczeń poszczególnych poziomów jasności odniesiona jest do liczby wszystkich pikseli obrazu: $H_{\mathbf{f}} = \sum_{m=0}^{M-1} h(m)$, służąc jako przybliżenie prawdopodobieństw wystąpienia kolejnych wartości alfabetu: $p(m) = h(m)/H_{\mathbf{f}}$, tak że rozumiemy $p(m) = p(a_m) = Pr(a_m) = Pr(f = a_m)$. Zbiór $\{p(m)\}_0^{M-1}$ taki że $\sum_{m=0}^{M-1} p(m) = 1$ wykorzystywany jest w analizie statystycznej do estymacji funkcji gęstości prawdopodobieństwa metodą częstościową, a to z kolei pozwala konstruować stochastyczne metody modelowania i przetwarzania obrazów.

Problem wyrównania histogramu (ogólniej dopasowania postaci histogramu do zamierzonej formy) sprowadza się do przekształcenia $S : \mathbf{f} \rightarrow \mathbf{g}$ obrazu źródłowego, opisanego znormalizowanym histogramem $\{p_f(m)\}$, w obraz o histogramie równomiernym (*de facto* zbliżonym do równomiernego) lub innym zamierzonym $\{p_g(n)\}$. W tym celu wykorzystuje się skumulowany histogram obrazu jako "sumator" kolejnych wartości $p_f(m)$ porównywanych z poziomami skumulowanego rozkładu prawdopodobieństw $p_g(n)$. Indeksy m i n przebiegają przez kolejne elementy alfabetów odpowiednio $A_{\mathbf{f}} = \{a_0, \ldots, a_{M-1}\}$ i $A_{\mathbf{g}} = \{b_0, \ldots, b_{N-1}\}$.

Precyzyjniej, skumulowany histogram estymuje skumulowane prawdopodobieństwo (inaczej dyskretną dystrybuantę) dla rozkładu prawdopodobieństw poziomów jasności postaci:

$$P(f \leqslant a_m) = P_f(a_m) = \sum_{k=0}^{m} p_f(k) \tag{3.7}$$

Na podstawie histogramu zamierzonego można analogicznie ustalić dyskretną dystrybuantę $P_g(b_n) = \sum_{l=0}^{n} p_g(l)$. Ponieważ dystrybuanty obu rozkładów są monotonicznie rosnące, dla ustalonego n można dobrać takie m, które zapewni równość obu dystrybuant: $P_g(b_n)$ oraz $P_f(a_m)$, co można rozpisać jako

$$\sum_{l=0}^{n} p_g(b_l) \cong \sum_{k=0}^{m} p_f(a_k) \tag{3.8}$$

Przybliżona równość pomiędzy skumulowanymi histogramami jest zwykle rozumiana jako $P_f(a_m)$ najbliższy $P_g(b_n)$ – nie sposób bowiem dokładnie dopasować skokowych zmian wartości dyskretnych dystrybuant obu rozkładów prawdopodobieństwa. Na tej podstawie można ustalić regułę wyznaczenia poziomów jasności obrazu przekształconego w sposób następujący: $b_n = P_g^{-1}(P_f(a_m)) = S(a_m)$.

W przypadku rozkładów ciągłych warunek równości dystrybuant

$$P_g(b_n) = P_f(a_m) \tag{3.9}$$

realizowany jest z dowolną precyzją, podobnie jak przekształcony histogram może dokładnie odpowiadać zamierzonemu. Zamiast alfabetów mamy wtedy przedziały $f = a_m \in [a_0, a_{M-1}]$ oraz $g = b_n \in [b_0, b_{N-1}]$, co daje rozwiniętą postać warunku (3.9): $\int_0^{b_n} p_g(g)dg = \int_0^{b_m} p_f(f)df$, czyli analogicznie $(g = b_n) = P_g^{-1}(P_f(a_m))$.

W przypadku równomiernego rozkładu $P_g(b_n)$ można przyjąć (dla uproszczenia rozważań), że określono jego postać przy znormalizowanych wartościach poziomów jasności $g \in [0,1]$, co daje $P(g \leqslant 1) = 1$, a $P(g \leqslant b_n) = P_g(b_n) = b_n$. Na podstawie (3.9) można wtedy zapisać, że znormalizowane

$$b_n = S(a_m) = P_f(a_m). \tag{3.10}$$

Bardziej ogólnie można zapisać $g = P_f(f) \cdot (b_{N-1} - b_0) + b_0$. Analityczna postać funkcji przekształceń obrazów według docelowych postaci ich histogramów w przypadku np.rozkładu wykładniczego postaci (za [154])

$$P_g(g) = \alpha \exp\{-\alpha(g - b_0)\} \tag{3.11}$$

wygląda następująco:

$$g = b_0 - 1/\alpha \ln[1 - P_f(f)] \tag{3.12}$$

zaś dla rozkładu Rayleigha

$$P_g(g) = \frac{g - b_0}{\alpha^2} \exp\Big\{-\frac{(g - b_0)^2}{2\alpha^2}\Big\} \tag{3.13}$$

mamy

$$g = b_0 + \Big[2\alpha^2 \ln(\frac{1}{1 - P_f(f)})\Big]^{1/2} \tag{3.14}$$

W przypadku rzeczywistych histogramów dyskretnych stosowanie zależności analitycznych jest rzadko przydatne. Korzystając z ustaleń (3.10) można jednak zaproponować prosty algorytm wyrównywania histogramu – algorytm 3.1.

Algorytm 3.1 *Wyrównanie histogramu obrazu*

1. Wyznacz histogram obrazu źródłowego $\mathbf{f}$, taki że
   $f(k,l) = a_m \Rightarrow h_f(m) = h_f(m) + 1$ na podstawie wartości wszystkich $H_\mathbf{f}$ pikseli obrazu, takich że $f(k,l) \in A_\mathbf{f}$;

2. Wyznacz źródłowy histogram skumulowany, taki że
   $P_f(a_m) = P_f(a_{m-1}) + h_f(m)$;

3. Oblicz wartości funkcji jasności obrazu docelowego, przyporządkowane poszczególnym a_m według zależności

   $$b_n = \Big\lceil P_f(a_m) \cdot b_{N-1} \Big\rfloor \tag{3.15}$$

   gdzie operator $\lceil \rfloor$ oznacza przybliżenie do najbliższego symbolu alfabetu $A_\mathbf{g}$;

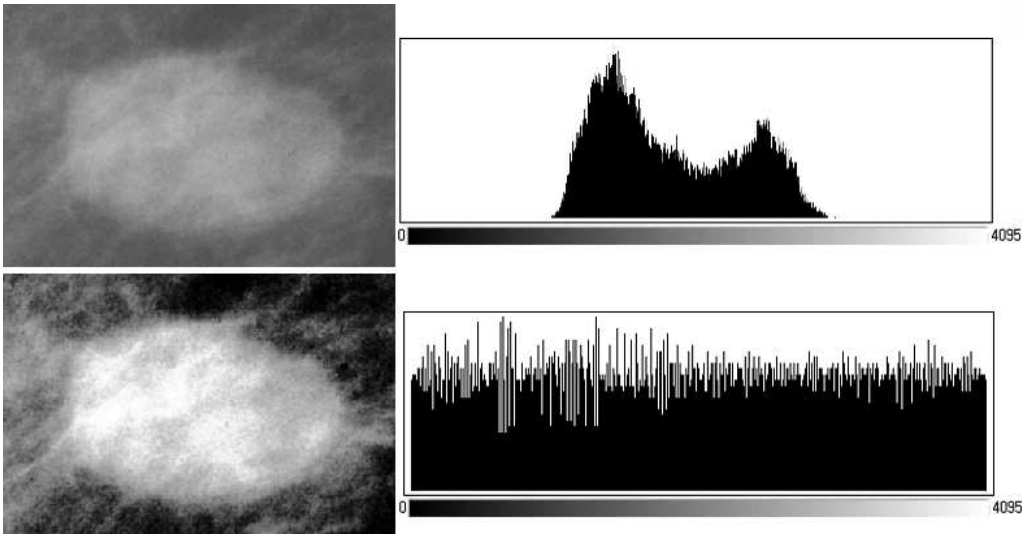
4. Wyznacz postać obrazu przetworzonego $\mathbf{g}$, tak że $\forall_{(k,l)} f(k,l) = a_m \Rightarrow g(k,l) = b_n$ według (3.15). Zakończ.

□

Przykładowy efekt wyrównywania histogramu przedstawiono na rys. 3.5. Pomimo tego, iż obraz źródłowy był dość dobrze skontrastowany i histogram niewiele odbiegał od równomiernego, uzyskano wyraźny efekt wyostrzenia kontrastu – co zresztą w tym przypadku niekoniecznie oznacza poprawę jakości z punktu widzenia obserwatora. Bardziej przekonujący efekt ulepszenia obrazu poprzez wyrównanie histogramu widać na rys. 3.6. Przekonujące wydobycie szeregu niewidocznych informacji uzyskano na rys. 3.7 - u góry, zaś mało korzystny efekt utraty subtelnych zróżnicowań jasności o charakterze ciągłym (rys. 3.7 - u dołu) wskazuje, że nie w każdym przypadku wyrównywanie histogramu jest korzystne.

Rysunek 3.5: Efekt wyrównania histogramu obrazu testowego lena – po lewej obraz źródłowy z histogramem, po prawej - obraz z ulepszonym kontrastem wskutek wyrównania histogramu.

Aby przekształcić histogram obrazu do innej niż równomierna, założonej z góry formy, należy złożyć dwa rodzaje przekształceń: obrazu źródłowego do pomocniczej formy pośredniej o histogramie w przybliżeniu równomiernym: $S_1 : \mathbf{f} \rightarrow \mathbf{p}_r$ oraz obrazu docelowego do analogicznej formy pośredniej: $S_2 : \mathbf{g} \rightarrow \mathbf{p}_r$. Wtedy uzyskamy $S_2^{-1} \cdot S_1 : \mathbf{f} \rightarrow \mathbf{g}$. Przykładowy efekt przekształcenia obrazu źródłowego według docelowego histogramu opisanego rozkładem Rayleigha (postaci (3.13)) zamieszczono na rys. 3.8. Rozciągnięcie informacji z zakresu poziomów jasnych



Rysunek 3.6: Wyrównywanie histogramu fragmentu mammogramu z guzem spikularnym – u góry obraz źródłowy z histogramem, u dołu - obraz z wyrównanym histogramem.

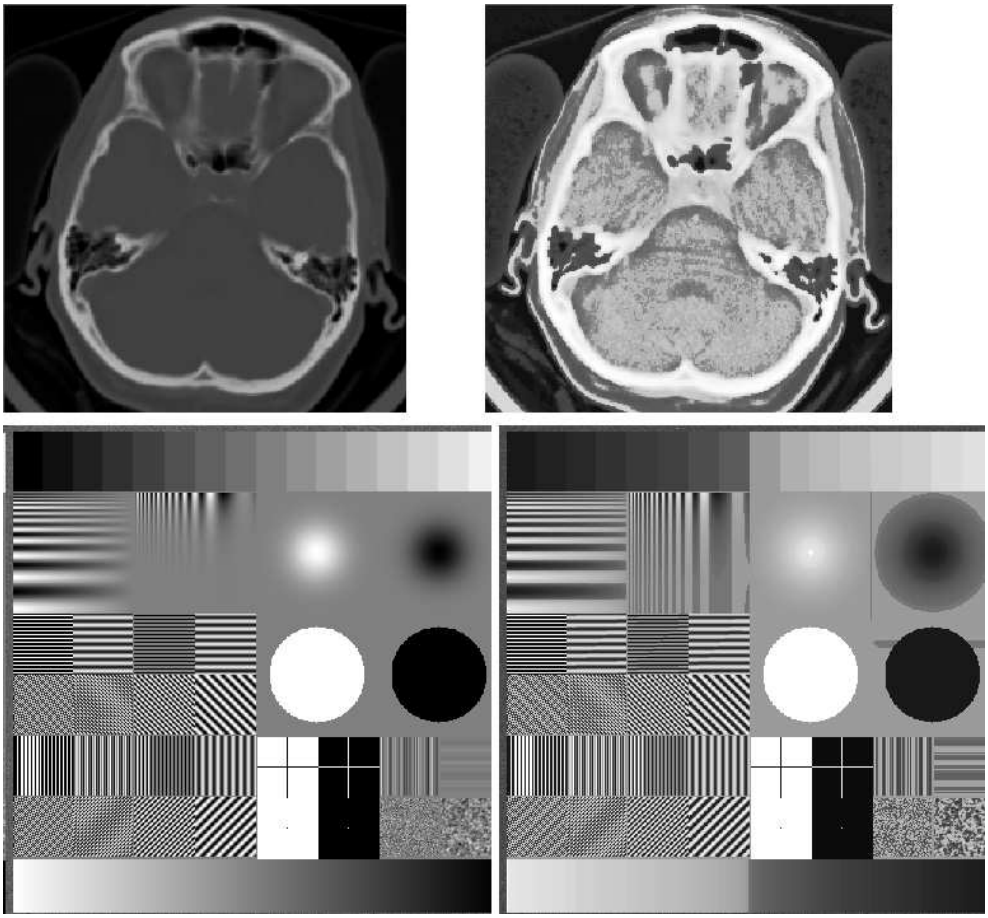
pozwala lepiej uwidocznić strukturę patologii - guza spikularnego, czyli obiektu diagnostycznego zainteresowania. Na rys. 3.9 znajduje się inny przykład uwidocznienia guza poprzez binaryzację bardzo wąskiego histogramu źródłowego, liczonego w wybranym regionie zainteresowania.

Innym sposobem korekty histogramu do zadanej empirycznie postaci jest odniesienie do kontrastowej charakterystyki wybranych regionów obrazu i na tej podstawie korygowanie histogramu całego obrazu np. według algorytmu 3.1. Na rys. 3.10 ukazano efekty wyrównywania histogramu na podstawie lokalnej statystyki wartości pikseli. W takim przypadku można uzyskać efekt silniejszego rozjaśnienia obrazu, czy wręcz niemal binaryzacji histogramu docelowego.

Jeszcze inną kategorię stanowią metody adaptacyjnego wyrównywania histogramu – AHE (*Adaptive Histogram Equalization*) [156] na podstawie lokalnych estymat histogramu liczonych w blokach przyległych lub zachodzących na siebie, o dobranych rozmiarach, z interpolacją zapeniającą ciągłość funkcji jasności na granicach obszarów przekształcanych niezależnie. Dodatkowo stosowane są metody ograniczania kontrastu (dynamiki) – CLAHE, czyli *Contrast-limited Adaptive Histogram Equalization* [157].

Filtracja liniowa - splotowa

Ważną kategorię metod ulepszania danych stanowią przekształcenia realizowane w przestrzeni danych za pomocą operatorów lokalnych. Obliczenia wykonywane na danych źródłowych zależą tutaj jedynie od wartości punktów sąsiednich, występujących w pewnym, najbliższym otoczeniu przekształcanej serii danych.



Rysunek 3.7: Efekt wyrównania histogramu obrazów testowych tomografii komputerowej oraz "target" – po lewej obraz źródłowy, po prawej - obraz ze zwiększonym kontrastem wskutek wyrównania histogramu.

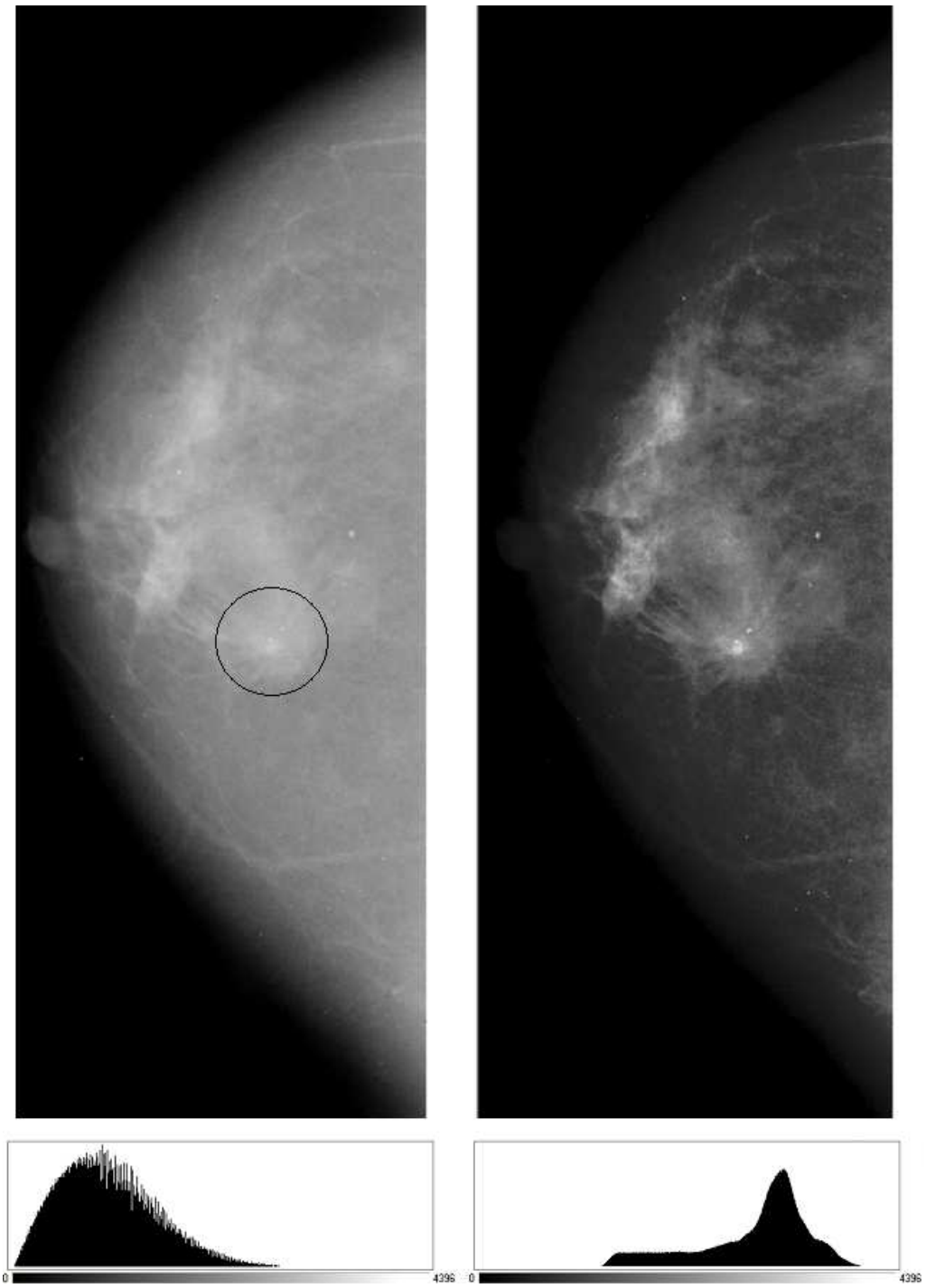
Szczególną rolę stanowią w tym przypadku operatory filtracji liniowej $\mathcal{O}_l$, tj. spełniającej dwa podstawowe warunki przekształceń liniowych

- addytywności: $\mathcal{O}_l(f+g) = \mathcal{O}_l(f) + \mathcal{O}_l(g)$,
- jednorodności: $\mathcal{O}_l(\alpha \cdot f) = \alpha \cdot \mathcal{O}_l(f)$.

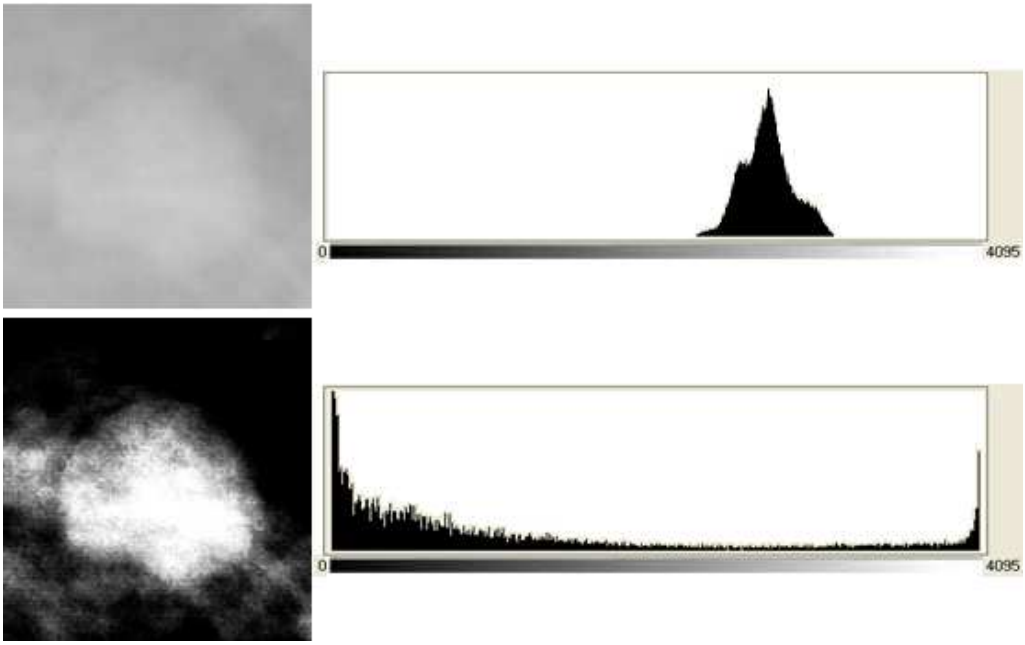
Filtracja liniowa w przestrzeni danych źródłowych jest przekształceniem lokalnym, kontekstowym, wykorzystującym operację splotu sygnału (funkcji sygnału f) ze skończoną funkcją odpowiedzi impulsowej filtru h. Operację splotu opisuje zależność

$$f * h(t) = \int_{D_h} f(t-x)h(x)\,\mathrm{d}x \tag{3.16}$$

gdzie D_h jest dziedziną określoności filtru (poza tą dziedziną odpowiedź impulsowa filtru jest zerowa).



Rysunek 3.8: Przekształcenie mammogramu z zaznaczonym guzem spikularnym w celu uzyskania postaci histogramu zbliżonej do rozkładu Rayleigha – po lewej obraz źródłowy z histogramem (liczonym jedynie dla obszaru sutka), po prawej - obraz przekształcony z dopasowanym histogramem. Przykład zaczerpnięty z [155].



Rysunek 3.9: Przekształcenie fragmentu mammogramu z prawie niewidocznym guzem celem binaryzacji histogramu docelowego – u góry obraz źródłowy z histogramem, poniżej - obraz przekształcony z dopasowanym histogramem.

W przypadku sygnałów cyfrowych $f(k),\ k \in \mathbb{Z}$ filtr jest rozumiany jako operator liniowy, niezmienniczy względem przesunięcia, przy czym przesunięcie jest ograniczone do dyskretnej siatki spróbkowanej dziedziny sygnału z przedziałem próbkowania równym 1. Pozwala to opisać filtr za pomocą zestawu (wektora) współczynników $\{h_n\}_0^{N-1}$, stanowiących dyskretną, skończoną – zwykle o niewielkich rozmiarach w stosunku do dziedziny filtrowanego sygnału $N << |\Omega_f|$, postać odpowiedzi impulsowej filtru.

Operację filtracji liniowej przedstawia więc zależność

$$f * h(k) = \sum_n f(k-n)h_n \tag{3.17}$$

Przy filtracji obrazów wykorzystuje się analogiczną – dwuwymiarową realizację splotu macierzy obrazu z maską – macierzą współczynników filtru $h_{m,n}$, przy czym nośnik filtru $(m,n) \in W_h$ jest zwarty, tj. domknięty i ograniczony do najbliższego sąsiedztwa w przestrzeni obrazu. Liniowa filtracja obrazów przekształca obraz źródłowy do postaci

$$g(k,l) = f * h(k,l) = \sum_{(m,n)\in W_h} f(i-m, j-n)h_{m,n} \tag{3.18}$$

Filtr można zdefiniować za pomocą macierzy współczynników wyrażających dyskretną odpowiedź impulsową filtru, o rozmiarze $(2R+1)\times(2R+1)$, indeksowanej

Rysunek 3.10: Efekt wyrównania histogramu obrazu testowego lena na podstawie lokalnego histogramu z obszaru wskazanego prostokątem – pod ulepszanymi (choć niekoniecznie ulepszonymi) obrazami zamieszczono ich przekształcone histogramy (w stosunku do źródłowego histogramu z rys. 3.5).

symetrycznie względem punktu centralnego $(0, 0)$

$$h_{m,n} = \begin{bmatrix} h_{-R,-R} & \dots & h_{0,-R} & \dots & h_{R,-R} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{-R,0} & \dots & h_{0,0} & \dots & h_{R,0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{-R,R} & \dots & h_{0,R} & \dots & h_{R,R} \end{bmatrix}$$

lub też za pomocą maski współczynników:

$h_{-R,-R}$	$\dots$	$h_{0,-R}$	$\dots$	$h_{R,-R}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$h_{-R,0}$	$\dots$	$h_{0,0}$	$\dots$	$h_{R,0}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$h_{-R,R}$	$\dots$	$h_{0,R}$	$\dots$	$h_{R,R}$

Istnieje wiele rodzajów filtrów, w tym przede wszystkim

- filtry dolnoprzepustowe, do wygładzania obrazu, redukcji szumów czy korekcji lokalnych nieciągłości funkcji jasności;

- filtry górnoprzepustowe, do usuwania składowej wolnozmiennej, do ekstrakcji szczegółów, gradientowe – do detekcji krawędzi i konturów;

- filtry pasmowo-przepustowe, wykorzystujące operatory laplasjanowe, maskowanie nieostrości (*unsharp masking*) czy filtracje kierunkową - do uwypuklania określonych cech, podkreślania krawędzi, uwidocznienia tekstury, wyostrzenia itp.

Filtry definiowane są przez ich maski, przy czym istotnych jest kilka wskazówek:

- wynik splotowych obliczeń według (3.18) należy zaokrąglić do najbliższej liczby całkowitej, gdyż $g(k,l) \in \mathbb{Z}$ jako wartość piksela;

- suma współczynników filtru zachowującego składową stałą winna wynosić 1 – dla wygody zwykle współczynniki filtru ustalane są jako liczby całkowite, a maskę filtru poprzedza mnożnik

$$\rho = \frac{1}{\sum_{(m,n)\in W_h} h_{m,n}} \tag{3.19}$$

- filtry odszumiające mają współczynniki dodatnie;

- suma współczynników filtrów górnoprzepustowych (wycinających składową stałą) wynosi 0;

- typowe kształty masek to bloki 3×3 lub 5×5, a także maski krzyżowe 5×5 lub prostokątne;

- przy filtracji na brzegach obrazu, kiedy to maska filtru pokrywa niezdefiniowany obszar sąsiedni pikseli granicznych, stosuje się zazwyczaj jedno z trzech typowych rozwiązań rozszerzenia dziedziny źródłowej (przyjmijmy $k = 1,\ldots,K, l = 1,\ldots,L$):

  - symetrycznie odbicie wartości pikseli względem granicy obrazu – $f(-1,\cdot) = f(1,\cdot)$, $f(-2,\cdot) = f(2,\cdot)$, $f(\cdot, L+1) = f(\cdot, L)$, $f(\cdot, L+2) = f(\cdot, L-1)$ itd.,
  - uzupełnienie zerami całego "brakującego" obszaru pokrytego przez maskę filtru,
  - cykliczne odbicie – piksele z początku są zawijane na koniec i odwrotnie, czyli $f(-1,\cdot) = f(K,\cdot)$, $f(-2,\cdot) = f(K-1,\cdot)$, $f(\cdot, L+1) = f(\cdot, 1)$, $f(\cdot, L+2) = f(\cdot, 2)$ itd.

Odszumianie. Podstawowym zastosowaniem filtracji splotowej jest odszumianie, czyli usuwanie bądź redukcja szumu maskującego użyteczną treść obrazową. Ze względu na losowy, nieobciążony charakter szumu najprostszą operacją odszumienia jest zastąpienie wartości pikseli średnią z pikseli sąsiednich w przestrzeni obrazu lub też średnią ważoną (z większą wagą przypisaną źródłowej wartości piksela i jego najbliższym sąsiadom). Procedurę tę można realizować filtracją splotową, dobierając współczynniki filtru jako wagi liczonej średniej i ustalając rozmiar maski filtru.

Ponieważ uniwersalny model szumu występującego w obrazach naturalnych zakłada wysoki stosunek energii sygnału użytecznego do szumu w zakresie niskich częstotliwości (składowe te dominują w typowej treści obrazowej), przy szumach zwykle dominujących przy częstotliwościach wyższych, receptą na redukcję szumu jest dolnoprzepustowa filtracja obrazu źródłowego z szumem akwizycji.

Przykładowe maski filtrów odszumiających (dolnoprzepustowych) mają następującą postać:

$$
\begin{aligned}
\text{uśredniający, równomierny:} \quad & h_{m,n} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \rho = 1/9 \\
\text{lekko nierównomierny:} \quad & h_{m,n} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \rho = 1/10 \\
\text{parametryczny ogólny:} \quad & h_{m,n} = \begin{bmatrix} 1 & p & 1 \\ p & p^2 & p \\ 1 & p & 1 \end{bmatrix}, \rho = \left(\frac{1}{p+2}\right)^2 \\
\text{parametryczny z } p = 2\text{:} \quad & h_{m,n} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \rho = 1/16 \\
\text{gaussowski z } \sigma = 1\text{:} \quad & h_{m,n} = \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}, \rho = 1/273
\end{aligned}
\tag{3.20}
$$

Przykładowe, nie zawsze korzystne efekty filtracji odszumiającej pokazano na rys. 3.11. Kryterium doboru filtrów jest kompromisem pomiędzy "siłą" odszumiania (większą redukcją szumów wskutek przesuwania zakresu widma zachowanego w kierunku coraz niższych częstotliwości), a zachowaniem krawędzi i szczegółów zawierających niewątpliwie wysokie częstotliwości w niewielkich procentowo obszarach. Alternatywnym do metod liniowych rozwiązaniem jest stosowanie filtracji nieliniowej – przede wszystkim filtrów medianowych.

Rysunek 3.11: Efekty odszumiana za pomocą liniowej filtracji splotowej; kolejno od lewej do prawej, góra-dół – obraz źródłowy lena, lena z równomiernym szumem addytywnym o amplitudzie 50 (średniokwadratowa różnica względem oryginału wynosi $mse = 831$) oraz obraz ten przekształcony za pomocą, kolejno (patrz definicje (3.20)), równomiernego filtru uśredniającego 3×3 ($mse = 675$), równomiernego filtru uśredniającego 7×7 ($mse = 746$), filtru parametrycznego dla $p = 2$ ($mse = 669$) oraz filtru gaussowskiego z $\sigma = 1$ ($mse = 668$).

Detekcja krawędzi. Kolejnym, istotnym obszarem zastosowań liniowej filtracji splotowej jest detekcja krawędzi, wykorzystująca tzw. filtry gradientowe (górnoprzepustowe). Metody detekcji krawędzi mogą być wykorzystywane do analizy obrazów w tzw. segmentacji konturowej, zaś w zastosowaniach służących ulepszaniu danych są przykładem ekstraktora lub wzmacniacza treści istotnych, selekcji informacji celem ich lepszej percepcji, uwydatnienia całościowych konturów definiujących obiekty.

Krawędź w sensie interpretacji treści obrazowej jest rozumiana jako fragment konturu, czyli granicy rozdzielającej obiekty bądź podobiekty, co ma znaczenie w zrozumieniu treści obrazu. Zakładając w pierwszym przybliżeniu ciągły model funkcji jasności, krawędź to zbiór punktów obrazu, w których występuje istotna nieciągłość funkcji jasności lub określonej cechy funkcji jasności w odniesieniu do kontekstu najbliższego w przestrzeni obrazu otoczenia.

Liczenie gradientu, tj. pola wektorowego $\nabla f(x,y) = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$ o określonym kierunek i module w poszczególnych punktach obrazu stanowi użyteczną charakterystykę rozkładu potencjalnych krawędzi i konturów w obrazie. Pole wektorowe gradientu wykorzystywane jest w niektórych metodach segmentacji konturowej obrazów, np. w koncepcji bazującej na poziomicach *level sets* [158].

W lokalnych metodach detekcji punktów krawędziowych bardziej przydatna jest estymacja wartości jednowymiarowego gradientu w punkcie (x,y) wzdłuż linii normalnej w stosunku do domniemanego kierunku krawędzi nachylonej do poziomu współrzędnych ortogonalnych pod kątem θ jako (za [154])

$$|\nabla f(x,y)| = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta \tag{3.21}$$

W przypadku obrazów cyfrowych mamy do czynienia z dyskretnymi modelami krawędzi, kiedy to gradienty są aproksymowane operatorami różnicowymi. Istotne przy tym są dwa zasadnicze elementy:

- **nośnik filtru różnicowego** (gradientowego), który określa obszar wyliczania gradientu (mniejszy nośnik oznacza wykrywanie w pierwszej kolejności "punktowych", czyli krótkich krawędzi, narożników czy wręcz pojedynczych punktów);
- **czułość kierunkowa filtru**, czyli zdolność do wykrywania krawędzi układających się w dowolnym kierunku płaszczyzny obrazu, z możliwie wiernym zachowaniem gładkości wykrywanego konturu danego obiektu;

Określenie wartości (modułu) gradientu krawędzi w przypadku dyskretnym realizowane jest zwykle poprzez wyznaczenie różnicowych składowych gradientu krawędzi w kierunku poziomym i pionowym, czyli odpowiednio wzdłuż wiersza i kolumny, do których należy punkt o współrzędnych (k,l) jako

$$\mathcal{G}(k,l) = \sqrt{[\mathcal{G}_r(k,l)]^2 + [\mathcal{G}_c(k,l)]^2} \tag{3.22}$$

lub w przybliżonej, lecz prostszej obliczeniowo wersji

$$\mathcal{G}(k,l) = |\mathcal{G}_r(k,l)| + |\mathcal{G}_c(k,l)| \tag{3.23}$$

podczas gdy orientacja gradientu względem osi poziomej wynosi

$$\theta(k,l) = \arctan \frac{\mathcal{G}_c(k,l)}{\mathcal{G}_r(k,l)} \tag{3.24}$$

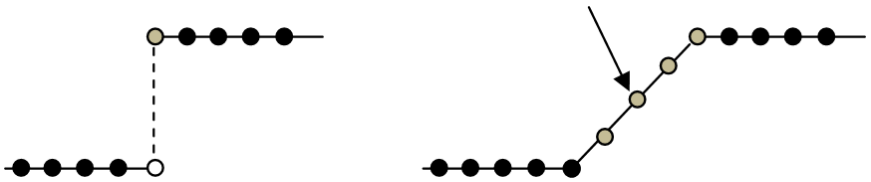
Wprowadzenie **selekcji wartości gradientu poprzez progowanie** pozwala uzyskać binarną mapę rozkładu punktów krawędziowych w obrazie.

Najprostszą formą wyznaczenia różnicowych składowych gradientu określają odpowiednio zależności $\mathcal{G}_r(k,l) = f(k,l) - f(i-1,j)$ oraz $\mathcal{G}_c(k,l) = f(k,l) - f(k,l+1)$. Można je obliczyć metodą filtracji splotowej za pomocą filtrów (według notacji przyjętej w (3.18) oraz (3.1.3), w tym założeń o nieparzystym rozmiarze maski filtru w obu wymiarach)

$$g^{(r)}_{m,n} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{oraz} \quad g^{(c)}_{m,n} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.25}$$

Średnia geometryczna oszacowanych różnicowo składowych poziomej i pionowej gradientu według (3.22) lub też suma modułów (3.23) pozwala oszacować wartość gradientu w każdym punkcie obrazu dając przestrzenny rozkład modułu gradientu $\mathcal{G}(k,l)$, przy czym $\mathcal{G}_r(k,l) = f * g^{(r)}(k,l)$ oraz $\mathcal{G}_c(k,l) = f * g^{(c)}(k,l)$.

Taka definicja estymatora składowych gradientu zachowuje największą możliwą lokalność obliczeń, w praktyce jednak działa skutecznie przede wszystkim w przypadku profilu krawędzi o charakterze skokowym, którego ciągłym modelem są osobliwe punkty krawędziowe – punkty nieciągłości, tj. nieciągłej zmiany wartości funkcji jasności (zobacz rys. 3.12).



Rysunek 3.12: Profile krawędzi w wersji dyskretnej (duże kropki) i ciągłej – po lewej krawędź skokowa, nieciągła, po prawej - krawędź ciągła, o łagodniejszym profilu typu *ramp*; prosty operator różnicowy liczy niezerowe gradienty w punktach oznaczonych jako szare kropki, co w przypadku krawędzi o mniejszym nachyleniu oznacza kilka punktów krawędziowych - niekiedy korzystnie jest ustalić jedynie środek krawędzi rozdzielającej obiekty (czyli punkt wskazany strzałką).

W przypadku krawędzi szerszych, bardziej rozmytych, o profilu charakteryzującym się mniejszym gradientem, opisanym np. funkcją *ramp* - rys. 3.12, operator

gradientowy według (3.25) wskaże wszystkie punkty profilu krawędzi. Da to nie zawsze korzystny efekt szerokiej krawędzi. Zwykle lepsze efekty uzyskuje się za pomocą filtrów premiujących jedynie środek szerszej krawędzi lub też punkt profilu o największym gradiencie. Innym problemem jest wybór jedynie tych punktów o istotnej wartości gradientu, które rzeczywiście należą do krawędzi rozdzielających obiekty, a nie są tylko lokalnymi wskaźnikami niejednorodności w rozkładzie funkcji jasności danego regionu. Poprawę skuteczności detekcji krawędzi można w niektórych przypadkach uzyskać poprzez stosowanie jednokierunkowych estymat krawędzi (według (3.21)) jako pojedynczych detektorów krawędzi poszukiwanych pod określonym kątem – np. filtrów $g^{(r)}$ oraz $g^{(c)}$ odpowiednio pod kątem $\theta = 0°$ i $\theta = 90°$.

Doskonalenie filtrów gradientowych może się więc odbywać w kilku zasadniczych kierunkach:

- zwiększenie skuteczności lokalizacji krawędzi o szerszym profilu i łagodniejszym nachyleniu, np. poprzez większy rozmiar maski i wprowadzenie centralnych wartości zerowych – zobacz filtry (3.26) oraz kolejne;

- zmniejszenie czułości na lokalne zmiany jasności obrazu nie będące krawędziami (a jedynie np. dynamiczną teksturą obiektu) poprzez liczenie gradientów uśrednionych (z ewentualnym ważeniem) po pewnym kontekście (uśrednienie odbywa się w kierunku prostopadłym do kierunku liczenia gradientu) – zobacz filtry Prewitta i Sobela definiowane przez (3.27) oraz (3.28);

- zwiększenie kątowej czułości detekcji poprzez sumowanie efektów filtracji za pomocą zestawu filtrów kierunkowych; do detekcji punktów krawędziowych może być stosowane oddzielne progowanie wartości jednokierunkowych gradientów dla poszczególnych kierunków lub też obliczana jest maksymalna wartość gradientu kierunkowego w punkcie, a odpowiadający mu kierunek wskazuje przebieg krawędzi; oddzielnie stosowane są też filtry do detekcji narożników, linii (rozdzielających obszar obiektów) czy pojedynczych punktów istotnych – zobacz (3.29)-(3.32).

Zestawy filtrów składowych gradientu oraz kierunkowych, często stosowanych do detekcji krawędzi są następujące:

- różnicowe z zerowym centrum

$$h^{(r)}_{m,n} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{oraz} \quad h^{(c)}_{m,n} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.26}$$

- Prewitta [159]

$$h^{(r)}_{m,n} = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ oraz } h^{(c)}_{m,n} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \text{ przy } \rho = 1/3 \tag{3.27}$$

- Sobela [160]

$$h^{(r)}_{m,n} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \text{ oraz } h^{(c)}_{m,n} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \text{ przy } \rho = 1/4 \tag{3.28}$$

- diagonalne Robertsa [161]

$$h^{(r)}_{m,n} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{oraz} \quad h^{(c)}_{m,n} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.29}$$

- diagonalne Sobela

$$h^{(d1)}_{m,n} = \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} \text{ oraz } h^{(d2)}_{m,n} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \text{ przy } \rho = 1/4 \tag{3.30}$$

- diagonalne linii

$$h^{(d1)}_{m,n} = \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{oraz} \quad h^{(d2)}_{m,n} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.31}$$

- kierunkowe Kirscha [162]

$$h^{(k1)}_{m,n} = \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}, h^{(k2)}_{m,n} = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}, h^{(k3)}_{m,n} = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix},$$
$$h^{(k3)}_{m,n} = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}, h^{(k5)}_{m,n} = \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}, h^{(k6)}_{m,n} = \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix},$$
$$h^{(k7)}_{m,n} = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, h^{(k8)}_{m,n} = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \text{ wszystkie przy } \rho = 1/15 \tag{3.32}$$

Liniowa filtracja jest pierwszym etapem detekcji krawędzi, po czym następuje zazwyczaj progowa – binarna klasyfikacja pikseli jako należące do krawędzi lub też nie. Generalnie, im wyższa wartość estymowanego gradientu, tym większe prawdopodobieństwo, że punkt należy do krawędzi. Niekiedy zależy to także od sąsiedztwa – jeżeli w określonym otoczeniu w przestrzeni obrazu znajdują się również punkty zaliczone do grupy krawędziowych, wtedy jest większa szansa na wykrycie ciągłej z natury krawędzi obiektu.

Dobór wartości progu wpływa na czułość metody detekcji, przy czym obniżanie wartości progu może jednocześnie pogorszyć specyficzność wykrywanych krawędzi – obok istotnych krawędzi mogą pojawić się wskazania fałszywe powodowane szumem lub też naturalną dynamiką tekstur obiektów występujących w obrazie lub zróżnicowaną charakterystyką tła. Można w tym celu wykorzystać np. lokalne estymatory wartości progu na podstawie sąsiedztwa punktów krawędziowych lub też globalne estymaty rozkładu gęstości prawdopodobieństwa wystąpienia krawędzi i jej braku według reguł klasyfikacji Bayesa. Przykładowy opis bardziej złożonej, hybrydowej metody progowania można znaleźć w [163].

Anizotropowe właściwości gradientu stanowią niekiedy utrudnienie przy wyznaczaniu wszystkich istotnych krawędzi w obrazie o nieznanej wcześniej charakterystyce – uzyskujemy bowiem rezultaty zależne od kąta obrotu (np. nieco obrócone ujecie kamery spowoduje zmiany w rozkładzie wykrywanych krawędzi).

Izotropowe, czyli niezmiennicze względem obrotu efekty detekcji krawędzi można uzyskać za pomocą pojedynczych, "bezkierunkowych" (tj. nie wyróżniających żadnego kierunku) **filtrów laplasjanowych**. Filtry te mają właściwość podkreślania czy też wzmacniania (eksponowania) krawędzi.

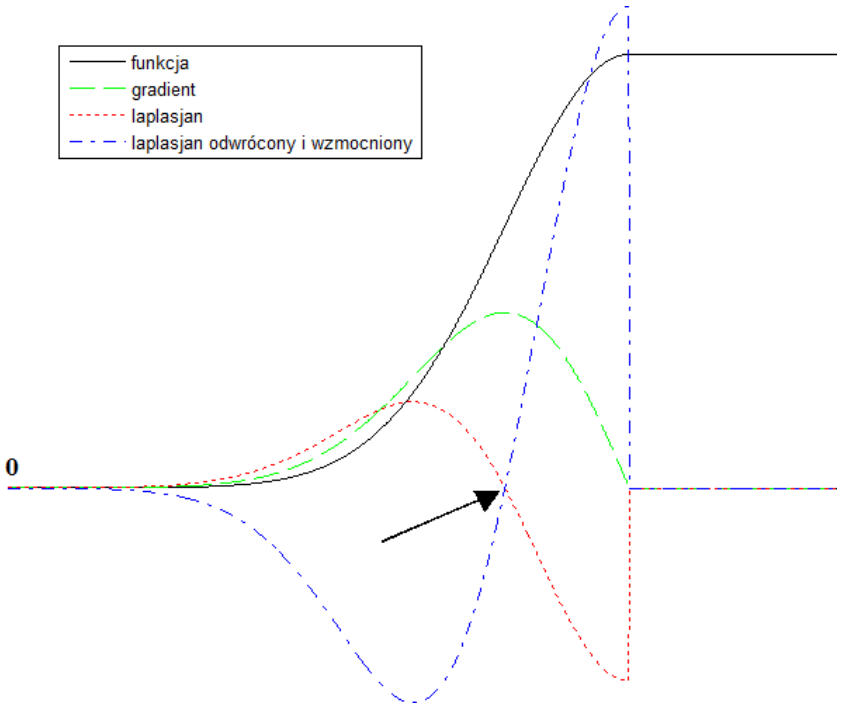
Podstawowa definicja laplasjanu (tj. operatora Laplace'a) sprowadza się do obliczenia skalarnej sumy drugich pochodnych cząstkowych dwukrotnie różniczkowalnej funkcji $f(x, y)$ (przypadek dwuwymiarowy):

$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.33}$$

Wykorzystanie analitycznej koncepcji zerowania się drugiej pochodnej w punktach przegięcia funkcji (tj. maksimach gradientu) pozwala dokładniej ustalić położenie środka krawędzi, a także poprzez splot z funkcją źródłową powoduje wzrost gradientu wokół punktu krawędzi - zobacz rys. 3.13.

Efekt detekcji i wzmocnienia krawędzi obrazów cyfrowych niezależnie od ich kierunku uzyskuje się za pomocą filtrów estymujących laplasjan na podstawie różnicowych przybliżeń drugiej pochodnej w ortogonalnych kierunkach kartezjańskiego układu współrzędnych obrazu. Najprościej można to zapisać jako

$$l_{m,n} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{3.34}$$



Rysunek 3.13: Przykładowy profil krawędzi (ciągły model połówki funkcji gaussowskiej), jego pierwsza i druga pochodna (model gradientu i laplasjanu) ze wskazaniem (strzałka) miejsca dziedziny ekstremum pochodnej i przejścia przez zero drugiej, tj. środka krawędzi.

przy czym niekiedy stosuje się dodatkowo współczynnik normalizujący $\rho = 1/4$. Wtedy operator Laplace'a $\mathcal{L}(k,l) = f * l(k,l)$.

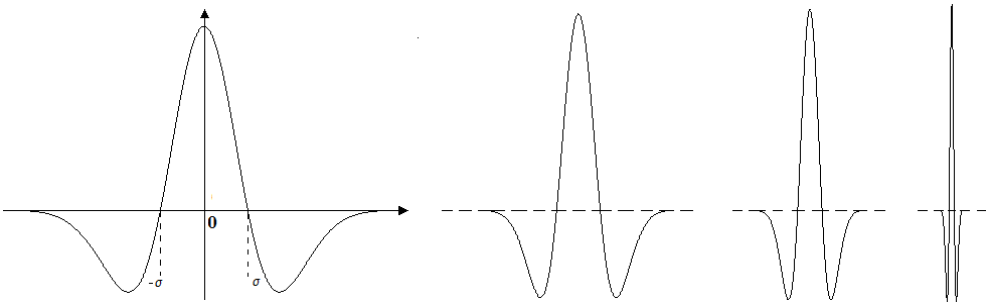
Poniżej przedstawiono kilka innych realizacji operatora Laplace' a stosowanych do detekcji krawędzi:

$$l_{m,n} = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix}, \quad l_{m,n} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix},$$
$$l_{m,n} = \begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}, \quad l_{m,n} = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \tag{3.35}$$

Często praktyką jest kaskadowe stosowania różnego typu filtrów, pozwalające uzyskać docelowo kilka efektów ulepszenia obrazów. Powszechnym zwyczajem w przypadku stosowania wysokoczęstotliwościowej obróbki obrazów naturalnych jest wcześniejsze ich odszumienie celem wzmocnienia jedynie treści użytecznej.

Korzystne w wielu zastosowaniach okazuje się złożenie operacji uśredniania filtrem gaussowskim (w celu redukcji szumu, a przez to fałszywych wskazań krawędzi) z izotropową filtracją laplasjanową, co korzystając z liniowości operacji

odpowiednich filtracji można zapisać jako: $\mathcal{L}oG(k,l) = f * l * h(k,l) = f * lh(k,l)$, gdzie postać filtru gaussowskiego $h_{m,n}$ przykładowo jak w (3.20), zaś postać ciągłej, jednowymiarowej odpowiedzi impulsowej filtru LoG, zależną od wartości odchylenia standardowego σ funkcji gaussowskiej pokazano na rys. 3.14.



Rysunek 3.14: Jednowymiarowa postać ciągłej odpowiedzi impulsowej filtru LoG, zależna od malejącej wartości odchylenia standardowego σ.

Cyfrowa postać dwuwymiarowych filtrów LoG może przybierać różne postacie, zależne przede wszystkim od σ i dokładności przybliżeń współczynników:

$$lh_{m,n} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{3.36}$$

$$lh_{m,n} = \begin{bmatrix} 0 & 0 & -1 & -2 & -2 & -2 & -1 & 0 & 0 \\ 0 & -2 & -3 & -5 & -5 & -5 & -3 & -2 & 0 \\ -1 & -3 & -5 & -3 & 0 & -3 & -5 & -3 & -1 \\ -2 & -5 & -3 & 12 & 23 & 12 & -3 & -5 & -2 \\ -2 & -5 & 0 & 23 & 40 & 23 & 0 & -5 & -2 \\ -2 & -5 & -3 & 12 & 23 & 12 & -3 & -5 & -2 \\ -1 & -3 & -5 & -3 & 0 & -3 & -5 & -3 & -1 \\ 0 & -2 & -3 & -5 & -5 & -5 & -3 & -2 & 0 \\ 0 & 0 & -1 & -2 & -2 & -2 & -1 & 0 & 0 \end{bmatrix} \tag{3.37}$$

Na rys. 3.15 i 3.16 przedstawiono przykładowe efekty filtracji służącej detekcji krawędzi.

Wzmacnianie krawędzi . Nieco odmiennym zastosowaniem operatorów Laplace'a jest podkreślanie krawędzi służące poprawie percepcji obrazu. Dzięki temu uzyskuje się efekt wyostrzenia obrazu – zobacz rys. 3.17.

Postać filtrów realizujących efekt wzmocnienia krawędzi jest zbliżona do (3.34) i (3.35) z tym że centralna w masce wartość współczynnika jest inkrementowana ze względu na konieczność zachowania zasadniczej treści obrazu wyrażonej w

Rysunek 3.15: Efekty splotowej filtracji służącej detekcji krawędzi (filtry gradientowe); kolejno od lewej do prawej, góra-dół – obraz źródłowy lena, po filtracji różnicowej (3.25), Prewitta (3.27), Sobela (3.28), Sobela z dodatkowymi filtrami diagonalnymi (3.30), Kirscha (3.32).

zakresie niskoczęstotliwościowym. Suma współczynników wynosi wtedy 1, co pozwala zachować dynamikę jasności obrazu źródłowego w obszarach wolnozmien-

Rysunek 3.16: Efekty splotowej filtracji służącej detekcji krawędzi (laplasjan); kolejno od lewej do prawej, góra-dół – obraz źródłowy lena po filtracji laplasjanowej według (3.34) oraz według (3.35) z wartością centralną 8, po filracji LoG 5×5 według (3.36) oraz LoG 9×9 (3.37).

nych, przykładowo:

$$h_{m,n} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{3.38}$$

$$h_{m,n} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{3.39}$$

$$h_{m,n} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 17 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{3.40}$$

Rysunek 3.17: Efekty splotowej filtracji służącej wzmocnieniu krawędzi za pomocą modyfikowanych filtrów laplasjanowych oraz maskowania nieostrości; kolejno od lewej do prawej, góra-dół – obraz źródłowy lena, lena po filtracji według (3.38), według (3.39), według (3.39), lena po maskowaniu nieostrości według (3.42) z filtrem laplasjanowym (według (3.35) z centralną 8) oraz według (3.41) z gaussowskim filtrem uśredniającym (3.20).

Silniejszy efekt wyostrzenia obrazu poprzez podkreślenie krawędzi i innych lokalnych zmian jasności o znaczącym gradiencie daje metoda maskowania nieostrości *unsharp masking*[164]. Zasadniczo jest ona realizowana poprzez

- wyznaczenie obrazu krawędzi wskutek odejmowanie od obrazu źródłowego nieostrej (rozmytej) jego maski: $f_{kraw}(k,l) = f - f_{mask}(k,l) = f - [f * h](k,l)$
- dodatnie do obrazu źródłowego obrazu krawędzi w odpowiedniej proporcji, czyli

$$f_{ostry}(k,l) = f + \alpha \cdot f_{kraw}(k,l) \tag{3.41}$$

Wzrost wartości stałej α oraz rozmiaru maski filtru dolnoprzepustowego $h_{m,n}$ nasila efekt wyostrzenia obrazu (rys. 3.17).

Bezpośrednią zasadę wzmocnienia treści wysokoczęstotliwościowej określa reguła:

$$g(k,l) = f + \alpha \cdot [f * g](k,l) = f + \alpha \cdot f_g(k,l) \tag{3.42}$$

z bezpośrednim wykorzystaniem określonych filtrów gradientowych $g_{m,n}$ do generacji obrazu krawędziowego.

Nieliniowa filtracja kontekstowa

Nieliniowa filtracja na bazie kontekstu wystąpienia danego piksela zamiast splotu wykorzystuje określone parametry lokalnego rozkładu wartości. Często jest to obliczanie mediany wartości pikseli określonego sąsiedztwa W_c (szerzej o medianowym uśrednianiu danych w [165, 166]) za pomocą przesuwanego okna filtru nieparzystego – wartość mediany zastępuje poziom jasności centralnego piksela tegoż okna. Analogicznie konstruowane są algorytmy filtracji z wykorzystaniem wartości maksimum, minimum lub modalnej (najczęściej występującej) w celu poprawy percepcji treści obrazowej. Można to zapisać ogólnie jako

$$g(k,l) = dist\_param\{f(i-m, j-n), \quad (m,n) \in W_c\} \tag{3.43}$$

gdzie *dist_param* przyjmuje postać 'median' (filtr medianowy), 'max' (filtr maksymalny), 'min' (filtr minimalny), 'mode' (filtr modalny).

W szczególności stosowane są

- filtracja medianowa w celu znaczącego odszumienia obrazu z możliwie małym efektem rozmycia krawędzi (lepiej radzi sobie z szumem impulsowym, nieco gorzej z gaussowskim, a przy szumie niesymetrycznym ulega zmianie średnia intensywność obrazu), najlepiej przy zachowaniu źródłowej wyrazistości krawędzi obiektów – zachowuje rozdzielczość oryginału, przy czym krawędź powinna obejmować przynajmniej połowę pikseli kontekstu, dlatego też rozmiar maski filtru powinien być dostosowany do wielkości krawędzi istotnych w obrazie; wartość piksela zastępowana jest medianą zbioru

wartości pikseli kontekstu (uproszczona procedura polega na ustawieniu pikseli sąsiedztwa w porządku wartości niemalejących – wartość środkowa tego szeregu stanowi poszukiwaną medianę); ze względu na dużą złożoność obliczeniową filtry te przybliżane są kombinacją filtrów maksymalnych i minimalnych [167];

- odszumiająca filtracja modalna, która dobrze redukuje niewielki szum w pobliżu krawędzi i wzmacnia krawędzie w silnie zaszumionym obszarze; może jednak powodować pewne zniekształcenia, przesunięcia krawędzi;

- filtracja adaptacyjna, czyli dostosowana do lokalnych właściwości obrazu; postać filtru zależy od lokalnej mediany, średniej lub wariancji, niekiedy dopuszczają interaktywną korekcje parametrów; możliwość dostosowania filtrów do lokalnych wymagań aplikacja okupiona jest jednak zwykle dużą złożonością obliczeniową; przykładowo

  - selektywna filtracja medianowa, gdzie przy obliczaniu mediany w określonym kontekście zbiór pikseli sąsiednich jest ograniczony do tych, których różnica poziomów jasności z pikselem centralnym jest niewiększa od ustalonego progu; daje to możliwość kontroli siły rozmycia krawędzi – mały próg silniej zachowuje krawędzie, jednak kosztem słabszego efektu odszumiania [172];
  - selektywna filtracja modalna, wyłączająca z lokalnej analizy wartości z określonego przedziału – liczenie wartości modalnej odbywa się wtedy jedynie w użytecznym zakresie wartości, podczas gdy w przypadkach całego kontekstu wypełnionego wartościami nieużytecznymi wstawiana jest wartość modalna najbliższego bloku; można w ten sposób dodatkowo wzmocnić redukcję szumu;
  - wyostrzanie krawędzi ekstremami jasności (*'extremum sharpening'*) – wartością operatora ekstremum jest minimum albo maximum z okna sąsiedztwa o środku w analizowanym punkcie w zależności od tego, które z nich jest bliższe wartości funkcji jasności w tym punkcie; jeśli wartość funkcji w punkcie jest dokładnie w połowie pomiędzy wartościami ekstremów, wartość pozostaje niezmieniona; rozmiar kontekstu powinien być dobierany zależnie od szerokości krawędzi w obrazie; zwykle operacja ta jest uzupełniona filtracja medianową, by nieco wygładzić podkreślone krawędzie [170, 171];

- filtry morfologiczne, będące złożeniem operacji maksimum oraz minimum w kontekście definiowanym przez element strukturujący; usuwają skutecznie drobne elementy zakłócające, niepotrzebne struktury, uzupełniają kształty niepełnych struktur, pomagają wyznaczyć zamknięte struktury, wyznaczyć

szkielet obiektów itd.; w typowej postaci nie zachowują średniej intensywności obrazu; podstawowe zastosowania w ulepszaniu obrazów to odszumianie (operatory otwarcia-zamknięcia, filtry kaskadowe) i wyostrzanie krawędzi (adaptacyjne przełączanie pomiędzy erozją i dylacją)[168]; mają także szerokie zastosowanie w analizie obrazów [169];

Możliwe jest kaskadowe stosowanie wybranych metod filtracji, np. odszumiającego filtru medianowego, dalej wyostrzenie ekstremami jasności czy selektywnej filtracji medianowej zwiększającego wyrazistość krawędzi. Przykładowe efekty stosownia wybranych filtrów nieliniowych przedstawiono na rys. 3.18.

Filtracje częstotliwościowe

Alternatywnym do przestrzennego sposobem filtracji są metody częstotliwościowe, co wynika przede wszystkim ze skutecznej formy analizy sygnałów oraz wygodnych metod projektowania filtrów za pomocą reprezentacji częstotliwościowej. Wykorzystuje się tutaj fakt, że liniowej filtracji splotowej w przestrzeni obrazu jest równoważna filtracja realizowana w fourierowskiej dziedzinie częstotliwościowej jako iloczyn transformat Fouriera sygnału i odpowiedzi impulsowej filtru.

W podrozdziale 2.2.2, strona 70, zdefiniowano trygonometryczne szeregi fourierowskie (2.1) w postaci trygonometrycznej oraz warunki dokładnego opisu sygnałów (warunki Dirichleta) za pomocą takiej reprezentacji.

Mając sygnał T okresowy $f(t)$ (tj. $f(t) = f(t+T)$) dla $t \in (-\infty, \infty)$, można go przedstawić korzystając z syntetycznej, zespolonej postaci wykładniczej szeregów Fouriera (przypomnienie (2.2))

$$f(t) = \sum_{k \in \mathbb{Z}} b_k e^{ji\omega_0 t} \tag{3.44}$$

gdzie $b_i = 1/T \int_{-T/2}^{T/2} f(t) e^{-jk\omega_0 t} dt$.

Transformacja Fouriera jest rozszerzeniem koncepcji wykładniczych szeregów Fouriera na klasę wszystkich sygnałów $f \in \mathrm{L}^2(\mathbb{R})$, w tym nieokresowych. Jest więc to przekształcenie $\mathcal{F} : \mathrm{L}^2(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$ zakładające szczególne rozumienie sygnałów nieokresowych.

Jeśli $f(t)$ jest nieokresowy i rozciąga się w całej dziedzinie liczb rzeczywistych, wtedy traktowany jest jako granicę skończonych sygnałów okresowych z $T \to \infty$, gdzie $\omega_0 \to 0$, co prowadzi do modelowego uciąglenia częstotliwości $k\omega_0 \to \omega$. Stąd zamiast sumy w (3.44) należy zapisać całkę, co prowadzi do ciągłego zbioru współczynników $F(\omega)$, takich że

$$F(\omega) = \int_{-\infty}^{\infty} f(t) cos\omega t dt + j \int_{-\infty}^{\infty} f(t) sin\omega t dt = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{3.45}$$

Rysunek 3.18: Przykładowe efekty stosowania filtrów nieliniowych; kolejno od lewej do prawej, góra-dół – obraz źródłowy lena, lena silnie pokryta szumem impulsowym ("sól i pieprz"), odszumienie za pomocą filtru medianowego o dużej masce 9×9 z silnym rozmyciem obrazu, nieskuteczność filtru medianowego o mniejszej masce 3×3, wykorzystanie selektywnego filtru modalnego zawężającego zakres wartości użytecznych, zastosowanie kaskady filtrów modalnego 2×2 z selekcją oraz medianowego 5×5.

Wtedy sygnał reprezentowany jest jako

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \tag{3.46}$$

Wyrażenie (3.45) jest **transformacją Fouriera** dająca ciągłe widmo częstotliwościowe sygnału, zaś (3.46) jest odwrotnym przekształceniem Fouriera.

Ponieważ funkcje rozwinięcia fourierowskiego, w tym przypadku funkcje bazowe transformacji Fouriera są zespolone, mamy więc transformatę $F(\omega) \in \mathbb{C}$, czyli zespolone widmo fourierowskie z częścią rzeczywistą $\Re(F(\omega))$ i urojoną $\Im(F(\omega))$, amplitudą $|F(\omega)|$ i fazą $\angle F(\omega)$:

$$F(\omega) = \Re(F(\omega)) + \Im(F(\omega)) = |F(\omega)| e^{j\angle F(\omega)} \tag{3.47}$$

gdzie $|F(\omega)| = \sqrt{\Re^2(F(\omega)) + \Im^2(F(\omega))}$ oraz $\angle F(\omega) = \arctan \frac{\Im(F(\omega))}{\Re(F(\omega))}$.

W przypadku sygnałów nieokresowych, określonych w skończonym przedziale $-[T/2, T/2]$ przyjmuje się okresowe (cykliczne) rozszerzenie (powielenie) tego sygnału na przedział $(-\infty, \infty)$ z okresem T. Wtedy, podobnie jak dla sygnałów okresowych rozciągniętych na całą dziedzinę $\mathbb{R}$, mając klasę sygnałów $f(t) \in \mathrm{L}_T^2$ transformacja przyjmuje postać (analogicznie do (3.44))

$$F(u) = 1/T \int_{-T/2}^{T/2} f(t) e^{-ju\frac{2\pi}{T}t} dt \tag{3.48}$$

gdzie $F(u) \in \mathrm{l}^2(\mathbb{Z})$, zaś odwrotna

$$f(t) = \sum_{u \in \mathbb{Z}} F(u) e^{ju\frac{2\pi}{T}t} \tag{3.49}$$

W przypadku sygnałów dyskretnych $f(k) \in \mathrm{l}^2(\mathbb{Z})$, rozumianych analogicznie jako okresowe w sensie przejścia granicznego, transformacja jest iloczynem skalarnym sygnału z bazą funkcji sinusoidalnych

$$F(\omega) = \sum_{k \in \mathbb{Z}} f(k) e^{-jk\omega} \tag{3.50}$$

gdzie $F(\omega) \in \mathrm{L}^2([-\pi, \pi])$, zaś odwrotna

$$f(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega) e^{jk\omega} d\omega \tag{3.51}$$

Najbardziej praktyczny, wykorzystywany w realiach systemów cyfrowych, sposób liczenia transformacji Fouriera dla dyskretnych sygnałów opisanych skończonym ciągiem wartości $(f(k))_{k=0}^{K-1} \in \mathbb{R}^K$ (z możliwością reprezentacji sygnałów zespolonych $(f(k))_{k=0}^{K-1} \in \mathbb{C}^K$) określają zależności formy prostej

$$F(u) = \sum_{k=0}^{K-1} f(k) e^{-\frac{j2\pi ku}{K}}, \quad u = 0, \ldots, K-1 \tag{3.52}$$

i odwrotnej

$$f(k) = \frac{1}{K} \sum_{u=0}^{K-1} F(u) e^{\frac{j2\pi ku}{K}}, \quad k = 0, \ldots, K-1 \tag{3.53}$$

Zakładamy przy tym cykliczne rozszerzenie $f(k)$ z okresem K. Wyrażenia (3.52) oraz (3.53) definiują tzw. **dyskretną transformację Fouriera**, w skrócie DFT (ang. *Discrete Fourier Transform*).

Dwuwymiarowa dyskretna transformacja Fouriera – 2D DFT, użyteczna w analizie częstotliwościowej obrazów postaci $f(k, l)$, przyjmuje separowalną postać poprzez sekwencyjne złożenie przekształcenia wzdłuż jednego kierunku (opisanego współrzędną k danych obrazowych), a później drugiego (opisanego współrzędną l). Poprzez analogię do (3.52) przyjmuje ona postać

$$F(u, v) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} f(k, l) e^{-\frac{j2\pi ku}{K}} e^{-\frac{j2\pi lv}{L}} \tag{3.54}$$

gdzie $u = 0, \ldots, K-1,\ v = 0, \ldots, L-1$. Ze względu na brak informacji lokalnych w fourierowskim widmie obrazu, w wielu zastosowaniach przekształcenie 2D DFT (lub zwykle jego szybsza wersja 2D FFT - *Fast Fourier Transform* [173]) realizowane jest przy blokowym podziale dziedziny. Transformata obliczana jest w każdym z bloków niezależnie, co pozwala zachować lokalny (o lokalności decyduje wielkość elementarnego bloku z pokrycia dziedziny) charakter analizy częstotliwościowych właściwości obrazu. Nawiązuje to do koncepcji okienkowej transformacji Fouriera (zobacz (2.4) w punkcie 2.2.2). W zależności od lokalnych cech widma sygnału można stosować dobrane metody filtracji.

Operacja filtracji obrazów w dziedzinie częstotliwości zakłada wykorzystanie filtrów cyfrowych o dobranych, selektywnych właściwościach, np. redukcji składowych wysokoczęstotliwościowych (dolnoprzepustowy filtr wygładzający) czy zachowania jedynie lokalnej informacji o charakterze szybkozmiennym (górnoprzepustowy filtr pasmowy). Stosowane są także filtry pasmowo-przepustowe, zachowujące cechy sygnału reprezentowane przez określony zakres widma (tj. częstotliwości pomiędzy dolną i górną częstotliwością graniczną lub wokół wybranej częstotliwości środkowej, z określoną szerokością pasma przepustowego). Filtry o dokładnie odwrotnych właściwościach, tłumiące sygnał w ustalonym zakresie częstotliwości nazywane są pasmowo-zaporowymi. W przypadku operacji lokalnych są to filtry o skończonej odpowiedzi impulsowej – SOI.

Charakterystyka częstotliwościowa filtru dyskretnego dokonywana jest za pomocą transformacji Fouriera jego skończonej odpowiedzi impulsowej $\{h_n\}_{n=0}^{N-1}$, tak że

$$H(u) = \sum_{n=0}^{N-1} h_n e^{-\frac{j2\pi nu}{N}} \tag{3.55}$$

gdzie $H(u)$ nazywana jest funkcją przenoszenia (inaczej transmitancją) filtru. Analogicznie do (3.54) oraz (3.55) określana jest dwuwymiarowa transmitancja $H(u,v)$ filtrów stosowanych do przetwarzania obrazów.

Liniowej filtracji splotowej $g(k,l) = f * h(k,l)$, realizowanej w przestrzeni obrazu za pomocą filtru $h_{m,n}$ według (3.18), odpowiada (jest równoważna - zgodnie z twierdzeniem o splocie: $\mathcal{F}\{f * h\} = \mathcal{F}\{f\} * \mathcal{F}\{h\}$ [174]) filtracja realizowana w fourierowskiej dziedzinie częstotliwościowej za pomocą iloczynu transformaty Fouriera sygnału i transmitancji filtru

$$G(u,v) = F(u,v) \cdot H(u,v) \tag{3.56}$$

Stąd obraz przetworzony

$$g(k,l) = \mathcal{F}^{-1}\{F(u,v) \cdot H(u,v)\} \tag{3.57}$$

Dyskretna postać funkcji przenoszenia filtru ma duże znaczenie implementacyjne, jednak w wielu przypadkach wygodniej jest analizować właściwości filtrów (o potencjalnie nieograniczonej odpowiedzi impulsowej - NOI) za pomocą ich ciągłej charakterystyki częstotliwościowej (analogicznie do (3.50))

$$H(\omega) = \sum_{n\in\mathbb{Z}} h_n e^{-jn\omega} \tag{3.58}$$

a w przypadku dwuwymiarowym

$$H(\omega_x, \omega_y) = \sum_{m\in\mathbb{Z}} \sum_{n\in\mathbb{Z}} h_{m,n} e^{-jm\omega_x} e^{-jn\omega_y} \tag{3.59}$$

Idealny filtr dolnoprzepustowy o pulsacji granicznej (odcięcia) ω_0 jest opisany przez moduł transmitancji (charakterystyka amplitudowa) postaci

$$|H(\omega)| = \begin{cases} 1 & \text{dla } \omega \leqslant \omega_0 \\ 0 & \text{w p.p.} \end{cases} \tag{3.60}$$

Daje on efekt przepuszczenia bez żądnych zmian wszystkich składowych sygnału o częstotliwościach z użytecznego zakresu widma (pasmo przepustowe), przy całkowitym usunięciu pozostałych (z pasma zaporowego), bez pasma przejściowego (zakres częstotliwości częściowo tłumionych, leżących na granicy zakresów przepustowego i zaporowego).

Transmitancja dolnoprzepustowego filtru idealnego w zakresie pasma przepustowego ma więc postać $H(\omega) = e^{-j2\Pi\frac{\omega}{\omega_0}}$ (w paśmie zaporowym jest zerowa), przy liniowej fazie $\angle H(\omega) = -\frac{2\Pi}{\omega_0}\omega$.

Niestety, takich filtrów nie sposób wykorzystać w praktyce (jest niereализowalny fizycznie) – jego skończone widmo (tj. przyjmujące niezerowe wartości jedynie w domkniętym obszarze dziedziny) implikuje NOI o postaci funkcji $\mathrm{sinc}x = \sin x/x$ (nie jest zachowany warunek przyczynowości, bo odpowiedź impulsowa filtru nie ma końca). Obcinanie odpowiedzi impulsowej do postaci SOI

powoduje odkształcenie zmierzonych według (3.61) i (3.62) postaci $H(d)$, pojawią się niekorzystne efekty zniekształceń fazowych.

Większość uniwersalnych filtrów obrazowych ma symetryczną charakterystykę częstotliwościową, co upraszcza definiowanie ich transmitancji jako $H(d)$, gdzie odległość od początku układu częstotliwościowych współrzędnych $d = \sqrt{\omega_x^2 + \omega_y^2}$. Nie wyklucza to oczywiście możliwości dobierania filtrów o anizotropowej charakterystyce przepustowej $H(\omega_x, \omega_y)$, dostosowanej do specyfiki przetwarzanych obrazów.

W przypadku idealnego filtru obrazowego, dolnoprzepustowe o pulsacji granicznej d_0 mamy więc analogiczną charakterystyka amplitudową postaci

$$|H_i(d)| = \begin{cases} 1 & \text{dla } d \leqslant d_0 \\ 0 & \text{w p.p.} \end{cases} \tag{3.61}$$

Odwrotne zadanie przepuszczania wysokich częstotliwości przy eliminacji niskich można uzyskać za pomocą analogicznego, idealnego filtru górnoprzepustowego:

$$|H_i(d)| = \begin{cases} 1 & \text{dla } d \geqslant d_0 \\ 0 & \text{w p.p.} \end{cases} \tag{3.62}$$

Stosowanie filtrów obrazowych o charakterystyce zbliżonej do gwałtownego skoku filtru idealnego jest źródłem – np. w filtracji dolnoprzepustowej – tzw. efektów pierścieniowych (*ringing effect*) powodowanych zjawiskiem Gibbsa (brakuje wyciętych z widma wysokich częstotliwości do reprezentacji krawędziowych skoków funkcji jasności), z powielaniem krawędzi i słabą redukcją szumu impulsowego (rys. 3.19). Takie właściwości próbuje się aproksymować za pomocą realnych filtrów o łagodniejszych charakterystykach w zakresie pasma przejściowego, najlepiej o znikomych oscylacjach w pasmach przepustowym i zaporowym.

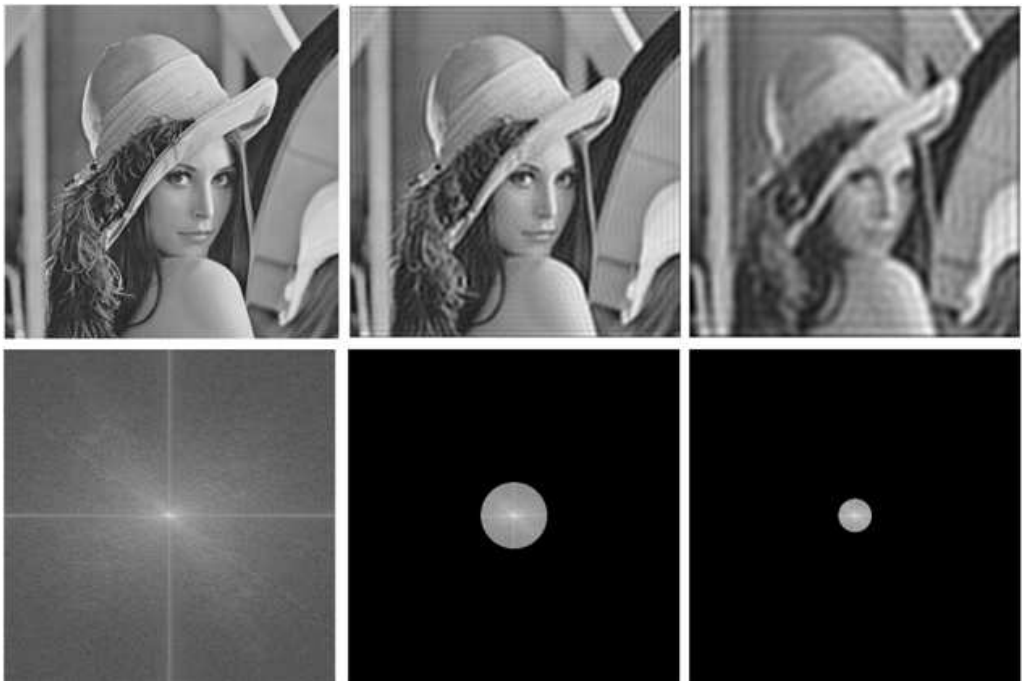
Przykładowe postacie charakterystyk częstotliwościowych filtrów stosowanych w przetwarzaniu obrazów są następujące:

a) filtry Butterwortha, o maksymalnie płaskiej charakterystyce amplitudowej w paśmie przepustowym, bez oscylacji (zafalowań) w pasmach przepustowym i zaporowym, o charakterystyce fazowej najbardziej zbliżonej do liniowej, jednak przy stosunkowo wolno opadającym widmie pasma przejściowego;

   - uśredniający (dolnoprzepustowy)

$$|H_B|(d) = \frac{1}{1 + (d/d_0)^{2n}} \tag{3.63}$$

     gdzie $n \in \mathbb{Z}$ jest rzędem filtru – wraz ze wzrostem n rośnie nachylenie charakterystyki w paśmie przejściowym, kosztem jednak coraz bardziej nieliniowej fazy; przy odpowiednio dużym n pojawiają się efekty Gibbsa;



Rysunek 3.19: Przykład filtracji w dziedzinie częstotliwościowej za pomocą dolnoprzepustowego filtru "idealnego" (aproksymowanego), z widocznym efektem pierścieniowym; kolejno w szeregu u góry obraz źródłowy, obraz po filtracji z wykorzystaniem transmitancji (3.61) przy rozmiarze d_0 równym 20% dziedziny obrazu oraz obraz po analogicznej filtracji ze zmniejszeniem względnego rozmiaru d_0 do 10%; w dolnym rzędzie zamieszczono częstotliwościowe widma amplitudowe odpowiednich obrazów, z jednoznacznie selektywną rolą filtru "idealnego".

- górnoprzepustowy

$$|G_B(d)| = \frac{1}{1 + (d_0/d)^{2n}} \tag{3.64}$$

c) filtry gaussowskie

- uśredniający

$$|H_g(d)| = e^{-(-1/2(d/d_0)^2)} \tag{3.65}$$

- górnoprzepustowy

$$|G_g(d)| = 1 - e^{-1/2(d/d_0)^2} \tag{3.66}$$

b) filtry wykładnicze rzędu n (uogólnienie filtrów gaussowskich)

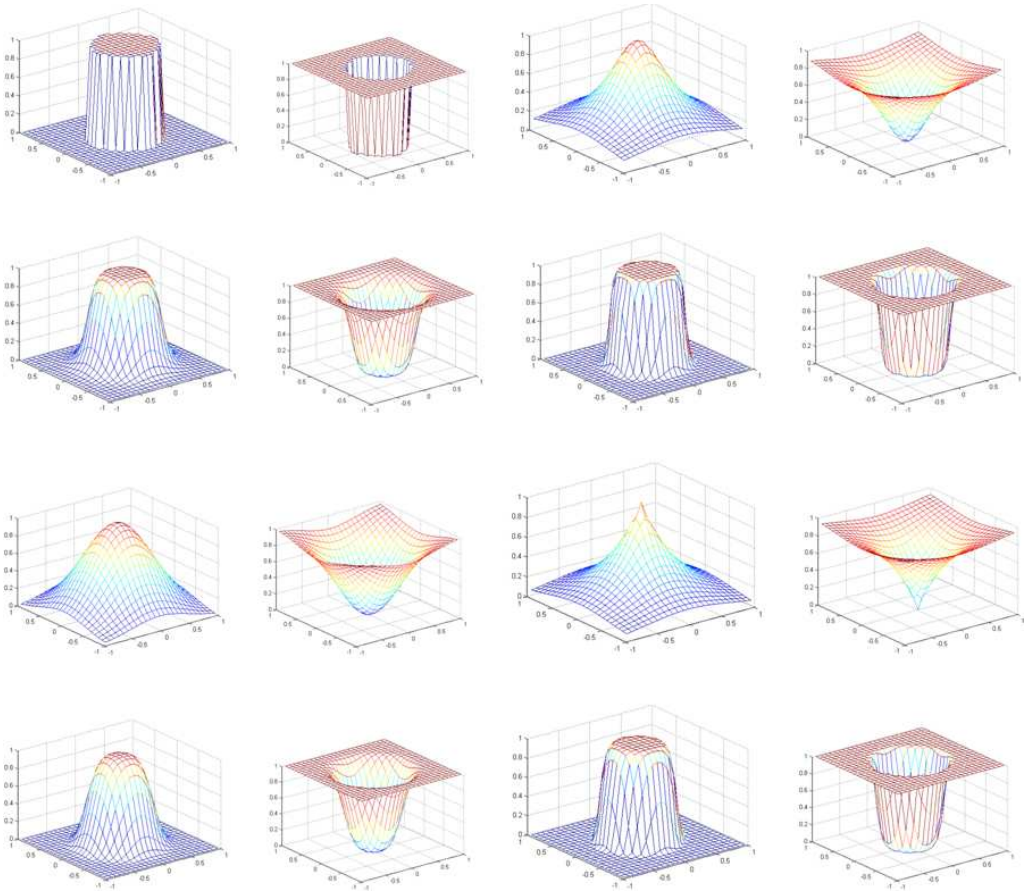
- uśredniający

$$|H_e(d)| = e^{-(d/d_0)^n} \tag{3.67}$$

- górnoprzepustowy

$$|G_e(d)| = 1 - e^{-(d/d_0)^n} \tag{3.68}$$

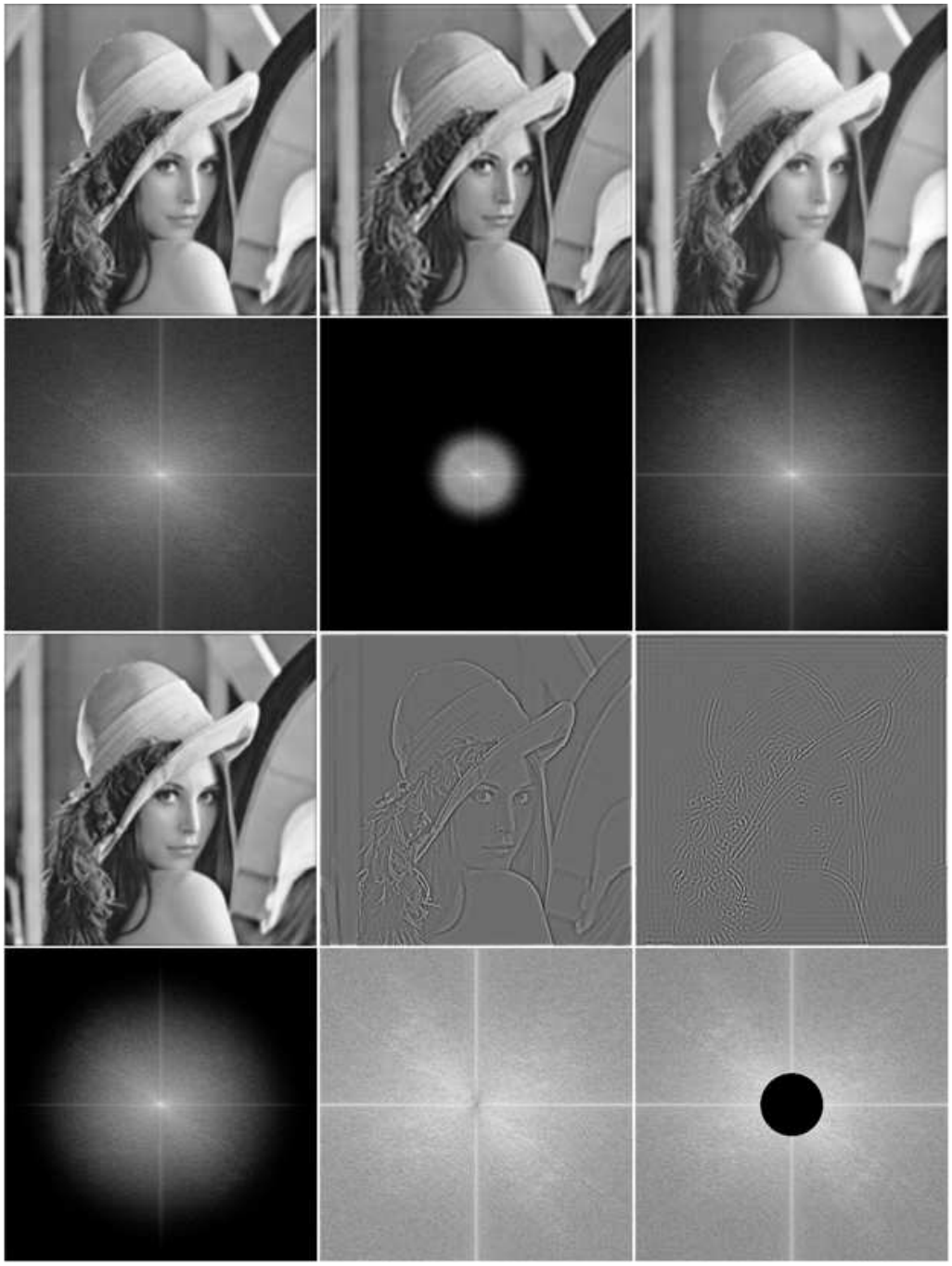
W praktyce przy przetwarzaniu obrazów stosowane są zwykle rzeczywitoliczbowe, przybliżające postacie transmitancji opisanych powyżej filtrów częstotliwościowych, tak że $H(d) \approx |H_i(d)|$. Zróżnicowanie charakterystyk amplitudowych tych filtrów ukazano na rys. 3.20, zaś efekty filtracji na rys. 3.21.



Rysunek 3.20: Charakterystyki amplitudowe wybranych filtrów – kolejno od lewej do prawej, z góry na dół pary filtrów dolno- i górnoprzepustowych: idealny, Butterwortha rzędu $n = 1$, $n = 4$, $n = 10$, gaussowski oraz wykładniczy rzędu $n = 1$, $n = 4$, $n = 10$.

Filtracja morfologii matematycznej

Podstawowa koncepcja morfologii matematycznej zakłada, że istnienie struktury geometrycznej (obiektów) zainteresowania nie jest zjawiskiem całkowicie obiektywnym. Struktura ta ujawnia się (jest odkrywana) dopiero poprzez relację pomiędzy nią a narzędziem badawczym, nazywanym elementem strukturującym



Rysunek 3.21: Efekty filtracji w dziedzinie częstotliwościowej; w dwóch rzędach zamieszczono przekształcone obrazy wraz z ich widmami amplitudowymi (pod spodem), odpowiednio dla filtrów dolnoprzepustowych Butterwortha z $n = 1$ i $n = 10$, wykładniczego z $n = 1$, Gaussa, a także górnoprzepustowych – Butterwortha z $n = 1$ oraz "idealnego" przy względnym rozmiarze d_0 równym 20%.

[169, 175]. Elementy te modyfikują kształt eksponując strukturę o określonych właściwościach, przy czym zwykle ich rozmiar jest znacząco mniejszy od rozmiaru obiektu badanego. Badanie obrazów za pomocą operacji morfologicznych może łączyć w sobie zarówno funkcje ulepszania obrazów, np. powiększanie mało widocznych szczegółów, jak też analizy – określenie szkieletu figury, segmentację obiektu itp.

Podstawy morfologii sięgają do teorii zbiorów, przy czym działania na zbiorach dokonywane są zwykle w przestrzeni dyskretnej $\mathbb{R}^N$ lub $\mathbb{Z}^N$, jako podzbiory przestrzeni euklidesowej[2] $\mathcal{E}$. Nieco inaczej definiowane operacje podstawowe odnoszą się także do funkcji, w tym dyskretnej funkcji jasności obrazu.

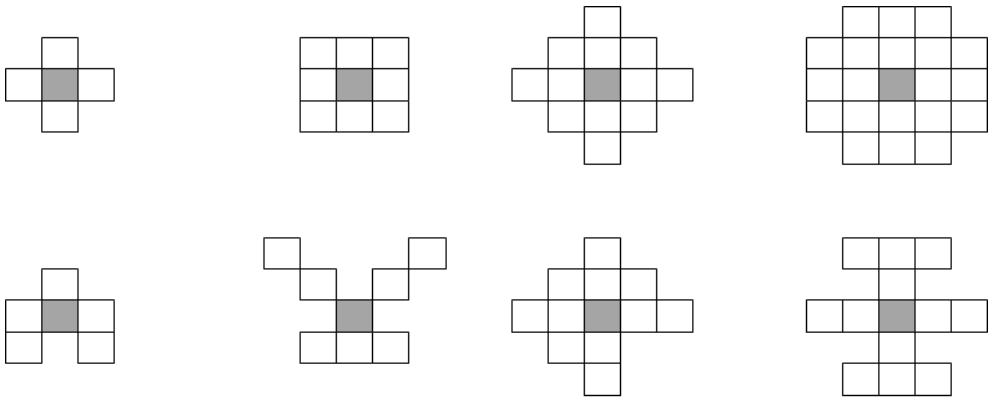
Filtracja morfologiczna jest operacją nieliniową, z binarną maską elementu strukturalnego, której kształt i wymiary określają sąsiedztwo punktu obrazu uwzględniane w procesie przetwarzania - zobacz rys. 3.22. Centralny punkt przemieszczanej maski identyfikuje przetwarzany punkt przestrzeni obrazu (tj. pragmatycznego zawężenia przestrzeni euklidesowej). W zależności od relacji punktu wraz z otoczeniem względem obiektu i jego tła, według przyjętej funkcji przetwarzania ustalana jest nowa wartość (albo przynależność) punktu. Przyjmuje się też w wersji podstawowej binarny opis obrazu, gdzie "1" oznacza piksele obiektu, a "0" etykietuje jego tło, czyli zbiór pikseli nie należących do obiektu. w takim przypadku punkty obrazu opisane są przynależnością do obiektu bądź tła oraz współrzędnymi. Obiekt zaś to zbiór punktów o określonych współrzędnych, czyli *de facto* zbiór współrzędnych wskazujących kolejne punktu obiektu. Operacja na punktach obiektu realizowane są więc jako działania na współrzędnych (punkty rozumiane są więc jako wektory zorientowane względem początku układu współrzędnych).

Wykorzystywane są dwa podstawowe rodzaje operacji: dylacja i erozja, przy czym dylacja w podstawowym rozumieniu jest izotropowym rozszerzeniem obiektu w stopniu określonym przez kształt i rozmiar elementu strukturującego, zaś erozja jest uszczupleniem zarysu obiektu przez narzędzie badawcze. Bardziej złożone operacje tworzone są na bazie tych dwóch przekształceń.

Wykorzystajmy operację sumy Minkowskiego (teoriomnogościowa) dwóch zbiorów:
$A \oplus B = \{a + b : \ a \in A, b \in B\}$ dla $A, B \in \mathcal{E}$, gdzie $\mathcal{E}$ to przestrzeń euklidesowa. Istotna jest też translacja (przesunięcie) zbioru A o wektor b: $A_b = A + b = \{a + b : \ a \in A\}$.

Dylacja obiektów obrazu jest operacją sumowania zbioru wszystkich (logicznych) punktów a obiektu A oraz punktów c maski elementu strukturującego C,

[2] przestrzeń euklidesowa służy tworzeniu naturalnych modeli świata rzeczywistego, jest "płaska" w odróżnieniu od przestrzeni budowanych np. na sferze; $n \geqslant 1$ wymiarowa przestrzeń euklidesowa jest uogólnieniem płaszczyzny i przestrzeni trójwymiarowej jako zbiór wszystkich uporządkowanych punktów $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$ wraz z odległością pomiędzy punktami czy wektorami określaną za pomocą iloczynu skalarnego.



0	1	1	1	0
0	0	1	0	0
1	1	1	1	1
0	0	1	0	0
0	1	1	0	0

Rysunek 3.22: Przykładowe elementy strukturujące, stosowane przy morfologicznej filtracji obrazów – w środku każdego z nich znajduje się szary punkt centralny, wskazujący piksel przetwarzany po nałożeniu maski na macierz obrazu; na dole logiczny opis nieregularnego kształtu maski na regularnym nośniku.

przy czym są to punkty z logiczną **1** opisane za pomocą swoich współrzędnych względem punktu centralnego. W efekcie otrzymujemy rozszerzony równomiernie obiekt. Możemy też rozważać dylację jako sumę otoczeń wszystkich punktów obiektu definiując otoczenie (kontekst) punktu a jako $C_a = \{a + c : c \in C\}$ (lub też sumę wszystkich translacji obiektu o punkty maski $A_c = \{a + c : a \in A\}$), wychodzącą poza obrys A na długość promienia elementu strukturującego. Jeszcze inny sposób to uzupełnienie obiektu A o te dodatkowe punkty przestrzeni, których otoczenie określone symetrycznym $C^s = \{t \in \mathcal{E} : -t \in C\}$ ma punkty wspólne z A.

Możemy więc określić dylację obiektu A za pomocą elementu strukturującego C jako

$$D_C(A) = A \oplus C = \bigcup_{a \in A} C_a = \{t \in \mathcal{E} | A \cap (C_t^s) \neq \emptyset\} \tag{3.69}$$

Erozja ma działanie przeciwne, odwołując się do operacji różnicy Minkowskiego zbiorów obiektu i kontekstu, czyli uszczuplenia obiektu o otoczenie punktów przestrzeni bezpośrednio przylegających do obiektu. Inaczej, operacja erozji wyznacza część wspólną zbiorów A_c, jest sumą punktów obiektu, których otoczenie zawiera się w A.

Erozja obiektu A za pomocą elementu strukturującego C definiowana jest

więc jako

$$E_C(A) = A \ominus C = \bigcap_{c \in C} A_c = \{a \in A | C_a \subseteq \ A\} = \{t \in E | C_t \subseteq \ A\} \tag{3.70}$$

Liczenie tych operacji morfologicznych na obrazie zawierającym potencjalnie większą liczbę rozdzielnych obiektów odbywa się analogicznie. Wszystkie etykietowane jedynką piksele traktowane mogą być jako punkty umownego, niekiedy niespójnego czy wielospójnego (składającego się z kilku rozdzielnych części) "obiektu".

W bardziej praktycznych rozważaniach, algorytm procedury dylacji może przebiegać następująco:

Algorytm 3.2 *Dylacja obrazu*

1. Zakładamy binarny opis obrazu źródłowego z etykietą $e(k,l) = \mathbf{1}$ przypisaną pikselom obiektu, przy wyzerowanych pikselach tła; ustalamy określoną maskę elementu strukturującego $c_{i,j}$, $i,j \in \{-I, -I+1, \ldots, I\}$, opisaną rozkładem logicznych jedynek na regularnym nośniku kwadratowym o boku $2I+1$ (wszystkie przykładowe maski z rys. 3.22 można zdefiniować na kwadracie 5×5); ustalamy też obraz wynikowy jako $\tilde{e}(k,l) = \mathbf{0}$ dla każdego piksela;

2. Do każdego piksela obrazu przykładamy punkt centralny elementu strukturującego $c_{0,0}$ definiując jego kontekst i wyznaczając logiczne pole kontekstu (włącznie z pikselem, zaś dla fragmentów kontekstów "wystających" poza obraz ustalamy etykietę zerową)

$$\mathcal{C}_e(k,l) = \Big\{ e(k+i, l+j) \wedge c_{i,j} : \ i,j \in \{-I, \ldots, I\} \Big\} \tag{3.71}$$

rozmiaru $2I+1 \times 2I+1$

3. Jeśli choć jeden z elementów $\mathcal{C}_e(k,l) = \mathbf{1}$, to przetwarzany piksel odpowiadający punktowi centralnemu przyjmuje wartość $\mathbf{1}$, czyli

$$\tilde{e}(k,l) = \bigvee \mathcal{C}_e(k,l) \tag{3.72}$$

4. Uzyskany binarny, poszerzony obraz obiektu jest wynikiem końcowym.

□

Algorytm erozji obrazu przebiega analogicznie, przy korekcie sposobu przetwarzania punktów obrazu.

Algorytm 3.3 *Erozja obrazu*

1. Założenia jak w algorytmie 3.2;

2. Do każdego piksela obrazu należącego do obiektu przykładamy punkt centralny elementu strukturującego $c_{0,0}$ definiując jego kontekst i wyznaczając logiczne pole kontekstu według (3.71);

3. Jeśli choć jeden z elementów $\mathcal{C}_e(k,l) = \mathbf{0}$, to przetwarzany piksel odpowiadający punktowi centralnemu przyjmuje wartość $\mathbf{0}$, czyli

$$\tilde{e}(k,l) = \bigwedge \mathcal{C}_e(k,l) \tag{3.73}$$

4. Uzyskany binarny, poszerzony obraz obiektu jest wynikiem końcowym.

□

Przyglądając się powyższym algorytmom i wynikającym z nich prostym implementacjom podstawowych operacji morfologicznych, nasuwa się nieco prostsza interpretacja erozji i dylacji – jako transformacji typu trafi–nie trafi (emphhit or miss). Według przyjętej konwencji logicznej, piksele oznaczone $\mathbf{1}$ definiują aktywne sąsiedztwo piksela centralnego (o szarym kolorze na rys. 3.22), jakby wzorzec przesuwający się po obrazie. Postać aktywnych pikseli sąsiedztwa decyduje o efekcie filtracji, czyli wartości przypisywanej pikselowi centralnemu.

W przypadku dylacji, jeśli choć jeden z pikseli sąsiedztwa ma etykietę $\mathbf{1}$, wtedy wartość piksela centralnego ustawiana jest również na $\mathbf{1}$. Przekształcenie to należy do grupy operatorów addytywnych, powodujących rozszerzenie obiektów (na zewnątrz lub do wewnątrz) zależnie od postaci maski sąsiedztwa. Bardziej subtelnym operatorem tej grupy jest wypełnienie wewnętrzne (*interior fill*), kiedy to dopiero zapalone (z etykietą $\mathbf{1}$) wszystkie piksele sąsiedztwa ustawiają $\tilde{e}(k,l) = \mathbf{1}$, czyli $\tilde{e}(k,l) = e(k,l) \vee \bigwedge \Big(\mathcal{C}_e(k,l) - e(k,l)\Big)$. Można w ten sposób bardziej precyzyjnie kształtować proces uzupełniania brakujących fragmentów obiektów.

Erozja zaś należy do grupy operatorów subtraktywnych, uszczuplających obiekty w określonych lokalnie okolicznościach. Logiczny iloczyn etykiet pikseli sąsiedztwa oraz piksela centralnego pozwala restrykcyjnie usunąć wszystkie piksele z sąsiedztwem nie mieszczącym się w obiekcie – (3.73). Można tę operację ograniczyć np. jedynie do usuwania pikseli izolowanych (np. stanowiących szum), wtedy $\tilde{e}(k,l) = e(k,l) \wedge \bigvee \Big(\mathcal{C}_e(k,l) - e(k,l)\Big)$.

Wspomniane operacje podstawowe mają swoje odpowiedniki dla obrazów wielopozomowych, z wielowartościową skalą danego komponentu funkcji jasności jako operatory maksimum i minimum liczone w kontekstach określonych przez element strukturujący. Dylacja realizowana jest za pomocą filtru maksymalnego, erozja zaś - minimalnego, przy założeniu konwencji jaśniejszego obiektu na ciemniejszym tle (w konwencji odwrotnej należy zamienić to przyporządkowanie – erozja z filtrem maksymalnym, a dylacja - minimalnym).

Zwiększając efekt filtracji można dobierać elementy strukturujące o większym nośniku, można także sekwencyjnie powtarzać operacje erozji czy dylacji, zmieniając przy tym maskę kontekstu. Przykładowo, iteracyjne stosowanie erozji przy

odpowiednich maskach dobieranych warunkowo i bezwarunkowo aż do uzyskania linii o jednopikselowej szerokości pozwala uzyskać przybliżenie szkieletu figury, czyli służy szkieletyzacji. Szkieletem figury nazywamy jej zbiór wszystkich punktów osiowych, tj. równoodległych od co najmniej dwóch punktów brzegowych figury.

Różnice pomiędzy efektem filtracji operatorem dylacji oraz erozji, uwydatniające zmiany rozszerzania i przycinania obiektów, a także odniesienie efektów dylacji i erozji do obrazu oryginalnego $\mathbf{e}$ definiuje klasę morfologicznych operatorów gradientowych:

- gradient morfologiczny: $GM_C(\mathbf{e}) = D_C(\mathbf{e}) - E_C(\mathbf{e})$,
- gradient wewnętrzny: $GW_C(\mathbf{e}) = \mathbf{e} - E_C(\mathbf{e})$,
- gradient zewnętrzny: $GZ_C(\mathbf{e}) = D_C(\mathbf{e}) - \mathbf{e}$.

Ponadto stosowanych jest szereg operatorów będących złożeniem, niekiedy kilkukrotnym tych dwóch podstawowych operacji. Wymienić tutaj należy przede wszystkim otwarcie do domknięcie, przekształcenia które stanowią łagodniejszą wersję erozji i dylacji zachowując generalnie kształt i wielkość obiektów o znaczących kształtach (tj. wyraźnie większych od elementu strukturującego), wygładzając kontury, zapewniając bardziej naturalny efekt przetwarzania obrazów. Mamy więc definicje

- otwarcie: $O_C(\mathbf{e}) = D_C(E_C(\mathbf{e}))$,
- zamknięcie (domknięcie): $Z_C(\mathbf{e}) = E(D(S), S)$,

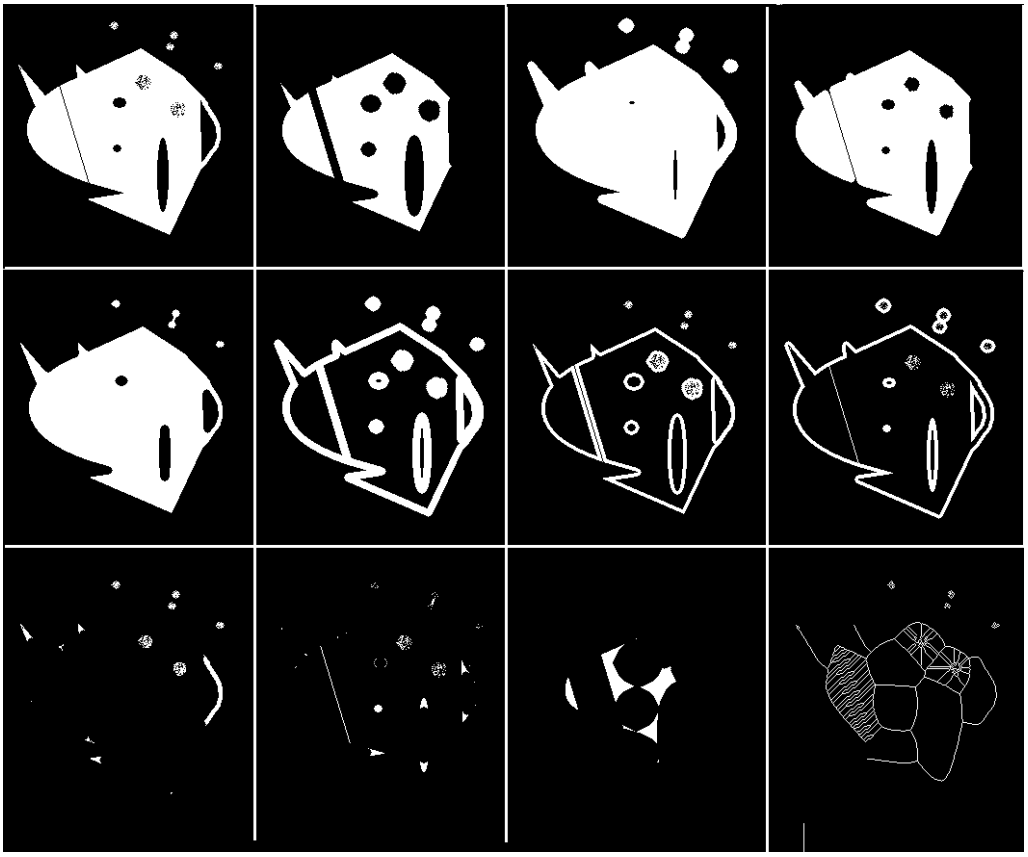
przy czym otwarcie usuwa drobne fragmenty figury, w których nie mieści się element strukturujący, drobne połączenia czy wypukłości. Domknięcie zaś wypełnia mniejsze od elementu strukturującego wklęsłości, wcięcia czy dziury. Uwypuklenie zmian dokonywanych przez otwarcie lub domknięcie, dające również bardziej naturalny efekt od operacji gradientowych uzyskuje się za pomocą dwóch operatorów

- *white top-hat*: $WTH = \mathbf{e} - O_C(\mathbf{e})$,
- *black top-hat*: $BTH = Z_C(\mathbf{e}) - \mathbf{e}$,

Przykładowe efekty filtracji morfologicznej na obrazach binarnych oraz wielopoziomowych pokazano odpowiednio na rys. 3.23 oraz rys. 3.24.

3.1.4 Modele obrazów

Ogólnie modelem nazywany jest opis (w formie schematu, zależności matematycznych, konstrukcji, rysunku technicznego itp.) ukazujący istotę (proces, działanie,



Rysunek 3.23: Przykładowe efekty filtracji morfologicznej obrazu binarnego, kolejno (od lewej do prawej, góra–dół): obraz źródłowy, po erozji, dylacji, otwarciu, domknięciu, gradiencie morfologicznym, gradiencie wewnętrznym, gradiencie zewnętrznym, operatorze WTH, BTH, pięciokrotnej erozji oraz szkieletyzacji; przyjęto stały element strukturujący – eliptyczny o promieniu 4 pikseli.

budowę, cechy, zależności) jakiegoś aspektu rzeczywistości (zjawiska, sytuacji, przedmiotu zainteresowania, itp.). Opis ten jest zwykle sparametryzowany (sformalizowany, ustrukturyzowany), przybliżony (tj. dopuszczający istnienie treści nieistotnych), najlepiej kompletny (tj. uwzględniający wszystkie istotne dla użytkownika elementy, obiekty, właściwości), a przy tym upakowany (inaczej zwarty, oszczędny w liczbie parametrów, złożoności obliczeniowej). Uwzględnia treść – zestaw pojęć, obiektów, reguł zależności, jej znaczenie użytkowe (przydatność) oraz formę (cechy wyrazu treści).

Modeli winien w pierwszej kolejności obejmować istotę (rdzeń) zjawiska, uwzględniając wszystkie konieczne właściwości w danym zastosowaniu. Rozszerzenia modelu dotyczą dołączania kolejnych cech mniej istotnych, najlepiej w porządku zgodnym z ustaloną hierarchią ich użyteczności.

Konstrukcja modelu jest zwykle kompromisem pomiędzy wiernością opisu, a

Rysunek 3.24: Przykładowe efekty filtracji morfologicznej obrazu testowego Lena, kolejno (od lewej do prawej, góra–dół): obraz źródłowy, po erozji za pomocą trzech masek (blokowej z promieniem 1 piksela, eliptycznej 3 pikseli, blokowej 4 pikseli, ukazanych w powiększeniu w górnym roku obrazków), dylacji, otwarciu, domknięciu, gradiencie morfologicznym, gradiencie wewnętrznym, gradiencie zewnętrznym, operatorze WTH, BTH, pięciokrotnej erozji oraz szkieletyzacji; dla pozostałych operacji zastosowano stały element strukturujący – eliptyczny o promieniu 3 pikseli.

jego złożonością.

Wśród często poszukiwanych, dodatkowych zalet modelu wymienić należy przede wszystkim:

- wysoka specyficzność (łatwe dostosowanie do uwarunkowań, wykorzystanie dostępnej, aktualnej wiedzy, elastyczność),
- hierarchiczność (uwzględnia różnego typu relacje pomiędzy pojęciami i obiektami, wprowadza uporządkowanie, progresję itp.),
- skalowalność (możliwość dostosowania do funkcjonalnych i sprzętowych wymagań odbiorcy),
- wysoka separowalność opisu treści użytecznej od szumów, artefaktów, itp. (ogólnie składników nieinformatywnych),
- adaptacyjność (w sensie lokalnym i globalnym).

Modelowanie obrazu jest procesem mającym na celu stworzenie reprezentacji danych obrazowych, która opisuje przede wszystkim zawartą w nim treść. W razie potrzeby model może także obejmować szumy i artefakty.

Rzetelny model służy analizie lub syntezie treści. Odkryta analitycznie treść pozwala zrozumieć przekaz informacji i odpowiednio go zinterpretować. Analogicznie, przy syntezie sceny realistycznej, np. na podstawie modeli "uczonych" z serii pomiarów, wizualizowane obiekty służą przekazowi treści w sposób możliwie wiarygodny i jednoznaczny. Ponieważ treść przekazu obrazowego to przede wszystkim obiekty, sposób opisu obiektów, ich właściwości (tekstur, krawędzi) oraz wzajemnych zależności (położenie, nakładanie, różnicowanie właściwości) jest elementem decydującym o ich skuteczności, różnicującym różne metody modelowania.

Model obrazu jest użyteczny w algorytmach rozpoznawania wzorców, graficznej konstrukcji obrazów lub ich fragmentów, ustalania zwartej reprezentacji w algorytmach kompresji, indeksowania lub innych. Reprezentacja informacji zastosowana w modelu stanowi w pierwszej kolejności o jego użyteczności, decyduje o jego walorach aplikacyjnych.

Metody modelowania obrazów można podzielić na kilka umownych kategorii:

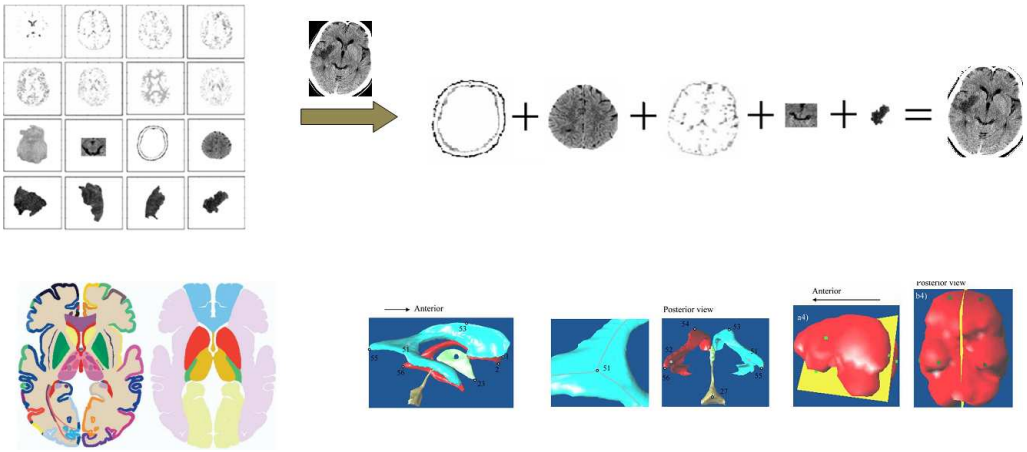
- semantyczne, czy bezpośrednio odnoszące się do wiedzy dziedzinowej, przewidywanych form manifestacji treści, informacji odczytywanej na poziomie abstrakcji użytkownika;
- probabilistyczne (stochastyczne), przybliżające obraz polem losowym, łańcuchami Markowa, rozkładami pozwalającymi generować losowe realizacje o zbliżonych statystycznie właściwościach;

- geometryczne, jako pasowane siatki prymitywów geometrycznych o ustalanym *a priori* formie (np. prostokątów);
- obiektowe, odnoszące się do bardziej złożonych kształtów elementów wzorcowych, ustalanych *posteriori* podczas procesu modelowania;
- funkcjonalne, wykorzystujące do opisu obrazów właściwej klasy funkcje.

Semantyczne modele obrazów

Wyobraźmy sobie zasadniczy schemat modelowania treści obrazowej, specyficzny, dostosowany do uwarunkowań metody obrazowania oraz morfologii badanych obiektów (struktur). Model taki nazwiemy semantycznym, bo najbardziej istotne w nim jest ustalenie znaczenia tworzących go obiektów, których zbiór nazywany słownikiem wyczerpuje możliwą treść. Złożenie z obiektów słownika, według koncepcji rozwinięcia, całej treści zawartej w obrazie następuje poprzez przypisanie im odpowiednich wag. Rozkład wag przypisanych poszczególnym obiektom stanowi informację obrazową, której analiza pozwala na właściwą interpretację obrazów i podjęcie właściwych decyzji warunkowanych odebraną informacją.

Przykład modelu semantycznego służącego analizie informacji zawartej w obrazach CT głowy pokazano na rys. 3.25.



Rysunek 3.25: Semantyczne modelowanie obrazów na przykładzie opisu treści obrazowych badań głowy za pomocą tomografii komputerowej; u góry zamieszczono koncepcję słownika struktur anatomicznych, które mogą posłużyć jako elementy składowe (komponenty) modelu mózgu wpasowywanego w realny wynik badania obrazowego; w dolnym rzędzie – dwa przykładowe fragmenty komputerowego atlasu mózgu, wykorzystywane w analizie semantycznej obrazów w Biomedical Imaging Laboratory, Singapur, a dalej anatomiczne punkty orientacyjne ludzkiego mózgu wykorzystywane w segmentacji wybranych struktur anatomicznych [176].

Geometryczne modele obrazów

Znane z zastosowań grafiki komputerowej siatki prymitywów rozpinają powierzchnie modeli obiektów przestrzennych. Ustalany aspekt geometryczny stosowanych, najprostszych form opisu treści jest wynikiem kompromisu pomiędzy prostotą obliczeń, elastycznością modelu, a jego skutecznością, tj. stosunkiem wierności przybliżenia do złożoności (liczby parametrów) modelu.

Elastyczność modelu oznacza łatwą jego modyfikację według szybkich algorytmów działających lokalnie, tj. w wąskim obszarze dziedziny. Liczba sąsiadujących oczek siatki (*de facto* parametrów zapewniających zachowanie cech modelu po modyfikacji), które musimy zmodyfikować doprecyzowując fragment pokrycia modelowanej sceny powinna być minimalna. Z drugiej strony taka właściwość modelu nie powinna być okupiona jego nadmierną złożonością.

Skuteczność modelu zapewnia możliwie silne podobieństwo prymitywów do kształtów dominujących w opisie treści obrazów, np. w zakresie wielokątowej złożoności (trójkąty, czworokąty, sześciokąty foremne itp.), gładkości konturów obiektów (sygnatury okręgów czy elips), kierunkowości (romby, równoległoboki, trapezy itp.). Możliwa adaptacja formy podstawowej do charakteru modelowanej treści wymaga jednak trafnego kryterium, szybkiego algorytmu dopasowania oraz zadawalającej stacjonarności form treściowych.

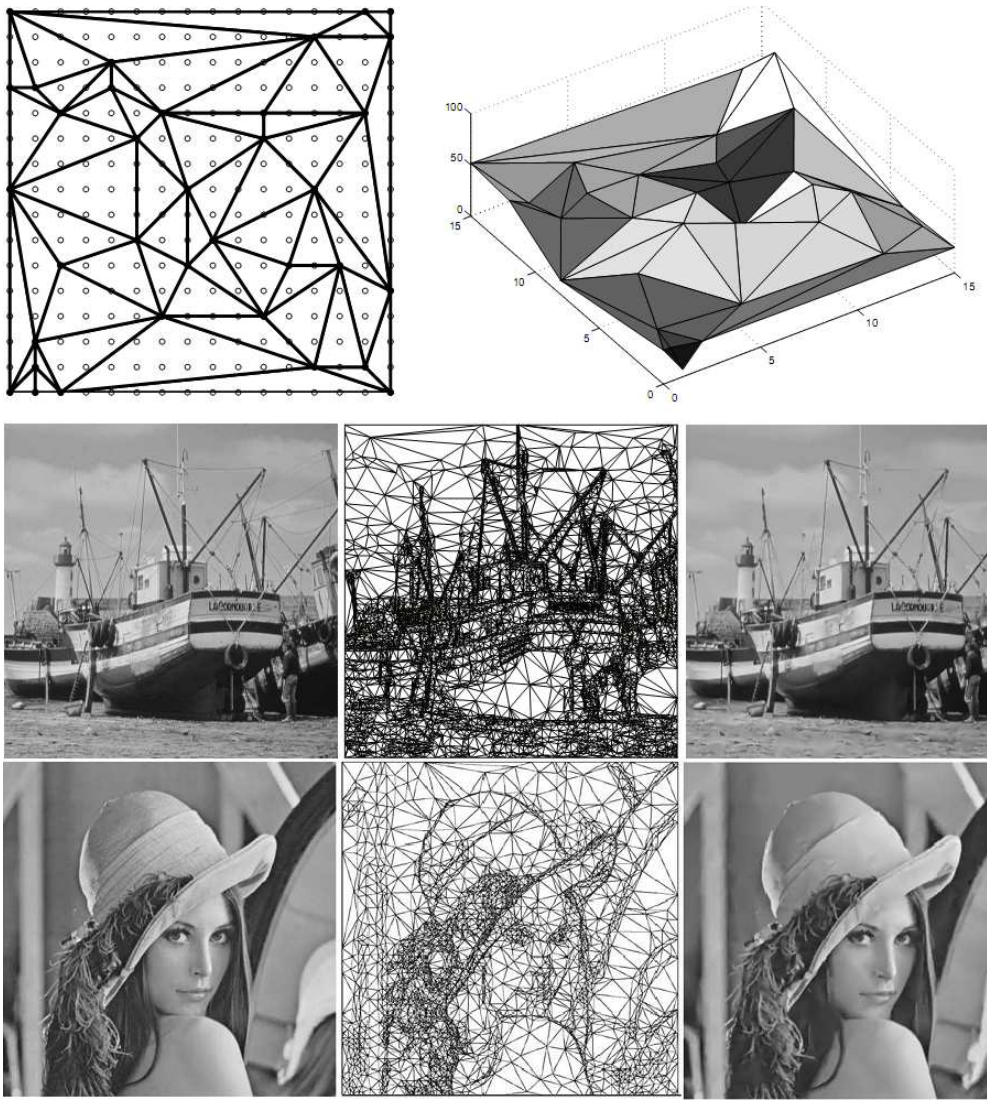
Przykładowe modelowanie obrazów za pomocą adaptacyjnej triangulacji przedstawiono na rys. 3.26. Mają one zastosowanie w kompresji obrazów [177], ale przede wszystkim w syntezie obrazów metodami grafiki komputerowej.

Obiektowe modele obrazów

Są naturalnym sposobem opisywania treści obrazów zawierających obiekty charakteryzowane teksturą, będące względem siebie w określonych relacjach, uzupełnione tłem. Obiekty są wyznaczane za pomocą metod segmentacji, rozpoznawania wzorców, wpasowywania znanych *a priori* modeli parametrycznych w lokalne struktury obrazu itp. Ustalane są także wzajemne relacje pomiędzy obiektami, częściami obiektów czy tłem.

Modele obiektów to często kształt opisany gładkim konturem, tekstura czy kolor regulowane zestawem parametrów z dopuszczalnym zakresem zmian, zwykle ograniczenia rozmiaru, niekiedy położenia, czasami odniesienia do cechy chropowatości, przezroczystości czy trzeciego wymiaru zobrazowanych obiektów.

Dobrym przykładem obiektowego myślenia o obrazach, w pewnym sensie koncepcją reprezentatywną są modele aktywnych konturów – *Active Contour Models* (ACM), inaczej *snakes* [178], kiedy to skalowalny, elastyczny wzorzec, wyznaczony na podstawie danych treningowych oraz innej wiedzy dostępnej *a priori* służy ustaleniu realnego obrysu obiektu o właściwościach (przede wszystkim kształcie) danej klasy. Wzorcowy kontur "porusza się" po obrazie wpasowując się w miejsce najbardziej dogodne o pożądanym profilu krawędzi opisywanego obiektu.



Rysunek 3.26: Geometryczne modele obrazów – przykłady adaptacyjnej triangulacji obszaru obrazu, wyznaczającej siatkę prymitywów, łączących tzw. punkty "znaczące" (narożniki, fragmenty wyraźnych krawędzi itp.); u góry po lewej – siatka trójkątów wyznaczających drobne regiony, które opisano z wykorzystaniem liniowej funkcji sklejanej (góra - po prawej). W kolejnych rzędach przykłady modelowania obrazów naturalnych: z lewej strony oryginały, potem siatki prymitywów i aproksymacje obrazów źródłowych na podstawie modeli (rysunki zaczerpnięto z http://drna.di.univr.it/papers/2010/Iskeatal/t1.pdf oraz [177]).

Parametry modelu określają w pierwszej kolejności jego elastyczność (bardziej sztywny – mniej podatny na lokalne zaburzenia kształtu, bardziej elastyczny – lepiej i szybciej wpasowuje się w lokalne zawirowania struktury) oraz podatność na zmiany lokalne, czyli z jaką dokładnością można dopasować kształt obiektu do

realiów zobrazowania (zwykle jest to ustalane drogą kompromisu ze złożonością obliczeniową metody).

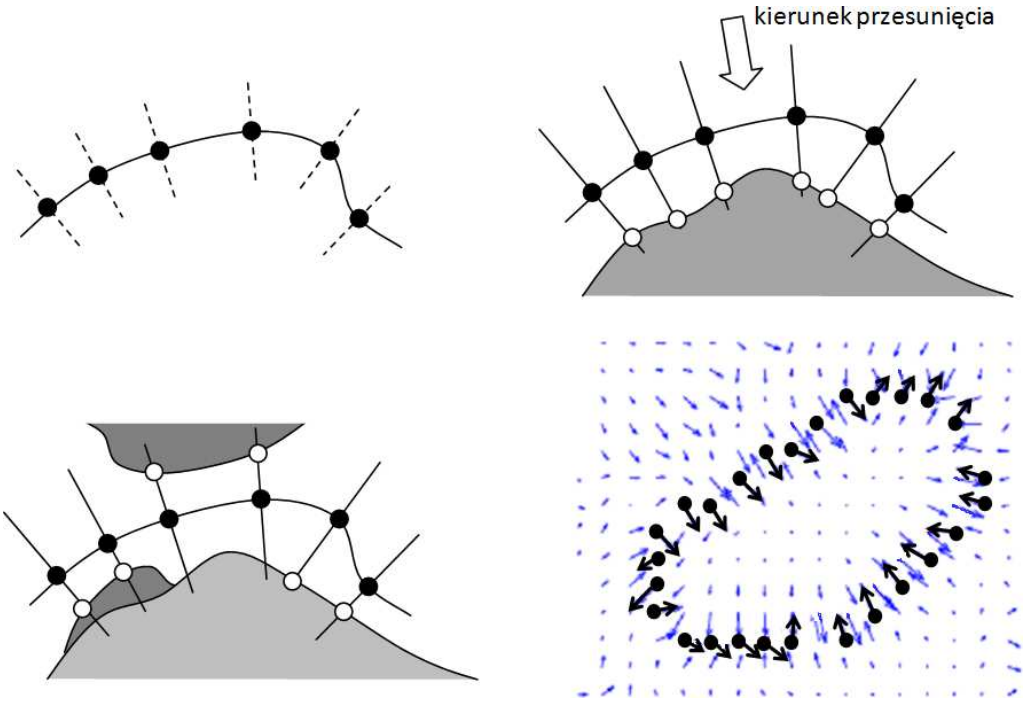
Kształt zamkniętego konturu obiektu kontrolowany jest za pomocą zbioru punktów węzłowych, odpowiednio gęsto i równomiernie rozłożonych wzdłuż konturu i "poruszających się" według dozwolonych ścieżek. Przemieszczający się po obrazie kontur obiektu jest zwykle gładką krzywą (np, sklejana), rozpinaną na podstawie przestrzennego rozkładu punktów węzłowych. Normalna do krzywej w każdym punkcie węzłowym wyznacza kierunek iteracyjnego przesuwania punktu, a nowe jego położenie winno w założeniu zbiegać coraz bliżej krawędzi opisywanego obiektu. W wersji podstawowej położenie to definiuje piksel – punkt liniowego sąsiedztwa wzdłuż kierunku normalnej, w którym występuje największy gradient lokalnego, przestrzennego rozkładu funkcji jasności (rys. 3.27).

Bardziej formalnie, proces optymalizacji kształtu i położenia konturu opisuje energetyczna funkcja kosztu, której minimalizacja w kolejnych krokach porusza i zniekształca krzywą. Pierwszy z członów funkcji kosztu, wewnętrzny, czyli dotyczący właściwości samej krzywej kontroluje jej regularność, gładkość rozstrzygając wpływ dwóch składników działających w przeciwnych kierunkach: odpychający (rozciągający) i przyciągający (kurczący). Ustalając parametry tego członu możemy spowodować usztywnione lub bardziej elastyczne zachowanie krzywej.

Drugi człon – obrazowy działa na podstawie analizy lokalnych gradientów, ogólniej kontroluje proces wpasowywania w obiekty obrazu o pożądanych właściwościach (przede wszystkim silnie zarysowanych krawędziach). Za jego pomocą można modelować kierunek i zakres przesuwania punktów konturu w kolejnych krokach dopasowania, a także ustalić warunek stopu.

Trzeci człon odnosi się do nakładaniu przez użytkownika dodatkowych ograniczeń, interaktywnych modyfikacji parametrów procesu modelowania obiektów czy też uwzględnienie specyficznej wiedzy określającej innego typu cechy opisywanych obiektów.

Iteracyjna procedura przemieszczania po obrazie konturu o pożądanych właściwościach, w kierunku większych gradientów ma wiele zalet, przede wszystkim stosunkowo mały koszt obliczeniowy, możliwość naturalnej interakcji z użytkownikiem czy wykorzystania wiedzy *a priori*. Ma też sporo ograniczeń – musi być zwykle kontrolowana przez użytkownika, tak co do uwarunkowań punktu startowego (obrót wzorca, skalowanie, wstępne nałożenie na modelowane struktury obrazu), jak też warunków stopu czy korekty parametrów modelu. Nie zawsze optymalnym jest punkt z największym gradientem (kontur może nam wtedy "uciec" do innego, znajdującego się w pobliżu obiektu - rys. 3.27), ze względu na ewentualne szumy wektor gradientu warto ustalać na podstawie większego, uśrednionego sąsiedztwa, albo odwoływać się całego kształtu profilu krawędzi wzdłuż normalnej itp. Pomocna jest tutaj metoda aktywnych kształtów – *Active Shape Models* (ASM) [179], które wykorzystuje statystyczne modele kształtu



Rysunek 3.27: Obiektowe modele obrazów – przykład dopasowania fragmentu konturu wzorca do realnych obiektów w obrazie; u góry po lewej zaznaczono czarnymi kropami punkty węzłowe wzorcowego konturu , a linią przerywaną - orientacyjne kierunki normalnych w punktach. U góry po prawej pokazano jeden krok przesunięcia konturu poprzez wyznaczenie nowych położeń punktów węzłowych (okręgi na ciągłej linii definiującej kierunek i zakres przeszukiwań punktów o największym lokalnym gradiencie); zaś poniżej po lewo – nieco problematyczna sytuacja przemieszczenia konturu w w kierunku dwóch różnych obiektów, co ukazuje jeden z problemów dobrego dopasowania konturu. Na dole po prawej zamieszczono przykładową mapę krawędzi pozwalającą określić prędkość przemieszczania się punktów konturu (czarne kropy) w kierunku najsilniejszych krawędzi, posługując się mniej schematyczną metodą zbiorów poziomic.

obiektów, iteracyjnie deformowanych, z możliwością wpasowania w krawędzie o zadanej rozkładem charakterystyce.

Można także odwołać się do całego rozkładu lokalnych rozkładu gradientów funkcji jasności, ich kierunku oraz wartości, i wyznaczyć tzw. mapę krawędzi $\vec{M}(x,y)$ (tj. zbiór wektorów zaczepionych w kolejnych punktach obrazu i wskazujących kierunek do znajdujących się w pobliżu krawędzi oraz ich siłę – rys. 3.27). Na tej podstawie określana jest prędkość poruszania się każdego z pikseli w kierunku najistotniejszej krawędzi (symbolizowanym jednostkowym wektorem normalnej do krawędzi $\vec{N}$) jako:

$$\vec{v}(x,y) = \frac{a_1 - a_2 c}{g(|\vec{M}(x,y)|)} \cdot \vec{N}(x,y) = v(x,y) \cdot \vec{N}(x,y) \tag{3.74}$$

gdzie g jest funkcją skalującą mapę krawędzi, c to parametr określający stopień krzywizny konturu, a a_1 i a_2 to stałe, wpływające na odpowiednio dynamizm przemieszczania się konturu oraz jego gładkość, dobierane zależnie od uwarunkowań zastosowania.

Wykorzystując wartości prędkości punktów według (3.74) można konstruować dopasowywany kontur obiektu $\mathcal{K}(t)$ w kolejnych chwilach czasowych na podstawie oszacowań prędkości punktów węzłowych konturu $v(x_\mathcal{K}, y_\mathcal{K}, t)$, co daje nam całościowy opis ruchu konturu w postaci $\mathcal{K}(t) = v(x_\mathcal{K}, y_\mathcal{K}, t) \cdot \vec{N}$.

Takie uogólnienie problemu dopasowania konturu obiektów z wykorzystaniem dodatkowego wymiaru czasu stało się podstawą metody poziomic (lub zbiorów poziomicowych) – *level sets*) [180, 181]. Wtedy funkcja jednowymiarowa modelowana jest domyślną krzywą, a krzywa – powierzchnią rozpinaną za pomocą siatki kartezjańskiej, definiowanej w interesującym nas obszarze przestrzeni.

W przypadku obrazów jest więc to metoda śledząca w czasie kształt konturu obiektu $\mathcal{K}(\mathbf{p}, t)$, opisanego wektorem parametrów $\mathbf{p}$, za pomocą powierzchni (ogólnie hiperpowierzchni dopuszczając wielowymiarowość dziedziny obrazów) postaci $\mathcal{S}(x, y, t) : \mathbb{R}^2 \times [0, T] \rightarrow \mathbb{R}$ jako

$$\mathcal{K}(\mathbf{p}, t) = \{(x, y) | \mathcal{S}(x, y, t) = 0\} \tag{3.75}$$

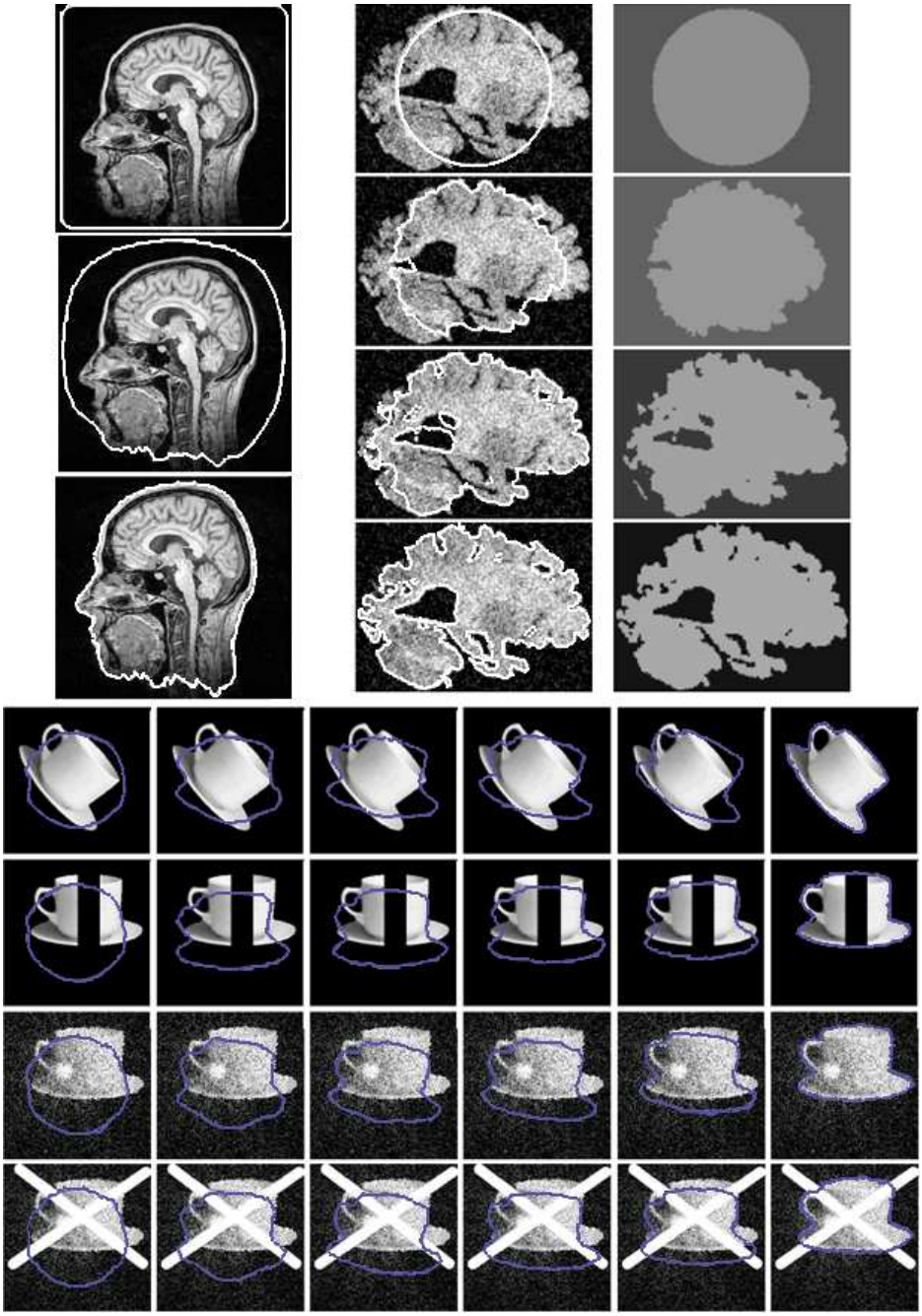
Taki model dobrze uwzględnia dynamiczny proces konturowego modelowania obiektów. Początkowa postać konturu to $\mathcal{K}(\mathbf{p}, 0) = \{(x, y) | \mathcal{S}(x, y, 0) = 0\}$.

Inaczej (3.75) można zapisać jako $\mathcal{S}(\mathcal{K}(t), t) = 0$. Hiperpowierzchnia $\mathcal{S}(x, y, t)$ może powstać na bazie czasowego rozkładu prędkości realnych punktów obrazu, zaś procedura wyznaczenia postaci konturu optymalnie wpasowanego w realne obiekty obrazów polega na zróżniczkowaniu po czasie modelu $\mathcal{S}(x, y, t)$ (liczeniu czasowych gradientów) pozwalających ustalić iterowany ruch punktów konturu. W kolejnych krokach (wartościach t) modyfikowany jest kształt konturu, zmianie ulega także postać hiperpowierzchni.

Przykładowe efekty modelowania kształtu obiektów metodą poziomic przedstawiono na rys. 3.28.

Skalowalne modele obrazów

Okazuje się, że geometryczny czy obiektowy opis obiektów może być wsparty, uzupełniony lub zastąpiony bardziej elastycznymi możliwościami geometrii analitycznej czy też potencjałem analizy funkcjonalnej. Kluczem jest zapewnienie skalowalności funkcjonalnego modelu obiektów, z zachowaniem przestrzennej i częstotliwościowej charakterystyki obrazów, dobrej rozdzielczości kątowej opisu informacji oraz niezmienniczości cech wobec przesunięć czy zmiany kontrastu. Nie należy też zapomnieć o skutecznym odseparowaniu szumu czy różnego typu zakłóceń (artefaktów) w przypadku danych rzeczywistych. Funkcjonalny schemat analizy wieloskalowej daje podstawy konstrukcji praktycznie nieograniczonej licz-



Rysunek 3.28: Przykłady obiektowego modelowania obrazów metodą poziomic. Kolejne fazy dopasowywania obrysu do realiów obrazowych obiektów, w tym dla danych złożonych (jak obraz tomograficzny głowy), będących w nietypowym położeniu, nieco odmiennym od typowego kształcie czy w warunkach silnego zaszumienia ukazują skuteczność metody; rysunek na podstawie http://www.wisdom.weizmann.ac.il/ boiman/reading/level_sets/levelset_tutorial.ppt.

by przybliżeń realnych obiektów oraz ich komponentów, wzajemnych relacji czy hierarchii zależności. Wydaje się, że takie narzędzie ma wystarczający potencjał opisu bardzo silnie zróżnicowanej klasy zmian, będącej przedmiotem interpretacji informacji obrazowej. Dlatego też skalowalne (inaczej wielorozdzielcze) modelowanie obrazów jest stale wykorzystywane w różnorodnych algorytmach obróbki obrazów.

Podstawowym narzędziem opisu sygnałów jest tutaj rozwinięcie w bazie funkcji – elementów przestrzeni funkcyjnej o określonych właściwościach, możliwie najlepiej pasujących do cech modelowanych sygnałów. Najpopularniejsze są dziś bazy falkowe, konstruowane tensorowo (jako złożenie dwóch jednowymiarowych przekształceń w kierunkach prostopadłych) w przestrzeni obrazu.

Falkowe metody przetwarzania obrazów wykorzystują przekształcenie danych obrazowych w nową dziedzinę wielu skal, podpasm częstotliwościowych oraz przestrzennego rozkładu danych obrazowych (zachowuje informację o położeniu). Wykorzystywana analiza wielorozdzielcza umożliwia dobrą charakterystykę sygnałów niestacjonarnych, w tym obrazów. Każdy ze współczynników opisuje jedynie lokalne właściwości obrazu. O tym, jak lokalne, decyduje rozmiar nośnika funkcji bazowej, czyli skojarzonego filtru służącego dekompozycji. Uzyskuje się to poprzez stosowanie funkcji bazowych przekształcenia o skończonym (dokładniej zwartym) nośniku.

Realizowana w transformacji falkowej wielorozdzielcza dekompozycja obrazu pozwala upakować, niejako "skoncentrować", energię sygnału w niewielkiej liczbie współczynników falkowych oraz uwypuklić cechy obrazu (takie jak rozkład konturów i krawędzi, własności tekstur i charakterystyka szumów), co daje większe możliwości analizy i klasyfikacji, selekcji informacji diagnostycznej czy poprawy jakości obrazów. Ważnym elementem takiej dekompozycji jest wybór podstawowej funkcji skalującej oraz falki matki (inaczej falki podstawowej), które określają właściwości bazy transformacji falkowej. Winny one uwzględniać cechy przetwarzanego obrazu, możliwie silnie upodabniając właściwości bazy do lokalnych trendów zmienności sygnału.

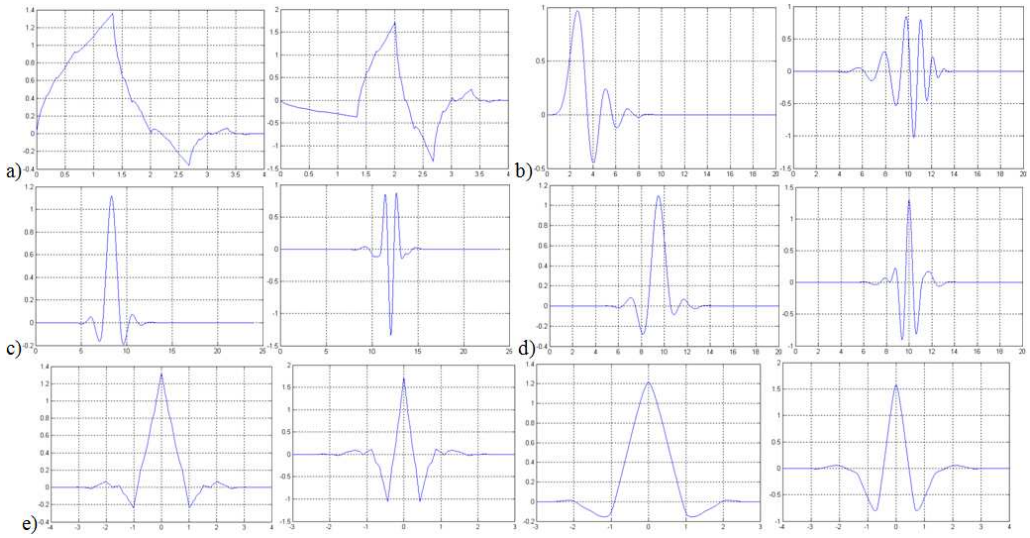
Wśród wieloskalowych reprezentacji obrazów, czy ogólnie sygnałów wyróżnić należy przede wszystkim ich rozwinięcia w bazach falkowych (ang. *wavelet*). Bazy te, rozszerzone za pomocą narzędzia pakietów falek w konstrukcje adaptacyjne, realizowane w konwencji tensorowej lub też definiowane w przestrzeni 2W (np. *contourlets*, *curvelets*, *wedgelets*, *beamlets*, *bandlets* czy też bazy falek zespolonych), z zachowaniem warunku ortogonalności bazy lub też przy słabszych warunkach, aż do ograniczenia jedynie liniowa niezależnością wektorów bazy, dają niemal nieograniczoną swobodę modelowania informacji obrazowej.

Falki Przyjrzyjmy się funkcji f, która opisuje pewne zjawisko rzeczywiste, sygnał, z którym związane jest wystąpienie określonej informacji. Funkcję tę można

scharakteryzować w sposób następujący:

- energia f jest skończona, tj. $\int |f(t)|^2 dt < \infty$;
- wartość średnia f wynosi zero, tj. $\int f(t)dt = 0$;
- silnie wyróżniona jest lokalizacja w czasie, tj. funkcja jest "lokalna" (nośnik jest zwarty, a funkcja – najlepiej rzadka w t, czyli z niewielką liczbą wartości niezerowych);
- kształt przypomina gasnące pobudzenie ośrodka, tj. falę z gasnącymi amplitudami kolejnych oscylacji oddalających się od zaburzenia centralnego.

Przykłady takich funkcji zamieszczono na rys. 3.29.



Rysunek 3.29: Przykładowe bazy funkcji skalujących i falek; kolejno: a) Daubechies 4 (z 2 momentami znikającymi), b) Daubechies 20 (z 10 momentami) [182], c) Coiflets 4 z bardziej symetryczną funkcją skalującą [183], d) Beylkin o dobrych właściwościach koncentracji energii sygnału [184], e) biortogonalna z funkcji sklejanych [185] – dwie pary funkcji skalujących i falkowych: do analizy i syntezy.

Jakkolwiek falki zwykło się rozpatrywać w kontekście transformacji falkowej, czyli przekształcenia sygnału w nową reprezentację (nowa dziedzina, nowy zbiór wartości), która wyraża informację sygnału w bardziej użytecznej formie, to warto zastanowić się na początku, jaka klasa funkcji wchodzi w grę przy projektowaniu bazy takiej transformacji.

Falkami są funkcje generowane z jednej funkcji matki ψ poprzez elementarne operacje skalowania (czyli zmiany skali czasu) przez s oraz przesunięcia o x:

$$\psi^{s,x}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-x}{s}\right), \qquad s \neq 0 \tag{3.76}$$

Falka-matka ψ jest funkcją $\psi \in \mathrm{L}^2(\mathbb{R})$ z zerową wartością średnią $\int \psi(t)dt = 0$, co wymusza co najmniej kilka oscylacji. Falka-matka jest znormalizowana $\|\psi\| = 1$ i skupiona w sąsiedztwie $t = 0$. Warunek na funkcję matkę może być też formułowany inaczej: $\int \frac{|F(\omega)|^2}{\omega} d\omega < \infty$, gdzie Ψ jest transformatą Fouriera funkcji ψ. Jeśli $\psi(t)$ zanika szybciej niż $|t|^{-1}$ dla $t \to \infty$, wówczas oba warunki są równoważne.

Falki pozostają także znormalizowane: $\|\psi^{s,x}\| = 1$. Kształt kolejnych funkcji falkowych zależy od parametru s. Jeśli $s < 1$, wówczas są to funkcje coraz węższe (następuje zmiana skali na coraz bardziej dokładną). Jeśli natomiast $s > 1$, to mamy do czynienia ze stopniowym rozszerzaniem funkcji matki (czyli przechodzeniem do skali bardziej zgrubnej).

Transformacja falkowa Aby dokonać analizy struktur sygnału o silnie zróżnicowanych rozmiarach konieczne jest wykorzystanie atomów przestrzeni czas-częstotliwość (tj. rodziny funkcji dobrze zlokalizowanych w czasie i w częstotliwości) o różnym nośniku w dziedzinie czasu. W transformacji falkowej następuje dekompozycja sygnału za pomocą bazy skalowanych i przesuwanych falek o takich właśnie cechach.

Ciągła transformacja falkowa funkcji $f \in \mathrm{L}^2(\mathbb{R})$ w dziedzinę skali s i czasu x ma postać:

$$\mathcal{W}_f(s,x) =< f, \psi^{s,x} >= \int f(t) \frac{1}{\sqrt{|s|}} \psi^* \left(\frac{t-x}{s} \right) dt \tag{3.77}$$

Reguła transformacji odwrotnej jest następująca:

$$f(t) = \frac{1}{C_\psi} \int \frac{1}{\sqrt{|s|}} W_f(s,x) \psi \left(\frac{t-x}{s} \right) ds dx \tag{3.78}$$

gdzie $C_\psi = \int |\psi(t)|^2 \frac{dt}{t} < \infty$.

Zasadniczą ideą transformacji falkowej jest więc przedstawienie dowolnej funkcji f jako superpozycji falek stanowiących jądro transformacji. Każde takie przekształcenie jest dekompozycją funkcji f na różne poziomy rozdzielczości czasowej. Jednym ze sposobów otrzymania falkowej reprezentacji funkcji f jest zapis w postaci całki po parametrach s i x rodziny funkcji falkowych $\psi^{s,x}$ z odpowiednimi współczynnikami. Ze względów praktycznych wygodniej jest wyrazić funkcję f w postaci dyskretnej superpozycji zastępując całkę operatorem sumowania. Wprowadza się wtedy zwykle dyskretyzację: $s = s_0^m, x = n x_0 s_0^m$, gdzie $m, n \in \mathbb{Z}$ oraz ustalone wartości $s_0 > 1, x_0 > 0$. Wówczas falkowa dekompozycja funkcji przedstawia się następująco:

$$f = \sum_{m,n \in \mathbb{Z}} c_{m,n}(f) \psi_{m,n}(t) \tag{3.79}$$

gdzie $\psi_{m,n}(t) = \psi^{s_0^m, n x_0 s_0^m}(t) = s_0^{-m/2} \psi(s_0^{-m} t - n x_0)$, a $c_{m,n}$ to współczynniki falkowe (w dziedzinie falkowej).

Dla wartości $s_0 = 2, x_0 = 1$ istnieje możliwość specjalnego doboru ψ, tak że $\psi_{m,n}$ tworzy bazę ortonormalną, pozwalającą wyznaczyć współczynniki falkowe

z zależności:

$$c_{m,n}(f) =< f, \psi_{m,n} >= \int f(t)\psi_{m,n}(t)dt \tag{3.80}$$

Skonstruowano wiele różnych ortonormalnych baz falkowych, przy czym zdecydowana większość mających znaczenie praktyczne nawiązuje do matematycznego narzędzia wprowadzonego przez Meyera [186] i Mallata [187], zwanego analizą wielorozdzielczą MRA (*multiresolution analysis*), która zapewnia jednoczesne występowanie wielu skal w reprezentacji sygnał. Narzędzie to może być dobrze wykorzystane do analizy sygnałów czy obrazów z zastosowaniem baz falkowych, pozwalając jednocześnie na konstrukcję szybkich algorytmów obliczeniowych.

Schemat wielorozdzielczy Zasadniczą ideą analizy wielorozdzielczej (MRA) jest dekompozycja funkcji na jedną reprezentację niskorozdzielczą (zgrubną) oraz sekwencję reprezentacji wysokorozdzielczych (szczegółowych). Wyznaczane są wielorozdzielcze aproksymacje danej funkcji (sygnału), będące aproksymacją tej funkcji z różną rozdzielczością. Formalna definicja aproksymacji wielorozdzielczej jest następująca:

Definicja 3.1 *O aproksymacji wielorozdzielczej (Mallat)*
Aproksymacją wielorozdzielczą jest ciąg $\{V_m\}_{m\in\mathbb{Z}}$ domkniętych podprzestrzeni $\mathrm{L}^2(\mathbb{R})$, takich że:

a) $$\underset{(m,n)\in\mathbb{Z}^2}{\forall} f(t) \in V_m \Longleftrightarrow f(t - n2^m) \in V_m \tag{3.81a}$$

b) $$\underset{m\in\mathbb{Z}}{\forall} V_{m+1} \subset V_m \tag{3.81b}$$

c) $$\underset{m\in\mathbb{Z}}{\forall} f(t) \in V_m \Longleftrightarrow f(\frac{t}{2}) \in V_{m+1} \tag{3.81c}$$

d) $$\lim_{m\to+\infty} V_m = \bigcap_{m=-\infty}^{+\infty} V_m = \{0\} \tag{3.81d}$$

e) $$\lim_{m\to-\infty} V_m = \overline{\bigcup_{m\in\mathbb{Z}} V_m} = \mathrm{L}^2(\mathbb{R}) \tag{3.81e}$$

f) istnieje funkcja $\phi \in V_0$ taka, że $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ jest bazą Riesza w centralnej podprzestrzeni V_0 (3.81f)

□

Z własności (3.81a) wynika niezmienniczość V_m względem dowolnego przesunięcia proporcjonalnego do skali 2^m. Aproksymacyjne cechy MRA wynikają z własności (3.81b), (3.81d) oraz (3.81e), przy czym (3.81c) podkreśla, że rozdzielczość funkcji f maleje w V_m przy wzroście m, a energia rzutów f na V_m może być wtedy dowolnie mała (rozdzielczość jest odwrotnością skali $s = 2^m$; jeśli rozdzielczość 2^{-m} dąży do 0, wtedy tracimy wszystkie szczegóły funkcji f). Każdą funkcję $f \in \mathrm{L}^2(\mathbb{R})$ można zaproksymować z dowolną dokładnością za pomocą rzutów na V_m (aproksymacja sygnału zbiega do oryginału f przy malejącym m).

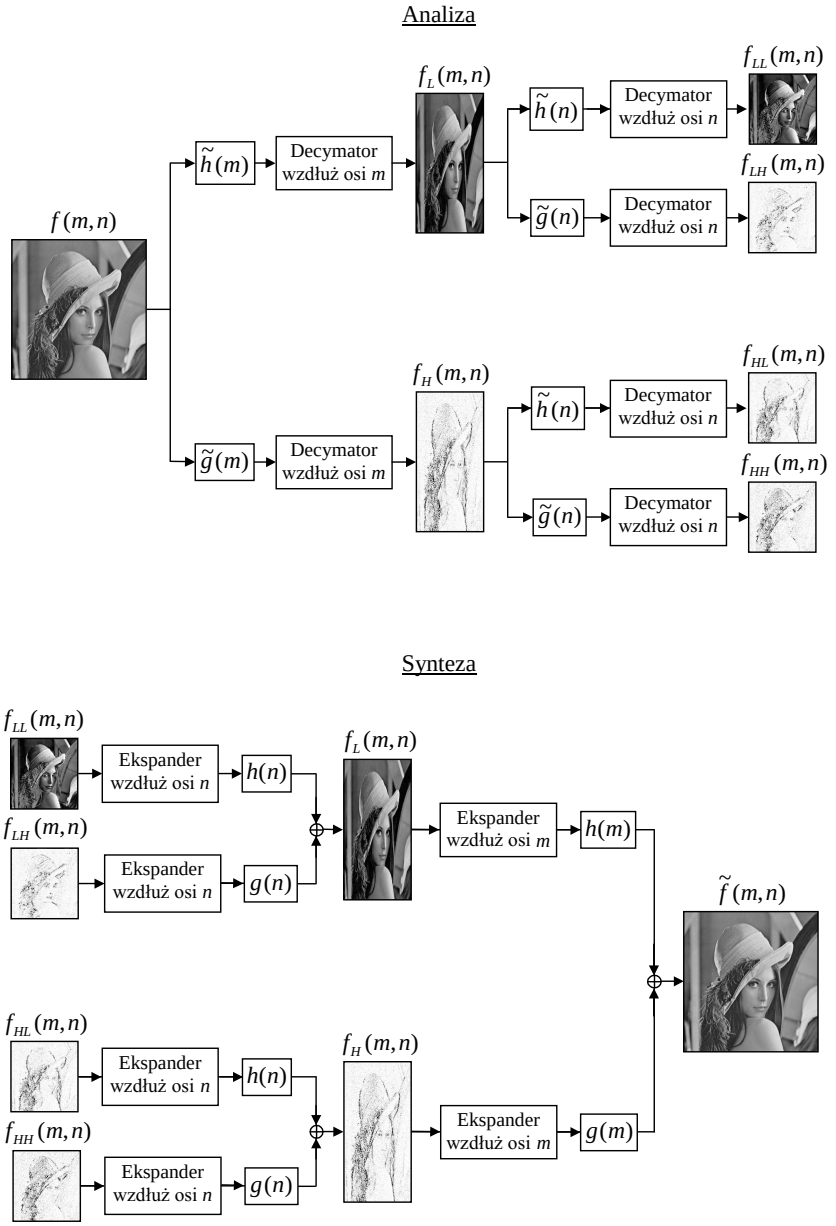
Falkowa dekompozycja obrazów Dekompozycję falkową składającą się z etapu analizy (przekształcenie w zbiór podpasm podczas kodowania) i syntezy (odtworzenie oryginału podczas dekodowania) obrazu przedstawiono na rys. 3.30. Można ją realizować na wiele sposobów poprzez dobór banku filtrów analizy $\tilde{h}, \tilde{g}$ i syntezy h, g (dolno- i górnoprzepustowych) stosowanych do przekształceń w obu kierunkach przestrzeni obrazu. Wykorzystywane są także inne sposoby podziału na podpasma (np. równomierny, *spacl*, *packet* lub *fbi* ze standardu kompresji JPEG2000, adaptacyjny z pakietami falek itp. [188].

Ze względu na niestacjonarny charakter modelu źródła informacji, dużą różnorodność cech różnego typu obrazów, różny poziom jakości przetwarzanych obrazów (stosunek sygnału do szumu, przestrzenna rozdzielczość, częstotliwościowe widmo sygnału itp.) brakuje skutecznych (tj. niezawodnych według przyjętego kryterium optymalności) metod doboru banków filtrów falkowych zależnie od zastosowania (przede wszystkim definicji sygnału, czyli informacji użytecznej diagnostycznie, oraz szumu).

Uzyskana struktura podpasm w postaci hierarchicznego drzewa dekompozycji Mallata została przedstawiona na rys. 3.31. Cztery podpasma składowych o najniższych częstotliwościach stanowią najwyższy poziom tego drzewa. Dane należące do tego poziomu nie mają rodzica i są rodzicami pierwszej generacji dla wszystkich skojarzonych przestrzennie współczynników. Każdy współczynnik, na poziomie różnym od podstawy drzewa, rozrasta się w grupę czterech współczynników kolejnego poziomu dokładniejszej skali, będąc z nimi w relacji rodzic-dzieci.

Pierwsze w hierarchii podpasmo najniższych częstotliwości LL_3 zawiera najwięcej informacji o obrazie (średnio na pojedynczy współczynnik). Potem występują kolejne podpasma największej skali: HL_3, LH_3 oraz HH_3, gdzie L oznacza podpasmo po filtracji dolnoprzepustowej, a H - górnoprzepustowej (najpierw po wierszach, potem po kolumnach - zgodnie z rys. 3.30). Zależności pomiędzy współczynnikami tych podpasm określa horyzontalna relacja drzewa dekompozycji, wyrażająca podobieństwa treści częstotliwościowych kolejnych podpasm tej samej skali w danym miejscu przestrzeni. Następnie w hierarchii są trzy podpasma drugiego poziomu drzewa: HL_2, LH_2, HH_2, których współczynniki pozostają w relacji rodzic-dzieci w stosunku do współczynników odpowiednich podpasm zarówno bardziej zgrubnej, jak i dokładniejszej skali. Na najniższym poziomie drzewa znajdują się trzy podpasma najdokładniejszej skali z indeksem 1, które nie mają węzłów potomnych.

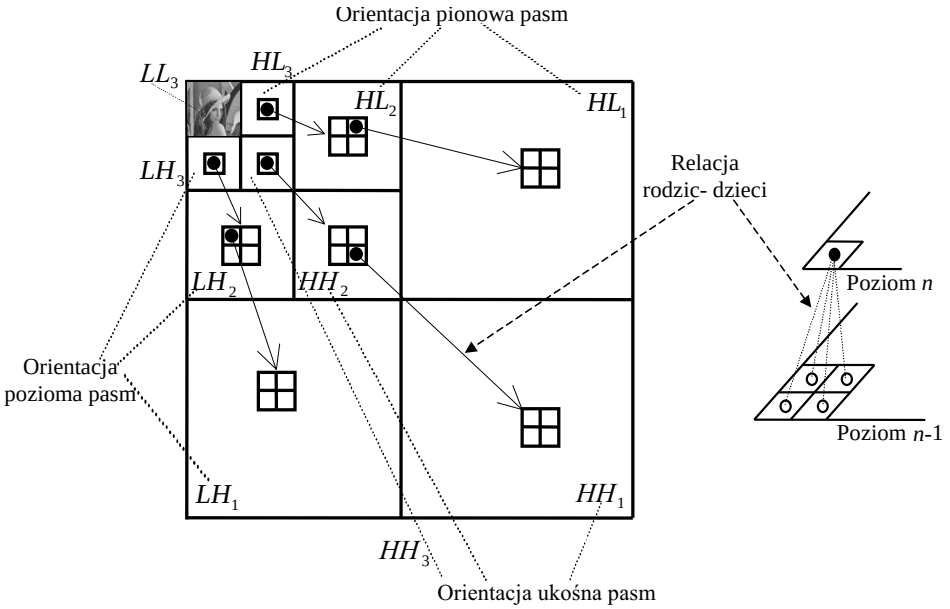
Semantyczne rozszerzenia modelu Zasadniczym celem stosowania bardziej zaawansowanych metod analizy wielorozdzielczej jest możliwie silne wydzielenie – ekstrakcja sygnału użytecznego, co podkreśla istotną rolę czynnika semantycznego. Sygnałem jest struktura, zbiór cech, region lub rozkład krawędzi, które reprezentują informację (tj. wpływają na proces interpretacji obrazu przez odbiorcę –



Rysunek 3.30: Falkowa analiza oraz synteza obrazu w algorytmach przetwarzania obrazów.

użytkownika). Elementami czy składnikami nieużytecznymi jest to wszystko, co się nie liczy, nie ma znaczenia w ocenie i interpretacji obrazów.

Falki pozwalają lepiej opisać rzeczywistość, a od trafnego opisu już tylko krok do zrozumienia, czyli odkrycia treści. Falkowe metody reprezentacji sygnałów skutecznie modelują sygnał - jego właściwości, istotę zawartej treści i formę charakterystycznych uwarunkowań jej występowania. Nowa reprezentacja ułatwia



Rysunek 3.31: Podstawowy schemat falkowej dekompozycji obrazu.

analizę i przetwarzanie sygnału, rozpoznanie treści czy przypisanie semantyki określonych cechom sygnału.

Poszukiwanie reprezentacji – skalowalnych modeli obrazów, w której ukryta/subtelna informacja obrazowa może zostać opisana i wyekstrahowana w procesie przybliżania, szacowania czy estymowania było i jest celem wielu badań z obszaru analizy obrazów. Często szczególnego znaczenia nabierają małe, lokalne struktury, drobne zmiany w charakterze dominującego tła czy wewnątrz większego obiektu. Metoda opisująca to istotne, aczkolwiek niewielkie i bardzo lokalne zaburzenie rozkładu energii sygnału ma zwykle znikomy udział w energetycznym widmie całego sygnału, a jego specyfika częstotliwościowa ginie w rozległym widmie całego obrazu.

Przy modelowaniu większych obiektów, analiza cech teksturowych lub też dających się ustalić zarysów może wystarczyć do ich różnicowania, wydzielenia czy ekstrakcji. Wieloskalowa dziedzina obrazów, w wielu przypadkach pozwala te cechy zdecydowanie bardziej uwydatnić.

Bardzo korzystną formą tej nowej dziedziny okazały się przeciwdziedziny przekształceń wieloskalowych, pozwalające analizować lokalne cechy sygnałów z różną rozdzielczością przybliżeń. Bazy tych przekształceń definiują postać szczególnie dobrze opisywalnych obiektów lokalnych. Umożliwia to charakterystyki zdecydowanie bardziej wyraziste, poprzez koncentrację energii nowej dziedziny właśnie wokół tych obiektów. Pozostaje jedynie dobrać postać bazy do lokalnych zaburzeń, które są ukryte w dziedzinie źródłowej, by wyrazić przypisaną im, poszukiwaną informację obrazową.

Istotą reprezentacji informacji w wielu skalach jest więc dobieranie bazy przekształceń do wzorców interpretowanych zmian celem skutecznego opisania zmian (sygnatur) obiektów ważnych. A że wzorce te są zwykle względne, jedynie orientacyjne, zmuszeni jesteśmy posługiwać się jedynie przybliżeniami, zarysami, zestawami cech różnicujących, do których w zależności od szeregu istotnych uwarunkowań dobierane są bazy przekształceń wieloskalowych. Kryteria tego doboru zależą od zastosowań, ale też od specyfiki konkretnego systemu obrazowania, wymagań użytkownika, a także konkretnej formy manifestacji informacji w danym przypadku. Dobór odpowiedniej bazy dokonywany jest mniej lub bardziej formalnie, z wykorzystaniem rachunku wariacyjnego i metod regularyzacji, z wykorzystaniem formalnej wiedzy *a priori* lub skutecznej adaptacji *a posteriori*, z możliwością dostosowania do niekiedy bardzo wysublimowanych, złożonych opisów poszukiwanego cech sygnału.

Probabilistyczne modele obrazów

Nie sposób stworzyć deterministycznej miary dopasowania modelu do realnego zbioru opisywanych danych, nawet przy bardzo dobrych, zwartych modelach o szerokiej różnorodności. Niepewność pozwala zmierzyć informację, pozwala także ja sklasyfikować rozpoznając znaczenie obiektów na różnych poziomach semantycznej hierarchii.

Do najczęściej stosowanych należą modele wykorzystujące modele Markowa, ukryte Modele Markowa, mieszanina rozkładów Gaussa, uzupełniona o modelowanie najbliższego kontekstu w obrazie.

Stochastyczne modele obrazów są naturalną konsekwencją probabilistycznej koncepcji źródeł informacji (zobacz p. 2.1.3), z informacją określaną poziomem niepewności odbiorcy. Dodatkowym uzasadnieniem losowych modeli obrazów jest stochastyczny charakter fizyki technik obrazowania oraz obecność szumów i artefaktów jako nieuniknionych "towarzyszy" procesu rejestracji obrazów rzeczywistych. Pojęcie szumów, zwykle niezależnych od sygnału, addytywnych, można rozszerzyć na inne składniki czy czynniki nieinformatywne, o charakterze przypadkowym z punktu widzenia oczekiwań odbiorcy informacji.

Nie sposób przy takich założeniach stworzyć miary deterministycznej, która jest skuteczna w dopasowaniu modelu do realnego zbioru opisywanych danych, nawet przy bardzo dobrych, zwartych modelach o szerokiej różnorodności. Niepewność pozwala zmierzyć informację, określać odległość funkcji losowych, pozwala także ją sklasyfikować rozpoznając znaczenie obiektów na różnych poziomach semantycznej hierarchii.

Stochastyczny model obrazu winien uwzględniać podział na regiony jako obiekty o swej specyfice, odmiennych cechach teksturowych i różnym znaczeniu. Metoda wyznaczenia modelu powinna uwzględniać przede wszystkim różnice w rozkładzie wartości pojedynczych pikseli (np. piksele o małych wartościach tworzą

region A, a piksele o dużych wartościach – region B), jak też zależności kontekstowe, czyli lokalne podobieństwo wartości grup pikseli leżących w najbliższym sąsiedztwie (np. piksele, których najbliżsi sąsiedzi mają bardzo zbliżone wartości tworzą region A, zaś piksele o sąsiedztwie bardzo różnicowym co do wartości – region B).

Modele te wykorzystują koncepcję pola losowego, dobierane klasy rozkładów, modele Markowa, ukryte Modele Markowa, mieszaninę rozkładów Gaussa, metody redukcji przestrzeni, itp. Globalny opis cech jest uzupełniany modelowaniem najbliższego kontekstu w obrazie. Podstawowe założenia ergodyczności i stacjonarności (np. w ustalonych wstępnie regionach symulujących obiekty) uzupełniane są szeregiem ograniczeń dotyczących rodzaju rozkładu, wartości parametrów rozkładu, sposobu grupowania, kryteriów przybliżeń, itd.

Obraz jako pole losowe. Zmienna losowa jako dowolna funkcja o wartościach rzeczywistych określona na zbiorze zdarzeń elementarnych, opisuje wartości zdarzeń. W przypadku opisu obrazu są to wartości zarejestrowanych pikseli $f(k,l)$ – przyjmijmy dla ustalenia uwagi, że są to wartości funkcji jasności, np. z zakresu 0–255 (dane bajtowe): $f(k,l) \in A_f = \{0,\ldots,255\}$.

Ciąg takich zdarzeń, opisanych wektorem uporządkowanych wartości pikseli z dziedziny obrazu można z kolei wyrazić za pomocą rodziny zmiennych losowych $\mathbf{F} = \{F_i\}$, gdzie indeks $i \in [1,\ldots,I]$ przy $I = K \cdot L$ przebiega naturalną dziedzinę obrazu $(k,l) \in \Omega_f$ – dwuwymiarową strukturę o wymiarach $K \times L$ – według porządku określonego bijekcją domkniętego przedziału liczb $[1,\ldots,I]$ w Ω_f, tak że $i = \xi(k,l)$. W najprostszym przypadku i wskazuje piksele wzdłuż kolejnych wierszy, chociaż rozkład zależności wartości pikseli sugeruje porządkowanie obrazu wzdłuż krzywych preferujących sąsiedztwo wielokierunkowe.

Wartości takiej funkcji losowej leżą w przestrzeni definiowanej przez szereg zmiennych losowych w ramach ich konkretnych realizacji – wartości f_i. Określony w ten sposób łańcuch losowy, czyli proces losowy zdefiniowany na dyskretnej przestrzeni stanów (wartości zmiennych losowych rodziny), nazywamy **losowym polem obrazu F** z konkretną realizacją – obrazem: $\mathbf{f} = (f_1, f_2, \ldots, f_I)$. $\mathbf{F}$ nazwijmy polem wartości (pikseli).

Uwzględniając obiektowy charakter informacji obrazowej, definiowany jest uzupełniający, analityczny model losowego pola obrazu. Każdy element dziedziny obrazu, tj. piksel, opisany jest tam etykietą e_i przynależności do regionu (obiektu) R_k, z wartościami wskazującymi na określony region $e_i = k \in A_e = \{1,\ldots,K\}$ (dla uproszczenia tło potraktujmy jako obiekt uzupełniający). Model $\mathbf{E} = \{E_i\}_{i=1}^{I}$ nazwijmy polem etykiet (pikseli) z wymagającą ustalenia realizacją $\mathbf{e} = (e_1, e_2, \ldots, e_I)$.

Probabilistyczny opis pola losowego stanowi łączny rozkład prawdopodobieństwa $P_{\mathbf{F}}(\mathbf{f})$, będący w przypadku dyskretnym funkcją rozkładu prawdopodobień-

stwa (dyskretną wersją funkcji gęstości prawdopodobieństwa). Wyznaczenie łącznego rozkładu prawdopodobieństwa na podstawie pojedynczych realizacji źródła informacji obrazowej jest praktycznie niemożliwie (brakuje generycznych teorii pozwalających wyznaczyć taki model). Wykorzystuje się wiedzę *a priori*, doprecyzowanie cech opisu ze względu na realizację określonych celów, uwzględnia pożądanych właściwości tworzonego opisu, itp., aby stworzyć parametryzowany, uproszczony rozkład prawdopodobieństwa dający wiarygodny opis informacji.

Najprostszy histogram. Silnie uproszczonym, a przy tym często stosowanym opisem właściwości obrazu jest histogram, czyli zliczenie liczby wystąpień każdej z możliwych wartości piksela – zobacz (3.5) wraz z opisem. Tworzona na jego bazie funkcja rozkładu prawdopodobieństwa wartości pikseli $p_f(\cdot)$ abstrahuje od lokalizacji wystąpienia tych wartości, traktując dane jako kolejne wartości zmiennej losowej modelującej obraz.

Przy uogólniających założeniach częstościowego szacowania prawdopodobieństw na podstawie możliwie licznych zbiorów danych, przy realizacji źródła o ustalonym alfabecie, stacjonarnego, po normalizacji sumy liczby wystąpień, uzyskujemy dyskretny rozkład prawdopodobieństw poszczególnych symboli alfabetu źródła (funkcję rozkładu prawdopodobieństwa).

Opisanie obrazu za pomocą histogramu ma dwa zasadnicze ograniczenia: brak wydzielenia obiektów kształtujących treść (wartości pikseli dowolnego obrazu naturalnego można losowo rozrzucić w polu obrazu, zachowując pokrycie zupełne, uzyskując zatarcie treści przy dokładnym zachowaniu postaci histogramu) oraz brak opisu lokalnych zależności pomiędzy sąsiednimi pikselami, które konstytuują obiekty.

Pierwsze z tych ograniczeń, przy dodatkowych założeniach dotyczących klasy rozkładów opisujących poszczególne obiekty, zostało wyeliminowane w modelu skończonej mieszaniny (*finite mixture model*) regionów (obiektów), w skrócie FMM. FMM można rozszerzyć o lokalną analizę rozkładu etykiet pikseli w najbliższym sąsiedztwie.

Lokalną charakterystykę danych poprzez ustalenie określonych, różnicujących właściwości regionów, można także zrealizować za pomocą lokalnych histogramów. Otrzymany zestaw histogramów lokalnych stanowi wraz z zarysem regionów bardziej złożony opis danych obrazowych.

Precyzyjniejszy opis kontekstowych zależności danych można realizować za pomocą modeli Markowa (jawnych czy ukrytych) rozszerzających pojęcie kontekstu na najbliższe sąsiedztwo w wielowymiarowej dziedzinie określoności danych obrazowych. bazujących na wyższych rzędów (pole losowe Markowa), w tym także modeli ukrytych.

Mieszanina regionów. Uwzględnienie w modelu podziału obrazu na regiony wymaga połączenie pól wartości i etykiet. Prawdopodobieństwo wartości określonego piksela ustalamy na podstawie możliwej jego przynależności do poszczególnych regionów $P(f_i, e_i = k) = P(f_i, i \in R_k) = P(f_i, R_k)$, gdzie $k = 1, \ldots, K$. Korzystając z prawdopodobieństwa warunkowego mamy:

$$P(f_i, R_k) = P(R_k)P(f_i|k) \tag{3.82}$$

zaś rozkład brzegowy dla $F_i = f_i \in A_f$ daje zależność:

$$P(f_i) = \sum_{k=1}^{K} P(f_i, k) = \sum_{k=1}^{K} P(R_k)P(f_i|R_k) \tag{3.83}$$

Ustalając oznaczenie $\pi_k = P(R_k)$, a więc $\pi_k > 0$ i $\sum_{k=1}^{K} \pi_k = 1$, otrzymujemy parametr modelu, który w pierwszym przybliżeniu charakteryzuje liczność poszczególnych regionów. Prawdopodobieństwo wartości poszczególnych pikseli obrazu f_i jest więc konstruowane na podstawie prawdopodobieństw warunkowanych ich przynależnością do poszczególnych regionów R_k:

$$P(f_i) = \sum_{k=1}^{K} \pi_k P_k(f_i) \tag{3.84}$$

Warunkowe prawdopodobieństwo przynależności piksela do regionu: $P_k(f_i) = P(f_i|R_k)$ poszukiwane jest zwykle w klasie rozkładów opisanych sparametryzowanym, uogólnionym rozkładem Gaussa:

$$P_k(f_i) = g_k(f_i \mid \mu_k, \sigma_k, \alpha) = g_k(f_i) = \frac{\alpha\beta_k}{2\Gamma(1/\alpha)} \exp[- \mid \beta_k(f_i - \mu_k) \mid^{\alpha}] \tag{3.85}$$

gdzie $\beta_k = \frac{1}{\sigma_k}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}$, zaś funkcja gamma $\Gamma(x) = \int_0^{\infty} u^{x-1}e^{-u}du$. Średnia μ_k oraz odchylenie standardowe σ_k określają statystyczne parametry rozkładu w poszczególnych regionach. Globalny parametr $\alpha > 0$ modeluje kształt rozkładu w sposób zgodny dla wszystkich regionów. Różnicowanie kształtu rozkładów w poszczególnych R_k jako α_k wymaga większych nakładów obliczeniowych i komplikacji procedury optymalizacyjnej modelu, nie dając przekonującej poprawy wiarygodności modelu.

Często stosowany rozkład normalny (Gaussa) ($\alpha = 2$ w równaniu (3.85)) daje następującą postać rozkładu:

$$P_k(f_i) = g_k(f_i \mid \mu_k, \sigma_k, \alpha = 2) = g_k^{(2)}(f_i) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(f_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{3.86}$$

Wykorzystując prawdopodobieństwo wartości piksela (3.84) z doprecyzowanym rozkładem prawdopodobieństw przynależności do regionów według (3.85) można zdefiniować model całego obrazu będącego mieszaniną regionów.

Opis obrazu. Według koncepcji FMM modelowany obraz $\mathbf{f}$ jest konkretną, zadaną realizacją szukanego pola $\mathbf{F}$ niezależnych zmiennych losowych: $\mathbf{f} \in \mathbf{F}$ o parametrach lokalnych i globalnych dobieranych według koncepcji największej wiarygodności (poszukiwanie takiego zbioru wartości parametrów modelu $\mathbf{F}$, by $\mathbf{f}$ był najbardziej prawdopodobny). Przyjmuje się więc, że wartość każdego piksela generowana za pomocą modelu źródła informacji obrazowej FMM generowana jest niezależnie, bez względu na uwarunkowania kontekstowe, stąd pojawia się iloczyn prawdopodobieństw wszystkich pikseli tworzących modelowany obraz. Warto zauważyć, że uprawniona jest charakterystyka takiego modelu za pomocą histogramu generowanych danych [189].

Funkcja rozkładu podobieństwa w modelu FMM z rozkładem gaussowskim, tj. modelu GMM (*Gaussian Mixture Model*) ma więc postać ogólną:

$$P(\mathbf{f}) = \prod_{i=1}^{I} \sum_{k=1}^{K} \pi_k g_k(f_i) \tag{3.87}$$

Model ten pozwala wygenerować (syntetyzować) przybliżony obraz $\tilde{\mathbf{f}} = \mathcal{G}(P(\mathbf{f})) \approx \mathbf{f}$.

Taka konstrukcja modelu nie zakłada spójności regionów, które należy według (3.83) rozumieć jedynie jako klasy różnicujące wartości poszczególnych. Konieczne jest uzupełnienie modelu w zakresie zależności wartości sąsiednich, w tym przede wszystkim należących do tego samego regionu. Właściwe rozdzielenie pikseli na spójne regiony warunkuje skuteczność tego modelu.

GMM opisuje przede wszystkim statystykę danych obrazowych, nie uwzględniając zależności pomiędzy pikselami. Bazuje na iloczynie prawdopodobieństw występowania wartości poszczególnych pikseli. Zakłada podział pikseli na regiony, czyli rozkład pola etykiet jako informację a priori.

Generalnie, FMM zakłada jako podstawowy model obrazu – pole losowe $\mathbf{F}$ zmiennych losowych niezależnych o jednakowym rozkładzie (*independent and identically distributed*) wewnątrz K regionów (obiektów) $R_k, k = 1, \ldots, K$. Liczba regionów jest globalnym parametrem modelu, wpływającym liniowo na jego złożoność.

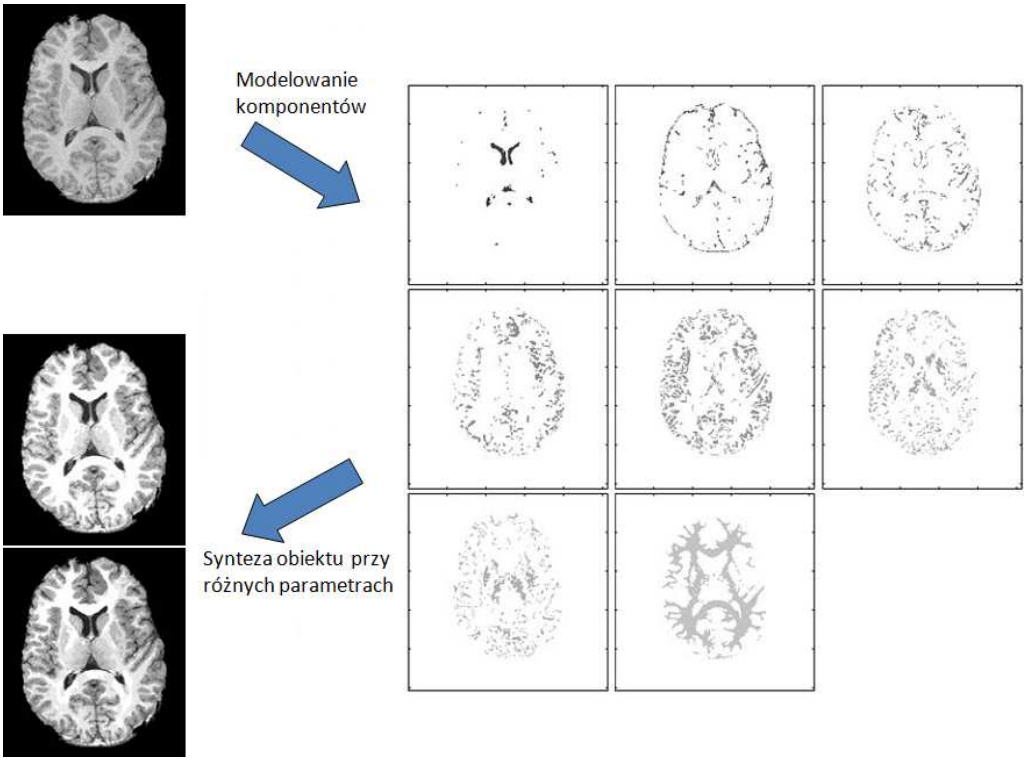
Pomimo ograniczeń, GMM znajduje szerokie zastosowanie w modelowaniu obrazów, zwykle uzupełniony o różne koncepcje opisu kontekstowego [190, 191].

Przykładowe efekty modelowania obrazów z wykorzystaniem mieszaniny regionów pokazano na rys. 3.32.

3.1.5 Metody analizy danych

Zróżnicowanie metod analizy danych, będące próbą pewnego uogólnienia i syntezy problemu analizy na przykładzie danych obrazowych przedstawiono na rys. 3.33.

Wśród metod analizy danych szczególne miejsce zajmują metody dotyczące obrazów. Tak trudne zadania jak rozpoznawanie wzorców określonej klasy obiek-



Rysunek 3.32: Przykład zastosowania modelu probabilistycznego gaussowskiej mieszaniny regionów do opisu tkanki mózgowia w obrazowych badaniach mózgu metodą rezonansu magnetycznego; na podstawie obrazu źródłowego (u góry po lewej) wydzielono regiony reprezentujące różne rodzaje tkanek mózgowia, w znacznym stopniu korelujące z semantycznym atlasem mózgu (rysunek na podstawie [191]).

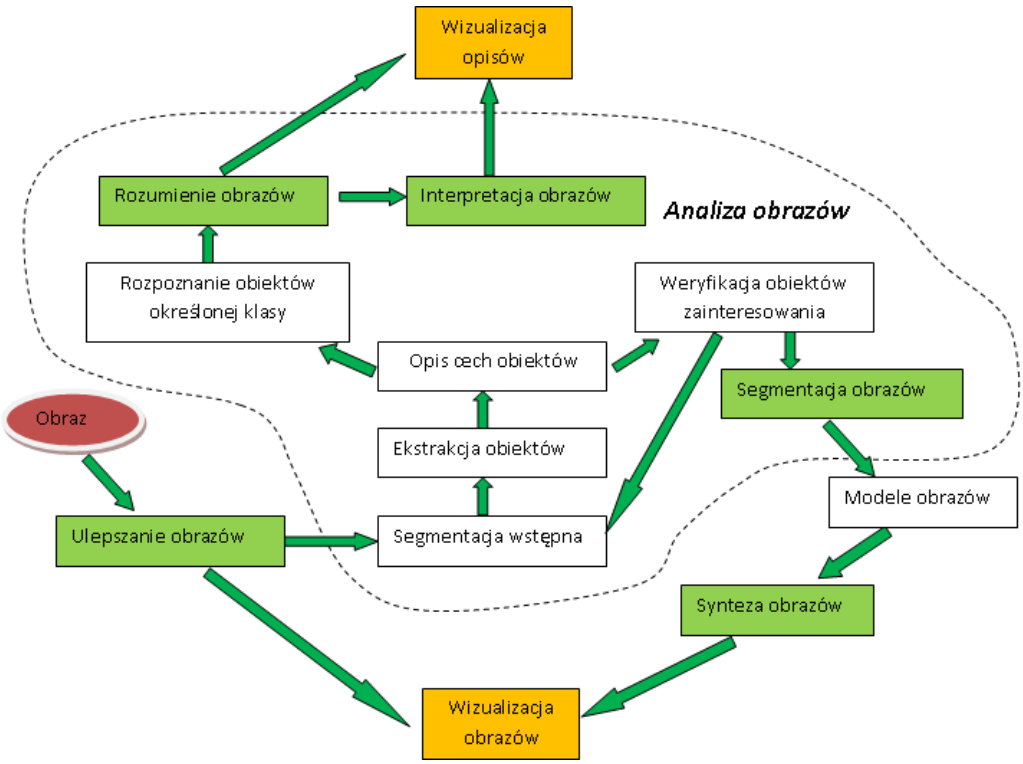
tów, liczenie deskryptorów różnicujących treściowo opisywane struktury czy próby automatycznej interpretacji obrazów stanowią prawdziwe wyzwanie w przypadku analizy złożonej informacji obrazowej.

Metody analizy obrazów można uporządkować, zgodnie z ich przeznaczeniem, według koncepcji służących:

- wydzieleniu składników treściowych, takich jak przestrzenne regiony, potencjalne obiekty, czy komponenty opisujące obrazy w innej dziedzinie (np. z wykorzystaniem skalowalnych modeli rozwinięć w bazach funkcji);

- opisowi cech wydzielonych składników, np. z wykorzystaniem procedur generacji czy selekcji cech według ustalonych kryteriów;

- rozpoznaniu określanej klasy obiektów, na bazie ustalonej przestrzeni cech i dobranych klasyfikatorów, lub też selekcji obiektów ze względu na ustaloną



Rysunek 3.33: Metody analizy danych na przykładzie zastosowań obrazowych; analiza ukazana została w kontekście innych form przetwarzania i wizualizacji obrazów.

hierarchię ich właściwości (np. przy wyszukiwaniu indeksowanych obrazów, zgodnie z wolą użytkownika formułowaną w zapytaniu);

- rozumieniu treści obrazów, poprzez odniesienie zestawu cech obliczeniowych do sformalizowanej wiedzy dziedzinowej, celem wspomagania interpretacji informacji przekazu obrazowego.

Podstawowym działaniem analitycznym jest wydzielenie regionów stanowiących potencjalne obszary zainteresowania, kandydatów na spodziewane obiekty czy też inne elementy, mogące mieć wpływ na użytkowanie obrazów. Zadanie to realizowane jest za pomocą różnych metod segmentacji obrazów, odwołujących się do zasadniczych koncepcji modelowania danych obrazowych.

Segmentacja

Segmentacja to wydzielenie z obrazu obszarów (inaczej regionów spójnych, czyli będących *w jednym kawałku*), charakteryzujących się:

a) **jednorodnością** względem określonej właściwości danych (atrybutu), przy czym kryterium jednorodności może dotyczyć zarówno wewnętrznych cech

obiektu (np. kolor czerwony w sensie zdefiniowanego deskryptora koloru), jak też mierzalnych granic obszarów (konturów) – w takim rozumieniu rozgranicza obszary wyraźna krawędź nawet jeśli poza obszarem granicznym wydzielone regiony są jednorodne;

b) **czytelnym znaczeniem** (funkcją semantyczną), wynikającym ze specyfiki zastosowań; ustalenie znaczenia segmentowanych obszarów może się odbywać na drodze interaktywnej, może też być realizowane metodami rozpoznawania obiektów, bazującymi na sformalizowanej wiedzy dziedzinowej.

Zwykle algorytmy segmentacji bazują na określonym kryterium jednorodności, dyktowanym aplikacją, a są weryfikowane według kryterium semantycznego. Niekiedy jednak stosuje się porządek odwrotny – inicjacja dotyczy wskazania ogólnych zarysów użytecznych dla użytkownika regionów, a następnie procedura wykorzystująca ustalone kryterium jednorodności dookreśla słabo dostrzegalne granice obiektów. Na etapie optymalizacji poszukiwane są skuteczne tzw. deskryptory semantyczne, czyli numeryczne wskaźniki znaczenia poszczególnych fragmentów obrazu, pozwalające łączyć grupy pikseli w wiarygodne regiony o ustalonym znaczeniu.

Celem segmentacji jest uproszczenie analizy danych obrazowych poprzez zmniejszenie stopnia złożoności obrazu. Obraz zastępowany jest opisem symbolicznym z wydzielonymi obszarami – segmentami, które mogą mieć pożądane cechy, np. spójność, zwartość, odpowiedni rozmiar czy kształt, pokrycie określonym wzorcem tekstury. Wydzielone segmenty otrzymują zwykle identyfikator – etykietę.

Segmentacja stanowi bardzo ważne ogniwo w analizie treści obrazów. Umożliwia wstępny opis czy uproszczenie sceny, określenie obszaru zainteresowań w dalszej analizie, wyznaczenie potencjalnych obiektów zainteresowania, rozpoznawanych w dalszej kolejności jako obiekty określonej klasy czy też anormalności.

Bardziej formalnie, segmentacją obrazów nazywany jest proces podziału całego obrazu $\mathbf{f}$ na skończony zbiór regionów – obiektów zainteresowania $R_1, R_2, \ldots, R_n$ oraz nieistotnego znaczeniowo tła B, przy czym podział ten odznacza się następującymi właściwościami:

a) $\mathbf{f} = R_1 \cup R_2 \cup \ldots \cup R_n \cup B$ – wydobyte regiony wraz z tłem znaczeniowym stanowią kompletny opis obrazu;

b) $\forall_{i \neq j} R_i \cap R_j = \emptyset$;

c) $\forall_i R_i \cap B = \emptyset$;

d) $\forall_i J_S(R_i)$=1;

e) $\forall_{i \neq j} J_S(R_i \cup R_j)$=0.

gdzie J_S jest binarnym operatorem ustalającym jednorodność i czytelne znaczenie regionu według przyjętych kryteriów ($J_S(R) = 1 \Leftrightarrow R$ jest jednorodne i sensowne). O ile weryfikacja jednorodności obszaru bazuje najczęściej na kryterium obliczeniowym dotyczącym istotnych cech obszaru, to określenie semantycznej jego funkcji sprowadza się zazwyczaj do identyfikacji obszaru jako określonego rodzaju obiektu rozpoznawanego przez użytkownika, co wymaga odwołania się do specyficznej wiedzy dziedzinowej, algorytmów rozpoznania klasy obiektu lub też obliczenia podobieństwa w stosunku do referencyjnych wzorców.

Segmentacja wstępna jest pierwszym, bardzo istotnym etapem analizy obrazów. Wydzielone wstępnie obszary służą w typowym schemacie ekstrakcji potencjalnych obiektów (etap reprezentacji obiektów), opisywanych następnie za pomocą zestawu cech (etap pomiaru - obliczania cech obiektów). Klasyfikacja obiektów (czyli etap kolejny) może mieć zwykle dwojaki charakter:

- weryfikacji wydzielonych regionów jako obiektów przekazu informacji w obrazie, poprzez ustalenie znaczenia (semantyki) obszarów, a w konsekwencji zaliczenie ich do kategorii regionów (jeśli $J_S(R_i) = 1$) bądź tła ($J_S(R_i) = 0$);

- wstępnego rozpoznania obszarów jako obiektów o istotnym znaczeniu w celu adaptacyjnej korekty wcześniejszych etapów poczynając od segmentacji wstępnej; ta procedury służy bardziej wiarygodnemu wyznaczeniu obiektów z obrazu w celu doskonalenia efektów analizy obrazów;

- rozpoznania obiektów jako należących do określonej klasy semantycznej.

Duża różnorodność stosowanych metod segmentacji sprawia, że można je klasyfikować na wiele sposobów. Cel segmentacji zależy często od rodzaju analizowanych danych, rodzaju wyrażonej treści, oczekiwań obserwatora oraz innych uwarunkowań aplikacji (ograniczenia czasowe, dokładność wyznaczenia granic obiektów itp.). Różne zastosowania prowadzą do zróżnicowanych algorytmów tej samej koncepcji segmentacji.

Podstawowy podział rozwiązań możliwych w zakresie segmentacji obejmuje:

- metody obszarowe, znajdujące regiony według obliczeniowego kryterium jednorodności,

- metody krawędziowe, bazujące na wykrywaniu krawędzi i aproksymacji konturów rozdzielających regiony o odmiennych właściwościach,

- metody probabilistyczne, gdzie wykorzystywany jest określony model lokalnych zależności danych, statystyczne właściwości globalne, szacowanie rozkładów łańcuchów i pól losowych opisujących poszczególne regiony itp., w tym przede wszystkim:

  - progowanie (*thresholding*) - wykorzystuje prostą operację porównywania wartości jasności pikseli obrazu z ustalonym progiem; podziału metod progowania dokonuje się w zależności od sposobu wyznaczania progu:
    * globalne, z progiem obliczonym na podstawie globalnych parametrów obrazu - takich jak histogram,
    * lokalne, gdzie próg ustalany jest na podstawie parametrów lokalnych obrazu w określonym kontekście,
    * adaptacyjne, z modyfikacją progu zależną od efektów progowania w obszarach poprzednich.
  - metoda mieszania regionów,
  - metody rozmyte, wykorzystujące elementy logiki rozmytej,

- metody rozpoznania, weryfikujące wydzielane obszary jako obiekty czytelnej klasy znaczeniowej na odpowiednim poziomie abstrakcji opisu obrazu, z zastosowaniem klasyfikatorów uczących się oraz mechanizmów doboru/selekcji cech;

- metody hybrydowe, łączące np. krawędziowe, obszarowe z opisem probabilistycznym jak niektóre wersje metod aktywnych kształtów, konturów czy poziomic.

Wśród obszarowych technik segmentacji (*region based*) można wyszczególnić następujące metody:

- grupowanie, inaczej klasteryzację (*clustering methods*), polegające na klasyfikacji obszarów na podstawie określonych parametrów charakteryzujących dane obszary; najbardziej powszechnymi algorytmami grupowania są:
  - metoda K średnich (*K-means*),
  - metoda rozmyta C średnich (*fuzzy C-means*),
  - hierarchiczne grupowanie (*hierarchical clustering*),
  - mieszanina rozkładów Gaussa.

- rozrost regionów (*region growing*), bazujący na założeniu, że piksele należące do wspólnego regionu mają podobne, w sensie ustalonego kryterium, właściwości; najbardziej popularne metody rozrostu to:
  - łączenie (*merging*) - polega na dołączaniu pikseli do punktów (miniregionów) początkowych o spójnym charakterze,
  - dzielenie (*splitting*) - zaczynając od zgrubnego podziału wstępnego, dokonuje się kolejnych podziałów dużych regionów, az do momentu,

kiedy regiony tworzone przez grupy pikseli osiągną zadowalający poziom ujednolicenia,

- łączenie i dzielenie - łączy opisane wyżej algorytmy,
- *hill climbing* - wykorzystuje fakt występowania w obrazie lokalnych maksimów intensywności; maksima odpowiadają ziarnom metody rozrostu.

- segmentacja wododziałowa - jasność pikseli w obrazie jest traktowana jako pewne odchylenie od poziomu odniesienia, jako wzniesienie ponad pewien podstawowy poziom; takie podejście pozwala traktować obraz jak powierzchnię topograficzną - obszary ciemne stanowią zagłębienia "terenu", podczas gdy obszary jasne odpowiadają wzniesieniom.

Techniki wykrywające krawędzie (*edge based*) można usystematyzować w następujący sposób:

- bazujące na pochodnych pierwszego rzędu funkcji obrazu – w obrazie wyszukuje się lokalne maksima oraz minima pierwszej pochodnej funkcji obrazu; gradient wskazuje kierunek największej zmiany wartości funkcji obrazu $f(x, y)$, amplituda tego wektora określa szybkość zachodzącej zmiany;

- bazujące na pochodnych drugiego rzędu funkcji obrazu – wykrywane są przejścia przez zero drugiej pochodnej funkcji obrazu; stosowane jest dyskretny operator Laplace'a będący sumą drugich pochodnych funkcji obrazu $f(x, y)$, poprzedzone często wygładzeniem obrazu za pomocą filtru gaussowskiego – zestawienie tych dwóch operacji nosi nazwę LoG (*Laplacian of Gaussian*);

- aktywnych konturów oraz aktywnych kształtów, bazujące na wzorcach obiektów zainteresowania wpasowywanych w realia danego obrazu metodami lokalnych gradientów czy też na podstawie dokładniejszej analizy kształtu profili krawędzi występujących w najbliższym otoczeniu;

- przeszukiwania grafów – w tej technice obraz rozpatrujemy jako graf, którego wierzchołkami są pojedyncze piksele, natomiast krawędzie grafu stanowią potencjalne krawędzie obrazu; używając gradientu lub innej operacji wykrywającej krawędzie, można obliczyć koszt połączeń wewnątrz grafu; krawędź wyznacza się odnajdując w grafie połączenia o niskiej wartości kosztu.

Rozpoznanie obiektów

Rozpoznawanie treści obrazowej sprowadza się zasadniczo do rozpoznania obiektów oraz określenia wzajemnych, wiążących je relacji. Właściwie rozpoznana treść

jest kluczowym etapem komputerowej analizy obrazów, bo umożliwia jej zrozumienie. Rozumienie to jest z kolei warunkiem odbioru pełnego przekazu informacji i jego interpretacji.

Skuteczność algorytmizacji i praktycznej realizacji całego tego procesu zależy w pierwszej kolejności od efektywnego rozpoznawania obiektów, ogólniej określonego rodzaju wzorców zależnych od zastosowania, np. twarzy lub jej części, guza w obrazie medycznym czy określonego rodzaju komórki w badaniach mikroskopowych. Rozpoznawanie polega często na tworzeniu dodatkowego opisu obrazu, który pozwoli lepiej różnicować poszczególne obszary, wydzielać z tła odmienne fragmenty czy formować obiekty z grup pikseli o zbliżonych cechach.

Rozpoznawanie może mieć różne cele:

1. przydzielenie wyznaczonych wcześniej obiektów (wzorców), np. za pomocą wybranej metody segmentacji, do określonej klasy, przy czym zbiór klas może być ustalony z góry, bądź też otwarty (np. grupowane wstępnie obiekty w części dają się rozpoznać jako danej klasy, w części zaś zajmują nie opisane dotąd obszary przestrzeni cech – na tej podstawie można rozszerzyć zbiór klas rozpoznawanych obiektów);

2. wyszukanie określonego rodzaju obiektu w obrazie, przy czym istotne może się okazać jedynie potwierdzenie jego obecności lub też wyznaczenie dokładnej jego lokalizacji, kształtu czy innej specyfiki;

3. ocenę grupy obiektów w obrazie pod kątem ich przynależności do wybranej klasy;

4. wydzielenie obiektów nienależących do klasy obiektów pożądanych, np. odrzucenie podróbek wartościowych rycin czy banknotów;

5. wyszukanie obiektu podobnego, kiedy to spośród wielu dostępnych przypadków należy znaleźć tylko te, które zawierają obiekt tożsamy z jedynym, dostępnym wzorcem, np. mając zdjęcie osoby chcemy ustalić jej tożsamość;

6. inne.

W pierwszym, klasycznym zastosowaniu rozpoznawania obiektów (wzorców), powstały na etapie segmentacji, symboliczny opis wstępny treści obrazu jest weryfikowany w dwóch kierunkach: samego występowania obiektu zainteresowania oraz bardziej szczegółowej identyfikacji rodzaju tego obiektu. Przykładowo, w specjalistycznym zastosowaniu medycznym polegającym na analizie mammogramów (tj. rentgenowskich obrazów sutka) segmentuje się wstępnie obszary jaśniejsze o zwartej strukturze, a następnie rozpoznaje się dwie klasy obiektów: guz lub obszar nieistotny. Rozpoznawanie może być także kontynuowane w kolejnym

etapie, kiedy to obiekty zidentyfikowane jako guz są weryfikowane jako zmiany złośliwe (patologia, która wymaga interwencji terapeutycznej) lub łagodne, z zaleceniem jedynie dalszej obserwacji i okresowej kontroli.

Rozpoznawanie bazujące na wcześniej wydzielonych regionach (potencjalnych lub realnych obiektach) zawiera zwykle dwa podstawowe etapy: określenie zestawu cech dobrze opisujących specyficzne właściwości analizowanych obiektów oraz klasyfikację. Optymalizowany zestaw cech ma zwiększyć skuteczność klasyfikacji. Służą temu takie etapy jak ekstrakcja cech, selekcja cech (np. określonej liczby najmniej skorelowanych właściwości obiektów), korekcja zestawu cech na bazie dostępnej wiedzy specjalistycznej (dziedzinowej) w celu wyznaczenia najsilniej różnicującej klasy obiektów przestrzeni cech.

Proces konstrukcji optymalnej przestrzeni cech w przypadku obiektów obrazowych nie podaje się prostym schematom, zwykle nie daje się opisać analitycznie czy innym, klasycznym formalizmem matematycznym. Często zaś czerpie z eksperymentu, intuicji dobrych heurystyk czy reguł statystycznych.

Cenne jest nierzadko wskazanie w modelu dziedziny ogólnych pojęć istotnych dla danego problemu interpretacji danych (np. obrazów), w drugim etapie są one uszczegóławiane aż do poziomu semantycznych cech dyskryminacyjnych. Etap końcowy to przypisanie cechom wizualnym matematycznych deskryptorów.

W zagadnieniu klasyfikacji zasadniczym i często decydującym o sukcesie zagadnieniem jest konstrukcja odpowiedniej przestrzeni deskryptorów matematycznych, charakteryzującej klasyfikowane obiekty. Proces ten powinien się rozpocząć od opisu obiektów za pomocą cech semantycznie istotnych i reprezentatywnych dla rozwiązywanego zadania.

Jeśli każdą cechę semantyczną opiszemy formułą obliczeniową lub algorytmem, to otrzymamy formalny opis obiektów w postaci zestawów N liczb, tzw. wektorów cech. Najczęściej z jedną cechą można związać kilka deskryptorów i nie wiadomo z góry, który z nich, czy też jaka ich kombinacja okaże się najbardziej efektywna w danym zadaniu klasyfikacji. Ponadto, przestrzeń cech rozpinana na bazie zestawu numerycznych deskryptorów opisuje często właściwości obiektów w sposób "nadmiarowy", za pomocą cech w pewnym stopniu skorelowanych.

Redukcja przestrzeni cech

Termin redukcja przestrzeni cech oznacza wybór ograniczonego, istotnego dla rozwiązania danego problemu analizy czy klasyfikacji danych, podzbioru początkowego zestawu cech w celu:

- poprawy efektywności klasyfikacji, przy czym efektywność ta (mierzona za pomocą takich parametrów jak czułość, specyficzność, trafność) zależy od kilku czynników:

  - wielkości i reprezentatywności zbioru uczącego, proporcji liczby przykładów do wymiaru przestrzeni cech,
  - jakości przestrzeni cech (liczby i efektywności deskryptorów cech obiektów) – włączenie do przestrzeni cech deskryptorów nieefektywnych obniża skuteczność klasyfikacji – efekt ten odgrywa szczególnie dużą rolę, gdy zbiór danych uczących jest niewielki w stosunku do liczby cech;
  - rodzaju klasyfikatora; klasyfikatory estymują swoje parametry na podstawie wektora cech określonych rozmiarów, dlatego przy ustalonym zbiorze uczącym wraz ze wzrostem liczby cech, wzrasta liczba parametrów do estymacji dla klasyfikatora, a wiec zwiększa się błąd estymacji – w konsekwencji, dla ustalonego zbioru uczącego zwiększa się błąd klasyfikacji;

- minimalizacji nakładów obliczeniowych związanych z ekstrakcją, klasyfikacją i przechowywaniem cech;

- poprawy przejrzystości procesu klasyfikacji, większego zrozumienia relacji między cechami obliczeniowymi, semantycznymi czy postrzegania rozpoznawanych obiektów, co sprzyja uogólnieniom otrzymywanych rezultatów czy też łatwości ich interpretacji;

- unikania problemów dużej wymiarowości analizowanych przestrzeni danych, związanych m.in. z rzadką reprezentacją danych w takich przestrzeniach, technikami przeszukiwania czy organizacji danych, statystyczną reprezentatywnością wydzielanych regionów itp.

Procedury redukcji wymiarowości przestrzeni cech stosowane są więc przed etapem klasyfikacji lub też iteracyjnie, naprzemiennie z klasyfikacją o dobieranej przestrzeni cech.

Redukcja przestrzeni cech obejmuje dwa zasadnicze podejścia:

- selekcja cech reprezentatywnych, inaczej dobór ograniczonego, istotnego dla rozwiązania danego problemu podzbioru początkowego zestawu cech, realizowana przede wszystkim za pomocą metod:
  - rankingowych, porządkujących cechy za pomocą ustalonej funkcji oceny (np. szacowanego współczynnika korelacji czy średniej informacji wzajemnej), liczonej na podstawie rozkładu wartości poszczególnych cech oraz docelowych przyporządkowań (klas wyjściowych) i odrzucającej cechy o najniżej ocenie końcowej; wybór cech odbywa się niezależnie od metody klasyfikacji na podstawie oceny ich zdolności przewidywania właściwej klasy opisywanych cechami obiektów;

  - przeszukiwania z kryterium dokładnej klasyfikacji (*wrapper method*), gdzie kolejne kombinacje możliwych podzbiorów zestawu cech są weryfikowane według przyjętej strategii, na podstawie miar dokładności wykonywanej klasyfikacji (wynik zależy więc od zastosowanego klasyfikatora); procedura jest powtarzana aż do spełnienia zdefiniowanego w metodzie kryterium stopu; główną wadą tego typu rozwiązania jest wysoki koszt obliczeniowy, zwłaszcza dla dużych zestawów cech;
  - filtracji – tutaj podzbiory cech wybierane są na etapie wstępnym, oceniane pojedynczo, a funkcja oceny bazuje jedynie na właściwościach danych, niezależnie od metody klasyfikacji, na podstawie ich związku z daną klasą; związek ten jest oceniany za pomocą różnych, często heurystycznych kryteriów (np. poziom średniej informacji wzajemnej) i wyrażany np. przy użyciu wag – wynikiem działania jednej z grup algorytmów (tzw. *rankers*) jest lista podzbiorów cech wraz z przypisanymi im wagami; prostą metodą wyboru optymalnego zbioru cech na podstawie wskazanych wag jest eliminacja cech na podstawie uzyskanej dokładności klasyfikacji; zaletą metod tej grupy jest szybkość działania nawet dla dużych zbiorów cech, przy czym metody te nie dają pewności co do optymalności rozwiązań;
  - metody wbudowane (*embedded*) – w tej grupie algorytmów selekcja cech jest wykonywana w fazie uczenia klasyfikatora i są zazwyczaj dopasowane do mechanizmów uczenia klasyfikatora;

- ekstrakcja cech poprzez transformację przestrzeni o wyższym wymiarze w przestrzeń o zredukowanej wymiarowości – następuje więc zmiana charakteru cech nowej przestrzeni; stosowane są w tym przypadku najczęściej: metoda analizy składowych głównych PCA (*Principal Component Analysis*), zbliżona w koncepcji transformacja Karhunen'a-Loeve'a KLT, liniowa analiza dyskryminacyjna LDA (*Linear Discriminant Analysis*), połączenie KLT+LDA, analiza składowych niezależnych ICA (*Independent Component Analysis*), a także nieliniowa wersja PCA – Kernel PCA czy lokalna PCA – LPCA; wykorzystywane mogą być także rozwinięcia w bazach ortogonalnych, falkowych, ogólnie wieloskalowych itp.

Klasyfikacja

Zasadniczym celem klasyfikacji jest przyporządkowanie analizowanych obiektów zainteresowania do określonej klasy (kategorii, grupy itp.) znaczeniowej. Skutkiem klasyfikacji jest rozpoznanie treści (na różnym poziomie abstrakcji) przekazu w zakresie zdefiniowanym przez mechanizm klasyfikacji. Klasy mogą mieć ścisłą definicję znaczeniową, jak np. "kot" czy "patologia", przybliżoną (np. "obcy obiekt" w zastosowaniach monitoringu, kiedy to celem jest jedynie wskazanie przedmiotu

dalszej analizy) lub też abstrakcyjną (np. wyszukujemy z rysunku wszystkie czworokąty jako obiekty o określonych cechach geometrycznych czy topologicznych, bez funkcji znaczeniowej, celem jedynie uporządkowania posiadanych zasobów).

Istnieje wiele metod klasyfikacji, od rozwiązujących bezbłędnie problem liniowo separowalnych cech obiektów, rozstrzyganie na podstawie danych *a priori* rozkładów prawdopodobieństwa rozdzielanych cech metodą największej wiarygodności, aż po metody określające przynależność obiektów w przypadku silnego, wzajemnego wymieszania rozdzielalnych znaczeniowo klas w numerycznej przestrzeni jedynie dostępnych cech. Całe to zagadnienie można uogólnić jako poszukiwanie optymalnej hiperpłaszczyzny czy hiperpowierzchni rozdzielającej klasy przynależności obiektów zainteresowania.

By jednak lepiej zrozumieć rolę metod klasyfikacji w różnego typu zastosowaniach *rozpoznających*, trzeba odnieść się do kategorii *systemów uczących się* na podstawie danych reprezentujących dany problem. Uczenie to rozumiane jest przede wszystkim jako odkrywanie albo poznawanie specyfiki informacji zawartej w dostępnych danych, aby na tej podstawie można było przeprowadzić różnego typu wnioskowanie w warunkach niepewności. Można więc mówić tutaj o systemach inteligentnej analizy danych, przy czym można je podzielić na uczące się bez nadzoru (czyli tylko na podstawie dostępnych danych, opisanych wektorami cech) oraz pod nadzorem (inaczej z nauczycielem lub – na przykładach). Nadzorowane uczenie wykorzystuje dodatkowy zbiór danych (atrybutów, zmiennych) objaśniających podstawowy wektor danych treningowych (objaśnianych – próba ucząca). W tym przypadku chodzi o okrycie zależności czy reguły wiążącej oba rodzaje danych, służącej wnioskowaniu.

W przypadku klasyfikacji obiekty $\mathcal{T}_i$ opisane za pomocą wektora cech $\mathbf{x}_i = \mathbf{x}(\mathcal{T}_i) \in \mathbb{R}^n$ przypisywane są do jednej z wcześniej ustalonych klas decyzyjnych. Danymi objaśniającymi jest w tym przypadku zbiór etykiet (pojedynczych atrybutów jakościowych) $\phi_i = \phi(\mathcal{T}_i) \in \{1, 2, \ldots, k\}$ określających przydział obiektów do jednej z k klas. Próba ucząca składa się więc z ciągu par $(\mathbf{x}_i, \phi_i)$ służących uczeniu – konstruowaniu klasyfikatora.

Docelowo klasyfikator powinien poprawnie wskazać (przewidzieć) przynależność obserwowanych obiektów do właściwej klasy na podstawie wektorów opisujących ich cech. Ważnym elementem systemu uczącego się jest więc sposób oceny końcowej skuteczności klasyfikacji, jak też szacowanie błędu klasyfikacji na kolejnych etapach procesu uczenia klasyfikatora. Zwykle wykorzystywane miary czułości, specyficzności czy trafności liczone są względem referencyjnego zbioru etykiet zbioru uczącego (na etapie uczenia) czy też zbioru testowego, składającego się z przypadków odmiennych, które nie należą do zbioru uczącego (na etapie weryfikacji klasyfikatora). Problem klasyfikacji pod nadzorem jest często rozwiązywany metodami analizy dyskryminacyjnej. Innym przykładem uczenia się pod nadzorem o charakterze ilościowym jest metoda regresji liniowej, kiedy to na pod-

stawie danych treningowych metoda uczy się wartości parametrów estymowanej funkcji regresji.

Uczenie bez nadzoru bazuje jedynie na zbiorze przykładowych obiektów opisanych jedynie wektorami cech $\mathbf{x}_i$ nieetykietowanych. System uczący się służy uporządkowaniu i opisaniu obserwowanych danych, wykrycia struktury danych, wzajemnych zależności, regularności, inaczej – do ich objaśnienia. Klasyfikacja bez nadzoru, nazywana też analizą skupień, klasteryzacją czy grupowaniem, sprowadza się do wyznaczenia zwykle rozłącznych klastrów czy skupień danych, gromadzących w jednej grupie dane o zbliżonych cechach. Niekiedy chodzi o wskazanie obiektów nietypowych, o cechach odbiegających od pozostałych. Ważną rolę w tym przypadku odkrywa kryterium grupowania, sposób inicjalizacji czy też określania aktualnej przynależności danych do skupisk w iterowanym procesie uczenia.

Projektowanie klasyfikatora pod nadzorem może niekiedy obejmować wstępną fazę grupowania, czy rozdzielenia zbioru danych uczących (treningowych) na rozłączne podzbiory przynależności do określonej liczby klas metodą uczenia bez nadzoru. W metodzie grupowania k średnich następuje grupowanie na zasadzie iteracyjnego przeplatania metod najbliższego sąsiada (przypisanie do reprezentacyjnego przedstawiciela najbliższej z k klas według przyjętej reguły odległościowej) oraz centroidu (środek ciężkości obiektów aktualnie przypisanych danej klasie staje się nowym reprezentantem klasy). Ustalone według tej procedury, przy określonym kryterium stopu, minimalno-odległościowe granice przedziałów (regionów) poszczególnych klas mogą stanowić definicje klas użytych do klasyfikacji obiektów testowych (rozpoznawanych) metodą najbliższego reprezentanta klas.

Modyfikacja tej metody, znana jako metoda grupowania c średnich, wykorzystuje rozmyty model przynależności obiektu do poszczególnych klas (prawdopodobna jest przynależność obiektu do każdej z możliwych klas). Procedura grupowania polega na iteracyjnej modyfikacji prawdopodobieństw przynależności obiektów do klas na bazie danych treningowych, a ostateczne przypisanie obiektów do klasy odbywa się na zasadzie największego prawdopodobieństwa (wiarygodności).

Zasadniczo wśród metod klasyfikacji pod nadzorem można wyróżnić metody należące do kategorii:

- klasyfikatorów bazujących na probabilistycznej teorii decyzji Bayesa, przy zastosowaniu znanych *a priori* rozkładów prawdopodobieństw (np. normalnych), ale też przy nieznanej postaci rozkładów (naiwny klasyfikator Bayesa, estymacja parametrów metodą największej wiarygodności, estymacja maksymalnej entropii czy mieszanina rozkładów);
- klasyfikatorów liniowych (zasada perceptronu, liniowy dyskryminator Fishera, metoda najmniejszych kwadratów czy metoda wektorów nośnych i

inne); ich zaletą jest prostota koncepcyjna i obliczeniowa, często efektem projektowanych klasyfikatorów są liniowe funkcje dyskryminacyjne;

- klasyfikatorów nieliniowych, będącym zazwyczaj uogólnieniem systemów liniowych (wielowarstwowy perceptron, sieci neuronowe, klasyfikatory wielomianowe, metoda wektorów nośnych z funkcjami jąder, drzewa decyzyjne, klasyfikatory łączone i inne); uogólniające rozszerzenie dotyczy też zwiększenia liczby klas w stosunku do podstawowej metody binarnej.

Liniowe funkcje dyskryminacyjne. Podstawowe zagadnienie klasyfikacji można sprawdzić do problemu przypisania obiektów opisanych wektorami cech (n-wymiarowym) do jednej z dwóch możliwych klas (binarny problem klasyfikacji) za pomocą liniowych funkcji dyskryminacyjnych. Przyjmijmy dla uproszczenia, że wartości wektorów cech obiektów obu klas są liniowo separowalne, czyli rozwiązanie za pomocą liniowych funkcji dyskryminacyjnych pozwoli skutecznie rozwiązać ten problem. Zaprojektowanie klasyfikatora sprawdza się w tym przypadku do wyznaczenia hiperpłaszczyzny decyzyjnej, rozdzielającej wektory obu klas, postaci

$$h(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = 0 \tag{3.88}$$

z wektorem wag $\mathbf{w} = [w_1, w_2, \ldots, w_n]$ określającym kierunek oraz progiem w_0 precyzującym lokalizację hiperpłaszczyzny. $\mathbf{w}$ i w_0 należy ustalić w procesie uczenia klasyfikatora. Sposoby wyznaczania $\mathbf{w}$ różnicują metody klasyfikacji. Warto zauważyć, że $\mathbf{w}^T$ rzutuje $\mathbf{x}$ na hiperpłaszczyznę, a $|h(\mathbf{x})|$ jest miarą euklidesowej odległości wektora cech od płaszczyzny decyzyjnej. Dla skutecznie wyznaczonej hiperpłaszczyzny z jednej strony mamy dodatnie wartości $h(\mathbf{x})$ – dla wektorów cech $\mathbf{x}$ obiektów jednej klasy, przyjmijmy Φ_1), zaś z drugiej – $h(\mathbf{x}) < 0$ dla obiektów drugiej klasy Φ_2). Jeśli założymy, że wektor wag dla skutecznie wyznaczonej hiperpłaszczyzny wynosi $\tilde{\mathbf{w}}$, wówczas ogólnie można zapisać (z dokładnością do zawsze możliwej korekty n-wymiarowej przestrzeni cech)

$$\begin{aligned} \tilde{\mathbf{w}}^T\mathbf{x} &> 0 \quad \forall \mathbf{x} \in \Phi_1 \\ \tilde{\mathbf{w}}^T\mathbf{x} &< 0 \quad \forall \mathbf{x} \in \Phi_2 \end{aligned} \tag{3.89}$$

Wyznaczenie skutecznego $\tilde{\mathbf{w}}$ jest klasycznym problemem optymalizacyjnym, a sposób poszukiwania rozwiązań wymaga zdefiniowania funkcji kosztu oraz algorytmu zmierzającego do minimalizacji wartości kosztu możliwych rozwiązań. I tak w przypadku algorytmów perceptronowych minimalizowana funkcja kosztu ma postać

$$\mathcal{L}_{\mathrm{P}}(\mathbf{w}) = \sum_{\mathbf{x} \in B} \left(\delta_x \mathbf{w}^T\mathbf{x}\right) \tag{3.90}$$

gdzie B jest zbiorem źle klasyfikowanych, za pomocą aktualnego $\mathbf{w}$, wektorów treningowych. Generalnie na podstawie (3.90) wyznaczamy wektor wag jako

$$\tilde{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}) \tag{3.91}$$

Zmienne $\delta_x = -1$ dla $\mathbf{x} \in \Phi_1$ (został zaliczony do Φ_2) oraz $\delta_x = 1$ dla $\mathbf{x} \in \Phi_2$ (odwrotnie). Łagodna, iteracyjna zmiana wartości wektora wag odbywa się według zależności $\mathbf{w}(i+1) = \mathbf{w}(i) - \kappa_i \sum_{\mathbf{x}\in B}(\delta_x \mathbf{x})$, startując z losowo dobranej wartości $\mathbf{w}(0)$ oraz odpowiednio korygowanej w kolejnych krokach wartości współczynnika κ_i (szybkość vs. dokładność przeszukiwań). Rozwiązanie nie jest jednoznaczne, gdyż hiperpłaszczyzn skutecznie rozdzielających liniowo separowane wartości wektorów cech obiektów należących do dwóch klas może być wiele (zobacz rys. 3.34, po lewej). Istnieje wiele modyfikacji algorytmu perceptronu (zobacz np. [196], rozdział 3.3).

W metodzie średniokwadratowej (*mean squares*) funkcja kosztu ma postać

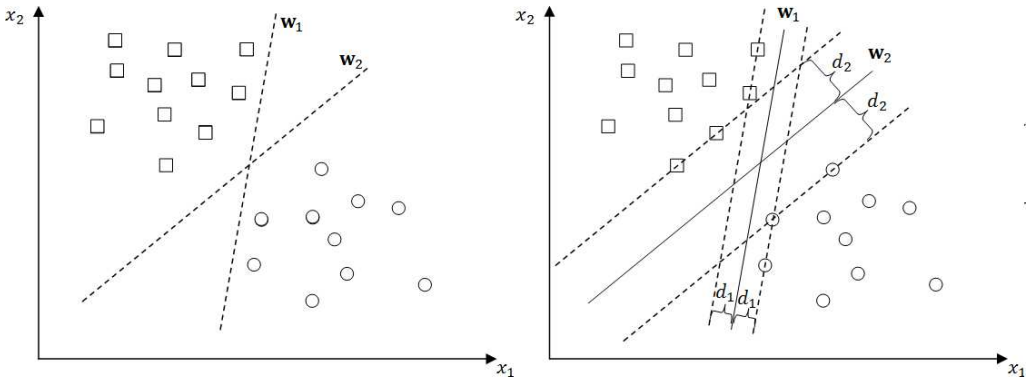
$$\mathcal{L}_{\mathrm{MS}}(\mathbf{w}) = E\left[|\phi(\mathbf{x}) - \mathbf{x}^T\mathbf{w}|^2\right] \tag{3.92}$$

gdzie pożądany efekt wskazań klasyfikatora $\phi(\mathbf{x}) = 1$ dla $\mathbf{x} \in \Phi_1$ oraz $\phi(\mathbf{x}) = -1$ dla $\mathbf{x} \in \Phi_2$. Policzenie pochodnych cząstkowych $\mathcal{L}(\mathbf{w})$ po wektorze wag i przyrównanie do zera pozwala wyznaczyć optymalny $\tilde{\mathbf{w}} = E\left[\mathbf{x}\mathbf{x}^T\right]^{-1} E[\mathbf{x}\phi(\mathbf{x})]$ z macierzą autokorelacji $E\left[\mathbf{x}\mathbf{x}^T\right]$ oraz kros-korelacji $E[\mathbf{x}\phi(\mathbf{x})]$ wektorów danych treningowych oraz spodziewanych rezultatów ich klasyfikacji. Liczenie tych macierzy wymaga wstępnej znajomości rozkładów prawdopodobieństwa odpowiednich zmiennych – uproszczeniem jest metoda najmniejszych kwadratów (*least squares*) z funkcją kosztu

$$\mathcal{L}_{\mathrm{LS}}(\mathbf{w}) = \sum_{i=1}^{t} \left(\phi_i(\mathbf{x}) - \mathbf{x}_i^T\mathbf{w}\right)^2 \tag{3.93}$$

gdzie $\phi_i(\mathbf{x}) \in \{-1, 1\}$ w przypadku dwóch klas, t – liczba wektorów treningowych. Rozwiązanie sprowadza się wtedy do rozwiązania układu równań normalnych z n niewiadomymi i macierzą korelacji próbek przybliżającą macierz autokorelacji.

Metoda wektorów nośnych. Opisane wyżej metody, wykorzystujące różne funkcje kosztu pozwalają wykreślać różne hiperpłaszczyzny dając rozwiązania problemu liniowego na zbiorze uczącym. Jednak ze względu na spodziewaną wysoką skuteczność nauczonego klasyfikatora na zbiorze zróżnicowanych przypadków testowych, należałoby dobrać taką jej postać, która zapewni maksymalny margines rozdzielający obie klasy przypadków. Chodzi o to, by margines błędu był możliwie duży na okoliczność klasyfikacji nieznanych przypadków. Dlatego jako bardziej użyteczną wybieramy hiperpłaszczyznę definiowaną przez kierunek $\mathbf{w}_2$ z rys. 3.34, ze względu na większą wartość symetrycznego (nie ma powodów, żeby uprzywilejowywać jedną z klas), obustronnego marginesu d_2.



Rysunek 3.34: Przykład liniowej separacji obiektów dwóch klas w przestrzeni cech (x_1, x_2) za pomocą różnych przebiegów prostej separującej (po lewej), przy czym prosta $\mathbf{w}_2$ zapewnia większy margines "bezpieczeństwa" $d_2 > d_1$, czyli jest bardziej odległa od najbliższych obiektów każdej z klas (po prawej).

Doprecyzowany problem optymalizacji sprowadza się więc do wyznaczenia takiej hiperpłaszczyzny separującej liniowo przypadki dwóch klas, która zapewni maksymalny margines w odniesieniu do przypadków uczących. Rozwiązanie tego zagadnienia, pozwalające jednoznacznie określić najbardziej użyteczną postać hiperpłaszczyzny zaproponowano w postaci metody (maszyny) wektorów nośnych – SVM (*Support Vector Machine*).

SVM jest przykładem klasyfikatora, który znajduje szerokie zastosowanie w aplikacjach multimedialnych – świadczą o tym aplikacje opisane w [199, 200] i wiele innych. Koncepcja metody wektorów nośnych (podpierających, wspierających, podtrzymujących), obejmująca w pierwszym okresie rozwoju rozważania teoretyczne, konstrukcję określonego modelu matematycznego, a następnie rozwijane z czasem realizacje algorytmiczne i atrakcyjne użytkowo implementacje kształtowała się na przestrzeni kilkudziesięciu lat (pierwsze istotne prace pochodzą z lat 60. zeszłego stulecia, kolejne to przede wszystkim [26, 194, 195]). Dziś metoda ta, badana eksperymentalnie w rosnącej skali zastosowań potwierdza swoją użyteczność [196, 197, 198].

Podstawą metody jest koncepcja przestrzeni decyzyjnej, która ulega podziałowi poprzez ustalenie granicy separującej obiekty o różnej przynależności klasowej. Na etapie uczenia klasyfikatora wyznaczana (rozpinana) jest hiperpłaszczyzna określonej przestrzeni cech, rozdzielająca z maksymalnym marginesem błędu przypadki uczące.

Wracając do rozważań bardziej formalnych, symetryczny margines określający "bezpieczną" strefę rozdzielającą obie klasy, definiowany jest jako odległość od hiperpłaszczyzny najbliższego punktu danej klasy tak że

$$d = \min_{\mathbf{x}\in\Phi_1} \frac{|h(\mathbf{x})|}{\|\mathbf{w}\|} = \min_{\mathbf{x}\in\Phi_2} \frac{|h(\mathbf{x})|}{\|\mathbf{w}\|} \tag{3.94}$$

Aby uprościć obliczenia znormalizujmy hiperpłaszczyznę, czyli przeskalujmy wartości wektora wag i progu tak, by wartości bezwzględne $h(\mathbf{x})$ w najbliższych hiperpłaszczyźnie punktach każdej z klas wynosiły 1. Wtedy obustronna szerokość marginesu, przy skalowanych wagach, wynosi $\frac{2}{\|\mathbf{w}\|}$. Minimalizowana funkcja kosztu konstruowana jest w taki sposób, by zmaksymalizować margines separacji przy wykorzystaniu normy kwadratowej, czyli

$$\mathcal{L}_{\mathrm{SVM}}(\mathbf{w}, w_0) = \frac{1}{2}\|\mathbf{w}\|^2 \tag{3.95}$$

przy ograniczeniu

$$\phi_i(h(\mathbf{x}_i)) \geqslant 1, \quad i = 1, \ldots, t \tag{3.96}$$

czyli że odległość każdego z obiektów winna być $\geqslant \frac{1}{\|\mathbf{w}\|}$ oraz że wartości $h(\mathbf{x})$ są dodatnie dla przypadków klasy 1 i ujemne dla klasy 2. Przy rozwiązywaniu (3.95) ograniczenie to przyjmuje postać

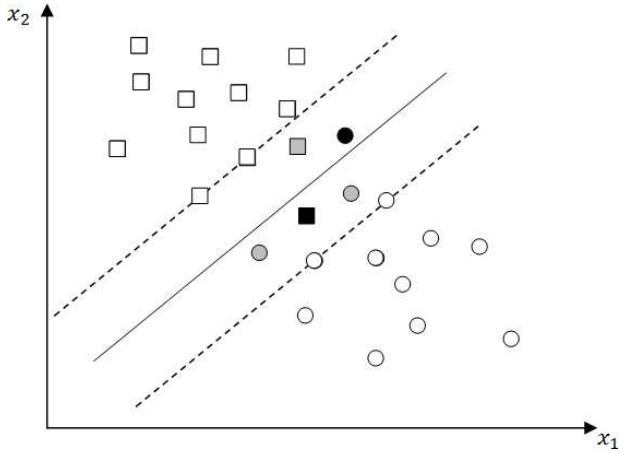
$$\lambda_i[\phi_i(h(\mathbf{x}_i)) - 1] = 0 \tag{3.97}$$

dla wartości mnożników Lagrange'a $\lambda_i \geqslant 0$.

Rozwiązaniem jest wektor wag $\tilde{\mathbf{w}} = \sum_{i=1}^{t} \lambda_i \phi_i \mathbf{x}_i$ wobec $\sum_{i=1}^{t} \lambda_i \phi_i$. Wartość w_0 może być wyznaczona na podstawie dowolnego z warunków (3.97) z niezerową λ_i (częściej – jako średnia po wszystkich takich wektorach).

Optymalna w sensie (3.95) hiperpłaszczyzna jest więc liniową kombinacją wektorów treningowych z niezerową wartością λ_i, czyli wektorów aktywnych przy ustalaniu $\tilde{\mathbf{w}}$. Okazuje się, że dla $\lambda_i \neq 0$ warunek (3.97) jest spełniony jedynie przy $h(\mathbf{x}_i) = \pm 1$, czyli dla wektorów uczących leżących na dwóch brzegowych hiperpłaszczyznach stanowiących granicę strefy "bezpieczeństwa" (po obu stronach wyznaczanej płaszczyzny decyzji w odległości marginesu – linie przerywane wokół prostej separującej klasy na rys. 3.34, z prawej). Te wektory treningowe noszą nazwę **wektorów nośnych**. Inne wektory, dla których $\lambda_i = 0$ leżą na zewnątrz tej strefy lub niekiedy mogą także leżeć na jej granicy (czyli na wspomnianych hiperpłaszczynach brzegowych [196]).

Generalnie metoda SVM rozwiązuje binarny (tj. dotyczący rozdzielenia dwóch klas) problem klasyfikacji. Jak pokazano, w najprostszym przypadku liniowo separowalnym wyznaczenie hiperpłaszczyzny separującej jest zadaniem optymalizacji kwadratowej. Określa się ją w iteracyjnym algorytmie uczącym, który minimalizuje dobraną funkcję kosztu. Metoda ta jest jednak nieskuteczna w przypadku, kiedy klasy nie są liniowo separowalne, tj. kiedy nie istnieje liniowa reguła klasyfikacji pozwalające bezbłędnie przyporządkować etykietę klas wszystkim danych treningowym (zobacz rys. 3.35). Uelastycznienie metody wymaga wówczas przedefiniowania funkcji kosztu oraz reguł ograniczających, które są skutkiem modyfikacji zasadniczej koncepcji klasyfikatora SVM, która dopuszcza występowanie przypadków wewnątrz marginesu separacji, a nawet błędne decyzji klasyfikacji. W



Rysunek 3.35: Przykład liniowo nieseparowalnego rozkładu wartości cech obiektów dwóch w przestrzeni decyzyjnej, kiedy to wewnątrz marginesu separacji pojawiają się przypadki treningowe (punkty szare i czarne), przy czym niektóre z nich są błędnie klasyfikowane za pomocą ustalonej hiperpłaszczyzny decyzyjnej (punkty czarne).

nowej sytuacji celem uczenia jest maksymalizacja marginesu separacji klas przy jednoczesnej minimalizacji dopuszczalnych przekroczeń granic tego marginesu.

Dopuszczalna możliwość popełnienia błędu przy klasyfikacji wyraża się w osłabieniu ograniczenia (3.96) do postaci

$$\phi_i(h(\mathbf{x}_i)) \geqslant 1 - \xi_i \tag{3.98}$$

gdzie $\xi_i \geqslant 0$ są wartościami nowej zmiennej "rozluźniającej" (*slack*) ξ. Wynika z tego, że obserwacja zostanie błędnie sklasyfikowana, jeśli dla danego przypadku wartość $\xi_i > 1$. Na etapie uczenia klasyfikatora minimalizowana jest wtedy zmodyfikowana w stosunku do (3.95) postać funkcji kosztu:

$$\mathcal{L}_{\mathrm{SVM}}(\mathbf{w}, w_0, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{t}\xi_i \tag{3.99}$$

gdzie stała C jest stałą - współczynnikiem kary. Celem optymalizacji jest ustalenie hiperpłaszczyzny na bazie właściwie dobranych wektorów nośnych, analogicznie jak w problemie liniowo separowalnym, dla której margines separacji będzie maksymalny, przy jednoczesnej minimalizacji liczby przekroczeń marginesu, tj. liczby przypadków, dla których $\xi_i > 0$.

Problem braku liniowej separowalności klas można zwykle skuteczniej rozwiązać stosując klasyfikatory nieliniowe. Najczęściej stosuje się wówczas nieliniowe przekształcenie – funkcję mapującą wektory cech obiektów w nową przestrzeń, zwiększając nierzadko jej wymiarowość, gdzie można skuteczniej zastosować metody liniowe i wyznaczyć separowalną hiperpłaszczyznę. Typowe postacie funkcji

mapujących (inaczej funkcji jądra) to funkcje wielomianowe, normalne, sigmoidalne, radialne RBF, tangensa hiperbolicznego nawet falkowe.

Klasyfikację według koncepcji SVM można również rozszerzyć na przypadek większej liczby klas jako zestaw problemów binarnych w konwencji jeden przeciw wszystkim (separacji przypadków danej klasy spośród wszystkich innych) lub też w konwencji jeden na jeden (na podstawie wyznaczonego w ten sposób zestawu hiperpłaszczyzn separujących ustala się ostateczne rozwiązanie metodą głosowania).

Rozumienie obrazów

Definicja procesu rozumienia danych nie jest prosta, bo odwołując się do zróżnicowanych ludzkich intuicji trudno jest precyzyjnie określić jej ramy w przełożeniu na formalne opisy, algorytmy, komputerowe metody analityczne. Ogólnie rozumieć przekaz informacji to "łapać, o co chodzi", pojmować sens, zdawać sobie sprawę z wymowy obserwowanej treści. Rozumienie to wyłowienie z obrazów informacji niezbędnej do ich użytkowania, czyli interpretacji odczytanego przekazu, a często też podjęcia na jego podstawie wiążących decyzji.

J. Hawkins [192] określił rozumienie sytuacji jako zdawania sobie sprawy z życiowo istotnych skutków danej sytuacji. Z kolei R. Penrose zwraca szczególną uwagę na rolę świadomości w procesie rozumienia danej sytuacji [193].

Istotne jest podkreślenie konieczności semantycznego rozumienia informacji [15]. Informacja służy odbiorcy w realizacji określonego celu, a przekaz danych dokonuje się zawsze w kontekście określonej treści. Treści danych przypisana jest funkcja semantyczna, która odzwierciedla w dużych stopniu ich użyteczność dla odbiorcy. Precyzyjnie określając semantykę rozpoznanych elementów, obiektów czy cech w przypadkach niejasnych – z informacją ukrytą i komunikując ją odbiorcy, dostarczamy informacji użytecznych w określonym zastosowaniu. Dobrem przykładem jest rozumienie informacji obrazowej.

Komputerowe rozumienie obrazów wspiera proces właściwej interpretacji semantycznej reprezentacji obrazów, tj. określonych (ustalonych, rozpoznanych) obiektów/regionów/cech – komponentów o przypisanej funkcji semantycznej. Interpretacja wyjaśnia "co rzeczywiście dzieje się" w obrazie, jaka jest wymowa informacji obrazowej na wyższym, określonym przez zastosowanie, poziomie abstrakcji. Prowadzi to do formułowania decyzji dotyczących dalszych działań, czyli robienia użytku z wyekstrahowanej, rozpoznanej, zrozumianej i właściwie ocenionej informacji z ustaloną semantyką.

Rozumienie obrazów to określenie znaczenia/wymowy pełnej informacji przekazywanej w obrazie (znaczenia całej sceny). Wykorzystywana jest tutaj wiedza dziedzinowa z zakresu relacji pomiędzy obiektami i ich cechami, okoliczności (kontekstu) ich występowania, dostrzegalne zarówno w samym obrazie, jak też wynikające z innych uwarunkowań procesu obrazowania.

Kluczowym warunkiem skutecznej interpretacji obrazów jest trafne rozpoznanie – odczytanie pełnej informacji poprzez **zrozumienie całego przekazu treści obrazowej**, w tym niekiedy nawet najdrobniejszych, słabo dostrzegalnych jego szczegółów. Przy skutecznym opracowaniu narzędzi wspomagania na plan pierwszy wysuwa się problem integracji cech wizualnych oraz obliczeniowych, opisujących regiony zainteresowań w kontekście wiarygodnych znaczeń rozpoznanych obiektów, struktur, drobnych detali, kompozycji treści przez synergii tych znaczeń do postaci odczytywanej informacji.

Komputerowe rozumienie obrazów jest rodzajem sprzęgu pomiędzy komputerem a człowiekiem, – rodzajem dostrojenia stosowanego modelu formalnego do fachowego, ale także psychofizycznego potencjału użytkownika, – symbolem dopasowania specjalisty do wyrafinowanego narzędzia i odwrotnie. Udane rozumienie obrazów stwarza możliwość wykorzystania całego potencjału dostępnej wiedzy i zdolności, wspierając kreatywność i umiejętność trafnego wnioskowania. Sukces zależy tutaj od obu stron, przy czym zazwyczaj decydujący wpływ ma postawa człowieka.

Tradycyjne pytanie stawiane w metodologii metod analizy obrazów: "jak policzyć to, co widać?", w przypadku współczesnych wymagań komputerowego wspomagania coraz częściej przyjmuje formę odwrotną: **"jak rozumieć i pokazać to, co da się policzyć?"**, czyli jak zrozumieć to, co pojawia się w warstwie numerycznego opisu danych. A często pojawiają się tam rzeczy na tyle istotne, że wpływają w znaczącym stopniu na informacyjny przekaz obrazu. Nowych metod należy poszukiwać na styku postrzegania i interpretacji, objętych regułami reprezentowania informacji i rozumienia reprezentacji.

Automatyczna interpretacja obrazów dotyczy wyjaśnienia przyczyn obserwowanych czy ekstrahowanych treści, nawiązuje więc do szerszej wiedzy doświadczalnej, specyfiki zastosowań, czy ogólnej wiedzy typowego użytkownika.

3.2 Grafika komputerowa

O grafice komputerowej można mówić w sposób różnorodny – że są to algorytmy i struktury danych specjalizowane obrazem lub że jest to obrazkowy (graficzny) język interakcji człowiek-komputer.

Jako dział informatyki grafika komputerowa wykorzystuje techniki komputerowe do syntezy obrazów (lub ich fragmentów) w procesie wizualizacji rzeczywistości w różnych celach użytkowych. W zastosowaniach grafiki zwykle chodzi o rzeczywistość mniej lub więcej realną (wirtualną, rozszerzoną, sztuczną, wizję artystyczną czy futurologiczną), opisywaną, kształtowaną czy przekształcaną w sposób ściśle zależny od zastosowania. Przykładowo, w przypadku wizualizacji danych rzeczywistych (naturalnych, pomierzonych, rejestrowanych w złożonych systemach obrazowania, np. w medycynie) może chodzić o możliwie wierne ukazanie rejestrowanej treści obrazowej jedynie w wybranych zakresach lub też znaczące uproszczenie przekazu obrazowego poprzez adaptację założonych *a prori* modeli treści użytkowej.

Synteza obrazów bazuje na modelowaniu scen, przy czym jest to zazwyczaj specyficzny rodzaj modelowania, dostosowany do przedmiotu i celów wizualizacji.

Z czasem skala zadań obejmowanych przez grafikę ulegała znacznemu rozszerzeniu lub ukonkretnieniu. Obecnie obejmuje przede wszystkim:

- różne formy wizualizacji danych czy doboru dogodnych uwarunkowań przekazu informacji obrazowej,
- animacje, film, ogólniej kształtowanie strumieni wizyjnych w multimediach,
- interfejsy użytkownika (środowisko graficzne),
- CAD/CAM rożnego typu (projektowanie inżynierskie, poligrafia, architektura etc.),
- gry w różnych konwencjach,
- symulatory, generatory przestrzeni wirtualnych, rozszerzonych (*augmented*),
- e-edukacja,
- medyczna diagnostyka obrazowa (wizualizacji badań tomografii, wirtualna endoskopia, obrazowanie multimodalne), chirurgia sterowana obrazem, protetyka, rehabilitacja, medycyna w internecie,
- ...

Istnieje szereg wygodnych narzędzi i środowisk obiektowego czy funkcjonalnego programowania, gwarantujących szybkie i efektowne realizacji prostych aplikacji graficznych. Warto tutaj wymienić przede wszystkim:

- Visualization Toolkit (VTK) http://www.vtk.org/; otwarty system do komputerowej grafiki 3W oraz przetwarzania i wizualizacji obrazów; zbudowany na bazie C++ z interfejsami w Tcl/Tk, Java, Python; dostosowany do wielu platform systemowych;

- Insight Segmentation and Registration Toolkit (ITK) http://www.itk.org/; otwarty system konstruowany na zasadach podobnych do VTK, dotyczących zaawansowanych, przyszłościowych metod analizy obrazów;

- Medical Image Processing and Visualization (MeVisLab) http://www.mevislab.de/; środowisko do tworzenia aplikacji w obszarze przetwarzania i wizualizacji obrazów medycznych; na bazie Open Inventor (obiekty grafiki na bazie OpenGL), Qt, Python, Javascript (interfejsy), integracja z modułami C++, VTK, ITK.

-

Szczególnie ważnych elementem zaawansowanych systemów grafiki jest sprzęt, m.in. specjalizowane procesory i akceleratory graficzne, z przetwarzaniem masowo-równoległem (GPU), różnego typu manipulatory, tablety, doskonalone, integrowane i zestawiane systemy monitorów etc.

3.3 Metody kompresji

Kompresję można sprowadzić do odwracalnego lub nieodwracalnego procesu redukcji długości reprezentacji danych. W opracowaniu nowoczesnych narzędzi kompresji danych, szczególnie obrazów, dużego znaczenia nabiera szersze rozumienie pojęcia kompresji jako wyznaczania efektywnej reprezentacji przesyłanej lub gromadzonej informacji. Wiarygodne definicje informacji, modele źródeł informacji oraz sposoby obliczania ilości informacji i optymalizacji kodów stają się tutaj zagadnieniem kluczowym.

3.3.1 Krótka charakterystyka współczesnych uwarunkowań

Jesteśmy świadkami "ery informacji" o technologicznych podstawach, którą charakteryzuje intensywny rozwój społeczeństwa informacyjnego, przekształcającego się stopniowo w społeczeństwo szeroko dostępnej (tzw. wolnej) wiedzy. Wobec rosnącej roli Internetu można mówić również o zjawisku tworzenia się społeczeństwa sieciowego (prace Castellsa [68]). Coraz większe znaczenie rozwoju technologicznego, zdominowanego przez zagadnienie efektywnej wymiany informacji oraz świata nauk przyrodniczych i ścisłych widać chociażby w propozycjach tzw. trzeciej kultury [69].

Szybki dostęp do istotnych informacji, efektywne jej przetworzenie oraz wyszukiwanie wiarygodnych źródeł wiedzy staje się decydującym o sukcesie składnikiem nowoczesnej działalności edukacyjnej, biznesowo-gospodarczej, organizacyjnej, społecznej. Wobec nadmiaru pseudoinformacji (tysiące reklam, setki bezużytecznych zachęt zapychających skrzynki pocztowe i kanały telewizyjne) coraz bardziej pożądane stają się narzędzia inteligentnie porządkujące przestrzeń informacyjną.

Sposobem na życie współczesnego człowieka będącego w ciągłym pośpiechu staje się oszczędzanie (wręcz wydzieranie) każdej wolnej chwili [70]. B. Pascal (1623-1662) pisząc list do przyjaciela wspomniał, że "nie ma czasu, aby napisać krócej". Pisał więc jak leci, przepraszał, że zabrakło mu czasu, by zaoszczędzić czas adresata. Poszanowanie czasu przekłada się na wzrost ilości informacji dostępnej (przekazywanej) w jednostce czasu w warunkach, jakimi dysponuje konkretny odbiorca (określone ramy czasowe, sprzętowe, finansowe). Oszczędność czasu odbiorcy informacji zapewnia przede wszystkim odpowiednia jakość informacji, tj. selekcja danych, które są istotne dla użytkownika, oraz efektywna jej reprezentacja (minimalizacja długości zapisu danych i czasu ich przekazywania, właściwa struktura informacji).

Technologiczne doskonalenie systemów rozpowszechniania informacji wymaga więc stosowania efektywnych metod kompresji danych, w tym koderów obrazów. Obecnie niebagatelną rolę odgrywa przekazywanie informacji za pomocą obrazu - przykładem niech będzie chociażby dominacja obrazkowych systemów operacyjnych typu Windows. Kompresja obrazów pozwala szybciej zobaczyć obraz o lepszej jakości, więcej informacji umieścić w danej objętości nośnika, szybciej wyszukać żądaną informację, porównać obszerne zasoby informacji obrazowej etc. W koderach obrazów wykorzystuje się obok elementów teorii informacji także zaawansowane metody numeryczne, teorię sygnałów, analizę funkcjonalną, psychowizualne modele percepcji czy inną wiedzę o człowieku i jego relacji do świata (zależnie od zastosowań).

Doskonalenie metod akwizycji obrazów, poprawa zdolności rozdzielczej systemów obrazowania, redukcja szumu i zwiększanie dynamiki sygnału użytecznego wymusza operowanie coraz większymi zbiorami danych obrazowych. Szybko rosnąca liczba obrazów cyfrowych przenoszących informację w coraz większej gamie zastosowań, takich jak fotografia cyfrowa, kamery cyfrowe na użytek domowy, obrazy satelitarne wykorzystywane w meteorologii, kartografii czy urbanistyce, systemy obrazowania biomedycznego itd., wymaga użycia efektywnych metod gromadzenia, indeksowania, przeglądania i wymiany tej informacji. W opinii wielu odbiorców i nadawców informacji obrazowej wzrost efektywności i użyteczności opracowywanych narzędzi kompresji obrazów jest zbyt wolny wobec wymagać współczesnych zastosowań. Przykładem może być rozwijany obecnie standard JPEG2000, w którym spełnienie żądań nowoczesnego użytkownika okupiono zbyt-

nią złożonością obliczeniową algorytmów i trudnościami w praktycznej realizacji założeń standardu.

3.3.2 Krótka historia rozwoju

Pierwowzory współczesnych metod kodowania można dostrzec już w podejmowanych od wieków próbach kształtowania różnych form oszczędnego opisu bogatych w formie treści (pojęć). W każdej społeczności podejmowano próby przekazania informacji, ostrzegając np. przed zagrożeniem za pomocą sygnałów dymnych czy dźwiękowych. Aby wyrazić emocje, utrwalić doświadczenia posługiwano się malowidłami, rysowano znaki na ciele, nacinano skórę. Wymyślono języki, by łatwiej się porozumieć. Słowom przypisano określone znaczenie, intensywnie gestykulowano, powstały różne formy mowy niewerbalnej, itp.

Dopiero jednak datowane na przełom lat czterdziestych i pięćdziesiątych prace Claude Shannona [21], jednego z największych naukowców XX wieku według m.in. Sloane'a [72], zawierały sformalizowaną podstawę służącą opracowaniu metod kompresji, dając początek intelektualnej dyscyplinie kompresji. Stanowią one swego rodzaju akt założycielski w rozwoju współczesnych metod kompresji.

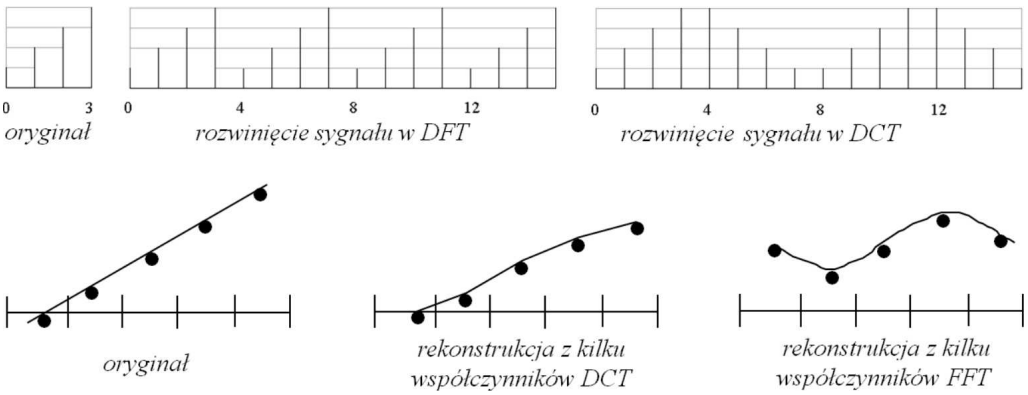
Sformułowane przez Shannona podstawy teorii informacji to kanał transmisji, probabilistyczne rozumienie informacji (zmienna losowa i łańcuch losowy opisujące źródło z pełną informacją statystyczną), entropia jako miara informacji, przyczyny nadmiarowości (redundancji), modele źródeł informacji, teoria zniekształceń źródeł informacji itd. Shannon przyczynił się także do powstania efektywnego algorytmu kodowania wykorzystującego analizę statystyczną sekwencji danych kompresowanych, znanego obecnie jako metoda Shannona-Fano [71].

Kolejnym istotnym wydarzeniem było opracowanie przez D.A. Huffmana optymalnej metody kodowania ciągu danych poprzez przyporządkowanie każdemu symbolowi alfabetu źródła modelującego dane wejściowe oddzielnego słowa kodowego o zmiennej długości (opublikowanie w 1952r.). Długość słów (w bitach) jest w przybliżeniu odwrotnie proporcjonalna do prawdopodobieństwa wystąpienia danego symbolu na wejściu kodera. Metoda jest wykorzystywana przez dziesiątki lat w różnych wersjach i odmianach jako uzupełnienie wstępnej dekompozycji oraz modelowania pośredniej reprezentacji danych (jak np. w standardzie JPEG).

W latach sześćdziesiątych i siedemdziesiątych opracowano wiele algorytmów kompresji stratnej polegających na wyznaczeniu i zachowaniu właściwości danych, które są decydujące w ich interpretacji i wykorzystaniu (np. tekstur i konturów w obrazach). Zwiększenie efektywności kompresji uzyskuje się kosztem znaczącej redukcji tła nie mającego wartości użytkowych. Metody te zwane ekstrakcyjnymi były stosowane głównie do kompresji obrazów medycznych. Określano w obrazie wszelką informację istotną diagnostycznie kodując odpowiednie regiony zainteresowań w sposób bezstratny, zaś pozostałe obszary, nieistotne w opinii specjalistów, kompresowano stratnie z dużym poziomem zniekształceń.

W latach 1977 i 1978 zostały opublikowane przez Lempela i Ziva dwa algorytmy bezstratnej kompresji, które stały się podstawą nowej grupy metod tzw. kodowania słownikowego [73, 74]. Metody te charakteryzuje faza modelowania dyskretnego (bazującego na identyczności ciągów symboli, bez modeli probabilistycznych), polegająca na wyznaczaniu słownika fraz powtarzających się ciągów danych. Słownik jest wtedy podstawową strukturą wykorzystywaną w procesie tworzenia reprezentacji kodowej. Jest ona konkatenacją kolejnych wskaźników pozycji słownika, które odpowiadają sukcesywnie kodowanym ciągom danych wejściowych. Od pierwszych liter nazwisk autorów oraz daty publikacji algorytmy te nazwano odpowiednio LZ77 i LZ78. Modyfikacje tych algorytmów, m.in. LZSS [75] i LZW [76] uzyskały duży walor użytkowy i znalazły zastosowanie w licznych archiwizerach, m.in w: Unix_Compress, ARC, PKZIP, LHA (LHarc), ARJ, RAR, 7-Zip, WinZip, PKArc, w formatach obrazów GIF i PNG, w metodzie Deflate.

W latach osiemdziesiątych XX wieku nastąpił intensywny rozwój technik kompresji. Na uwagę zasługują prace nad coraz doskonalszymi modelami adaptacyjnymi, a także rozwój metod kodowania transformacyjnego. Podejmowano próby wykorzystania różnych transformacji, np. Fouriera, Walsha-Hadamarda, sinusowej, Karhunena-Loevego, do upakowania energii sygnału w zredukowanej dziedzinie przekształcenia. Najlepsze rezultaty uzyskano dla dyskretnej transformaty kosinusowej DCT (*discrete cosine transform*), co znalazło odbicie w opracowanych na początku lat dziewięćdziesiątych standardach kompresji obrazów cyfrowych wielopoziomowych – JPEG i MPEG. Na rys. 3.36 przedstawiono schematyczne porównanie funkcji bazowych DCT i transformacji Fouriera, które ukazuje zaletę ciągłości kosinusowych funkcji bazowych przy granicach przedziału określoności, co zapewnia wyższą jakość rekonstruowanego po kwantyzacji sygnału rozwiniętego w przedziałach przyległych.



Rysunek 3.36: Porównanie baz przekształceń DCT i Fouriera w przypadku przedziałowego liczenia transformacji. Ciągłość funkcji DCT na granicach przedziałów (widoczna na rysunkach powyżej) powoduje dokładniejszą rekonstrukcję sygnału w przedziale (rysunki poniżej).

Ponadto opracowano wydajne implementacje kodowania arytmetycznego [77, 79], najefektywniejszego z kodów binarnych (stale optymalizowane szczególnie w zakresie modelowania, np. w ramach standardów JPEG2000, MPEG-4, JBIG-2 – zobacz [80, 82]). Kodery arytmetyczne wykorzystują kontekstowe modele zależności danych ograniczone niekiedy do alfabetu binarnego, uproszczone koncepcje obliczeń znacznie zmniejszające złożoność obliczeniową.

Przełom lat osiemdziesiątych i początek lat dziewięćdziesiątych to doskonalenie dwóch nowatorskich metod kompresji obrazów, bazujących na dość złożonym aparacie matematycznym. Pierwsza to metoda wykorzystująca przekształcenia fraktalne, pozwalająca uzyskać dużą skuteczność kompresji szczególnie dla obrazów naturalnych o niebogatej treści. Istotne zasługi w rozwoju tej techniki położyli między innymi Barnsley, Jacquin, Hurd. Drugą metodą, z którą związane są takie nazwiska jak: Mallat, Daubechies, Villasenor, Vetterli, Strang i wiele innych, jest kompresja na bazie wieloskalowych przekształceń falkowych (*wavelet transform*). Kompresja falkowa rozszerza znaną wcześniej koncepcję kodowania pasmowego (*subband coding*). Kodowanie falkowe, realizowane także w wersji bezstratnej za pomocą transformacji całkowitoliczbowych, zapewnia dużą efektywność kompresji sygnałów niestacjonarnych, w tym obrazów. Ma szereg zalet, takich jak: naturalnie uzyskiwany hierarchiczny opis informacji w przestrzeni wielorozdzielczej, możliwość łatwej i szybkiej adaptacji metody kodowania do lokalnej charakterystyki danych w różnej skali, z zachowaniem zależności z przestrzeni oryginalnej. Pozwala to zwiększyć efektywność kompresji obrazów do wartości często niemożliwych do uzyskania za pomocą innych metod.

W drugiej połowie lat dziewięćdziesiątych rośnie znaczenie metod kompresji obrazów statycznych (pojedynczych), które wykorzystują przekształcenia falkowe do dekompozycji danych źródłowych w różnych zastosowaniach. Przełomowa okazała się tutaj praca J.M. Shapiro [83] dotycząca algorytmu EZW (*embedded zerotree wavelet*), ukazująca efektywny sposób kodowania współczynników falkowych, znacznie zwiększający wydajność kompresji w stosunku do metod bazujących na DCT. W setkach publikacji przedstawiono coraz doskonalsze metody dekompozycji, kwantyzacji i kodowania współczynników falkowych, których efektem było opracowanie nowego standardu kompresji obrazów JPEG2000. Standard ten wykorzystuje koncepcję falkowej dekompozycji obrazów, elastyczny sposób kształtowania strumienia danych kodowanych w zależności od potrzeb użytkownika, dokładną kontrolę długości tego strumienia i optymalizację stopnia zniekształceń, metodę kompresji stratnej-do-bezstratnej (*lossy-to-lossless*) umożliwiającą efektywną kompresję stratną, sukcesywnie przechodzącą w bezstratną po dołączeniu wszystkich informacji do sekwencji wyjściowej kodera i wiele innych. Koder według JPEG2000 pozwala w wielu przypadkach zwiększyć blisko dwukrotnie efektywność kompresji w stosunku do kodera standardu JPEG. Kolejne części standardu JPEG2000 dotyczą różnych obszarów zastosowań (m.in.

transmisji bezprzewodowej, interaktywnych protokołów wymiany informacji obrazowej, zabezpieczeń praw własności w strumieniu kodowym, kina domowego i telewizji cyfrowej, indeksowania i opisu za pomocą deskryptorów obrazów).

Zastosowanie metod falkowych do kompresji wideo napotyka na pewne ograniczenia. Stosowane obecnie kodeki wykorzystują najczęściej transformację DCT z blokową estymacją i kompensacją ruchu. Do tej grupy należą doskonalone od początku lat dziewięćdziesiątych standardy multimedialne: H.261, MPEG-1, MPEG-2, H.263, H.263+, aż po H.264 (MPEG-4, część 10). Ponadto, prace nad algorytmami kompresji drugiej generacji (z obiektową analizą scen i bardziej elastyczną kompensacją ruchu) doprowadziły do opracowania nowego kodeka wideo w ramach MPEG-4, część II. W kolejnych przedstawicielach rodziny standardów MPEG skoncentrowano się bardziej na zagadnieniu indeksowania danych multimedialnych, opracowaniu deskryptorów opisujących treść i formę danych (MPEG-7). W roku 2000 rozpoczęto prace nad standardem MPEG-21, który jest "wielkim obrazem" ogromnej infrastruktury wymiany i konsumpcji treści multimedialnych, uwzględniającym mnogość istniejących już i rozwijanych narzędzi reprezentowania danych, określając ich wzajemne relacje i porządkując całą przestrzeń multimediów. Aktualny stan prac nad standardami JPEG i MPEG można śledzić na stronach odpowiednio http://www.jpeg.org/ i http://www.mpeg.org/.

3.3.3 Ograniczenia

Głównym ograniczeniem efektywności metod kompresji są zbyt uproszczone modele źródeł informacji wskutek dalece niedoskonałego modelowania rzeczywistości (założenia stacjonarności, gaussowskiej statystyki źródeł). Względność "optymalnych" dziś rozwiązań wynika także ze sposobu definiowania informacji bez odniesienia do warstwy semantycznej. Często nie sposób przełożyć posiadanej wiedzy a priori, dotyczącej analizowanego procesu, na parametry modelu statystycznego. Nie ma jednoznacznych definicji nadmiarowości, która może występować tak na poziomie pojedynczych danych, jak też obiektowym i semantycznym. Opisując informację trudno jest stwierdzić, jak zmiana parametrów modelu obiektu decyduje o utracie przez niego "tożsamości" wpływającej na ilość przesyłanej informacji. Wymaga to definicji pojęcia informacji na wyższym poziomie abstrakcji, a jest to przecież poziom użytkowy typowego odbiorcy.

Interesująca staje się postać reprezentacji danych optymalnej nie w kontekście przyjętych założeń, ale wobec bogatej rzeczywistości form, jakie przyjmuje informacja we współczesnym świecie. Rozważmy prosty przykład. W telewizyjnym teleturnieju uczestnicy zabawy odkrywają kolejno fragmenty obrazu próbując możliwie szybko rozpoznać jego treść. Niekiedy potrzeba bardzo niewiele odsłoniętych elementów, by zidentyfikować znany obraz. Cyfrowy obraz Mona Lisa skompresowany według standardu JPEG z zachowaniem wysokiej jakości rekonstrukcji to 150 kilobajtów danych. Jeśli zaczniemy stopniowo odsłaniać ten obraz

rekonstruując progresywnie kolejne bity skompresowanej reprezentacji, to już po dekompresji 1000 bajtów większość specjalistów potrafi rozpoznać ten obraz. Do rozpoznania obrazu Abrahama Lincolna w podobnym teście przeprowadzonym przez L. Harmona wystarczyło 756 bitów [101].

Problemem jest włączenie wiedzy o świecie dostępnej a priori w poszukiwanie optymalnej reprezentacji informacji obrazowej. Zniekształcona, cyfrowa rekonstrukcja Mona Lisy według JPEG jest tylko przybliżeniem, aproksymacją oryginału. O poziomie stratności (nieodwracalności) metody kompresji decyduje poziom aproksymacji danych źródłowych wymagany przez odbiorcę. Niekiedy informację może stanowić jedynie adres internetowy (wskaźnik odpowiedniego obiektu) lub też tekst "Obraz Mona Lisa" powodując reakcję odbiorcy adekwatną do upodobań, posiadanej wiedzy i dostępnych środków (np. wizytę w muzeum Louvre, obejrzenie wysokiej jakości reprodukcji posiadanej w domu czy też książki o dziełach sztuki etc). W innym przypadku odbiorca wymaga reprodukcji najwyższej jakości, co daje kompresja bezstratna cyfrowej postaci obrazu o najwyższej jakości. Występuje tutaj duże podobieństwo z zagadnieniem indeksowania.

3.3.4 Możliwości udoskonaleń

Możliwe jest bardziej ogólne podejście do zagadnień informacji, kompresji, granic efektywności kodowania, indeksowania. Informacja definiowana jest wtedy jako w dużym stopniu odwołanie do ogólnej wiedzy odbiorcy, pojęcie nadmiarowości konstruowane jest nie jako przewidywalna statystyczna zależność danych w obrębie przekazywanego zbioru (strumienia), ale w znacznie szerszej skali - jako nawiązanie do zasobów wiedzy i środków dostępnych odbiorcy, do realnej semantyki danych. Takie podejście pozwala opracować doskonalsze narzędzia kompresji, czego przykładem są próby optymalizacji kompresji obrazów w ramach standardu JPEG2000.

Zasadniczą cechą współczesnych metod kompresji jest elastyczność, zdolność doboru ilości, jakości i postaci informacji wynikowej (wyjściowej) w zależności od definiowanych potrzeb odbiorcy. Opracowanie skutecznego algorytmu kodowania wymaga zwykle optymalizacji wielokryterialnej, wykorzystania mechanizmów adaptacji do lokalnych cech sygnału (zbioru danych), a czasami nawet interakcji zmieniających procedurę kompresji w czasie rzeczywistym.

Dla obrazów stosuje się szereg metod eliminacji nadmiarowości przestrzennej poprzez dekompozycję obrazu do postaci, która jest bardziej podatna na kodowanie binarne. Są wśród nich metody predykcyjne, dekompozycji falkowej i pasmowej, blokowej, skanujące piksele według określonego porządku, dzielące wartości pikseli obrazu na szereg map bitowych (*bit plane encoding*) z największą korelacją wśród najstarszych bitów i inne.

Ważną grupę stanowią algorytmy predykcyjne: dana (tj. wartość piksela) przeznaczona aktualnie do zakodowania jest przewidywana na podstawie danych z

sąsiedztwa (najczęściej najbliższych w przestrzeni obrazu, które są potencjalnie najbardziej z nią skorelowane), które pojawiły się już wcześniej w kodowanej sekwencji (warunek przyczynowości, konieczny do rekonstrukcji obrazu źródłowego podczas dekodowania). Kodowana jest jedynie niedokładność tego przewidywania. Sposób kodowania predykcyjnego prowadzący do praktycznych implementacji nazywany jest DPCM (*differential pulse code modulation*). Początki metod DPCM związane są z pracami prowadzonymi w Bell Laboratories [84]. Poważnym ograniczeniem ich efektywności jest niedoskonałość predykcji powodowana m.in. brakiem możliwości ukształtowania kontekstu (sąsiedztwa) predykcji otaczającego kodowany piksel z każdej strony (konieczność zachowania warunku przyczynowości).

Modyfikacją metod predykcyjnych są algorytmy interpolacyjne HINT (*Hierarchical Interpolation*) [85, 86, 87], w których zastosowano kodowanie kolejnych wersji obrazu o rosnącej rozdzielczości. Dzięki temu do predykcji wykorzystane są wartości pikseli otaczające dany piksel z każdej strony, jednak nie zawsze leżące w najbliższej odległości. Otrzymane w wyniku predykcji obrazy różnicowe są kodowane z wykorzystaniem metod entropijnych (na podstawie probabilistycznych modeli źródeł informacji).

Nowsze rozwiązania zawierają także rozbudowaną fazę modelowania statystycznych zależności danych w określonym kontekście (np. metoda CALIC [88]), które bezpośrednio sterują pracą koderów entropijnych. Wykorzystuje się także adaptacyjne kodery binarne, dostosowane do charakterystyki danych za pomocą alfabetu binarnego, w tym binarne kodery arytmetyczne, metody kodowania serii jedynek, kody Golomba.

W grupie najbardziej efektywnych bezstratnych metod kodowania obrazów wymienić należy przede wszystkim wspomniany koder CALIC oraz standardy JPEG-LS [89], JPEG2000 [90], JBIG [92], JBIG2 [93] oraz AVC z MPEG-4 [91].

Podsumowując, zbiór stosowanych rozwiązań fazy modelowania można podzielić na cztery zasadnicze grupy:

- z prostym modelem statycznym w metodach: RLE dla serii bitów (np. koder Z [95]) i bajtów (format PCX), Golomba, Huffmana i w kodowaniu arytmetycznym do kodowania reszt predykcyjnych;
- z rozbudowanym modelem probabilistycznym wyższych rzędów (np. w koderach arytmetycznych map bitowych),
- ze słownikiem (metody słownikowe np. w formatach GIF [96] i PNG [97]);
- ze wstępną dekompozycją danych, gdy tworzona jest pośrednia reprezentacja: metody predykcyjne i interpolacyjne, kodowanie map bitowych, skanowanie danych według określonego porządku (np. po krzywej Peano), całkowitoliczbowe kodowanie transformacyjne (np. w AVC [98]), całkowitoliczbowe dekompozycje falkowe i pasmowe (np. w JPEG2000).

Faza binarnego kodowania jest najczęściej realizowana w trzech wariantach:

- przypisanie słów kodowych pojedynczym symbolom alfabetu źródła (metody Huffmana i Shannona-Fano) - kody o zmiennej długości słów;
- przypisanie słów kodowych (często o stałej długości) ciągom symboli wejściowych o zmiennej długości (RLE, kodowanie słownikowe); modyfikacją tych metod są adaptacyjne kodery słownikowe o zmiennym rozmiarze słownika (a więc także indeksów), jak również RLE z kodem Golomba, gdzie różnej długości ciągom symboli odpowiadają sekwencje (słowa) o zmiennej liczbie bitów;
- utworzenie sekwencji kodowej w postaci jednego binarnego słowa kodowego wyznaczanego sukcesywnie dla całego ciągu wejściowego (kodowanie arytmetyczne).

3.3.5 Odwracalna kompresja danych

Bitowo bezstratna kompresja obrazów oznacza przede wszystkim tworzenie możliwie oszczędnej ich reprezentacji z wykorzystaniem zasad statystycznej teorii informacji do celów archiwizacji. Przez ostatnie ponad 10 lat, od czasu opracowania efektywnych koderów CALIC i według standardu JPEG-LS [94], nie nastąpił żaden przełom, jakościowy postęp w dziedzinie odwracalnej kompresji obrazów.Wspomniane kodery odwracalne na etapie modelowania źródeł informacji wykorzystują 2W (dwuwymiarową) charakterystykę danych w postaci efektywnych modeli predykcji.

Obserwuje się natomiast tendencję modelowania źródeł informacji na niższym poziomie bez pośrednich modeli sąsiednich wartości pikseli źródłowych. Uproszczenia prowadza do zamiany: zamiast obiektów piksele, zamiast pikseli - rozkłady map bitowych. Stosuje się modelowanie lokalnych kontekstów wybranych przestrzeni bitowych niekoniecznie skorelowanych z obrazową interpretacją danych, szacowanie zależności danych traktowanych bardziej elementarnie (z uproszczonym alfabetem źródła), a co za tym idzie bardziej uniwersalnie.

W koderach odwracalnych kluczową rolę odgrywają metody modelowania kontekstu przy kodowaniu binarnym (zwykle arytmetycznym). W PPM (*prediction with partial string matching*) [102] przy modelowaniu prawdopodobieństw warunkowych adaptacyjnego kodera arytmetycznego wykorzystywane są konteksty różnych rozmiarów. Alfabet linii prawdopodobieństw określonego kontekstu rozszerzany jest dynamicznie, tj. niezerowy podprzedział określonego symbolu pojawia się w linii prawdopodobieństw dopiero wówczas, gdy symbol wystąpi w sekwencji wejściowej poprzedzony tym kontekstem. Ponadto w każdej linii zarezerwowany jest podprzedział symbolu "PRZEŁĄCZ", który służy do sygnalizacji zmiany rozmiaru kontekstu (dobierany jest możliwie najdłuższy kontekst

zaczynając od kontekstu masymalnego rzędu m). Poszczególne odmiany metody PPM, dostosowane do danych tekstowych, obrazowych, innych, różnią się sposobem określania linii prawdopodobieństw i kontekstów (rozmiarem, kształtem, uproszczeniem statystyki w danym kontekście).

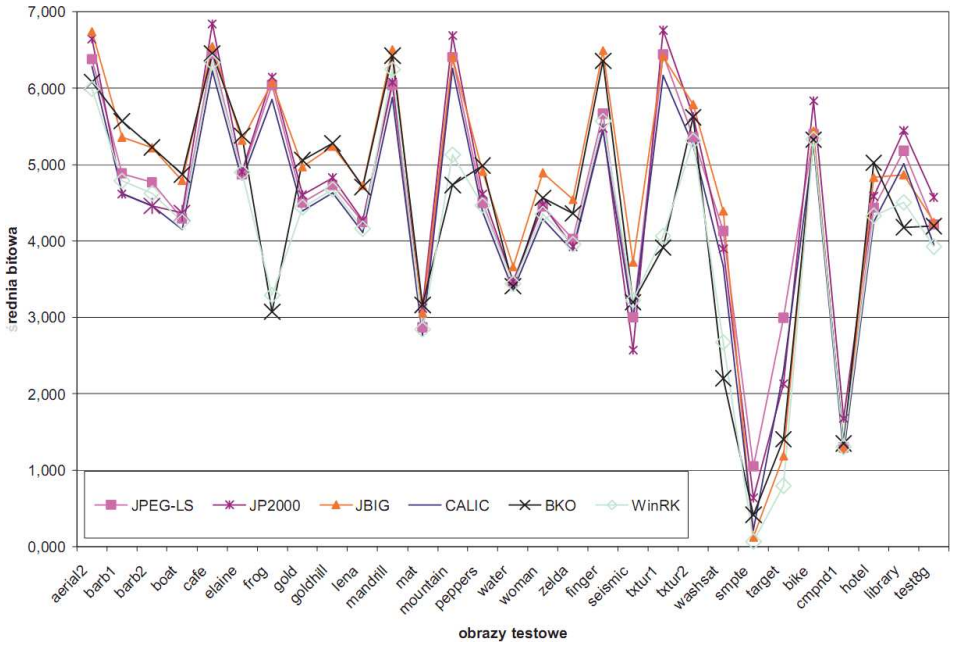
W przypadku, gdy efektywne opisanie jednym kontekstem złożonych lokalnych zależności danych źródłowych nie jest możliwe, stosowane jest niekiedy ważenie kilku modeli prawdopodobieństw bazujących na różnych kontekstach. Jeśli rozkład $P_S^{C_1}$ określono na podstawie kontekstu C_1, a $P_S^{C_2}$ na podstawie C_2, wówczas do kodowania można wykorzystać ważony rozkład prawdopodobieństw $\left(P_S^{C_1} + P_S^{C_2}\right)/2$. Metoda ważenia rozkładów prawdopodobieństw z wykorzystaniem struktury drzewa CTW (*context-tree weighting*) [103] pozwala wykorzystać wszystkie możliwe konteksty i modele kontekstu ograniczone rozmiarem m. CTW bazuje na binarnym alfabecie źródła informacji bez względu na rodzaj kodowanych danych (obraz, dźwięk, tekst itp.). Wymaga to wstępnej konwersji ciągu symboli nad alfabetem A_S w ciąg bitowy (tj. na wstępnej serializacji, jak w standardzie JBIG, czy też w metodzie BKO [108]), kodowany z wykorzystaniem binarnej struktury drzewa określającej konteksty i prawdopodobieństwa modeli źródła z pamięcią i bez do sterowania kodowaniem arytmetycznym.

Do określenia prawdopodobieństw ciągu symboli wykorzystuje się tzw. prawdopodobieństwo blokowe całego bloku bitów, obliczane na podstawie częstości dotychczasowych (w modelu przyczynowym) wystąpień zer i jedynek zakładając model bez pamięci. Zależności pomiędzy danymi uwzględnione zostały w tzw. prawdopodobieństwach ważonych obliczanych w węzłach tworzonego drzewa binarnego.

Inaczej wyznaczane jest też prawdopodobieństwo symboli, tj. według wyrażenia: $P_e(s = a_1) = \frac{n_1+1/2}{n_1+n2+1}$, gdzie alfabet źródła $A_S = \{a_1, a_2\}$, n_1, n_2 - liczba wystąpień odpowiednio a_1 i a_2 w kodowanym dotychczas ciągu źródłowym. $P_e(s = a_2)$ definiowane jest analogicznie.

Coraz większą rolę w praktycznych zastosowaniach odgrywają uniwersalne narzędzia do archiwizacji danych (tzw. archiwizery) wykorzystujące podobne metody statystycznego modelowania źródeł informacji z doborem parametrów zależnie od typu danych. Wykazują one dużą efektywność kompresji skutecznie konkurując z rozwiązaniami dedykowanymi np. kompresji obrazów. Dowodem tego są wyniki licznych eksperymentów (zobacz np. [104]), a także przedstawione poniżej rezultaty przeprowadzonych testów własnych (rys. 3.37).

Uniwersalny archiwizer WinRK (wykorzystujący odmianę metody PPM do modelowania kontekstu) pozwolił w kilku przypadkach na wyraźne zmniejszenie długości zakodowanych danych obrazowych (średnio o blisko 6%) w stosunku do najefektywniejszego kodera obrazów CALIC. W stosunku do standardów JPEG-LS i JPEG2000 poprawa efektywności sięgnęła odpowiednio 10% i 11%. Nowa wersja WinRK (niestety na obecnym etapie realizacji nieco zbyt złożona oblicze-



Rysunek 3.37: Porównanie efektywności bezstratnych koderów obrazów. Wykorzystano 29 obrazów testowych w 8 bitowej skali szarości (z prac standaryzacyjnych komitetu JPEG), o nazwach jak na rysunku. Wartości średnich bitowych zbioru wszystkich obrazów testowych dla poszczególnych koderów wynoszą odpowiednio: 4,55 bnp (JPEG_LS [105]), 4,60 bnp (JPEG2000 VM8.6), 4,75 bnp (JBIG [106]), 4,35 bnp (CALIC [107]), 4,36 (BKO [108]) oraz 4,1 bnp (WinRK v. 2.1.6 [109]). Skrót bnp oznacza "bitów na piksel".

niowo) z algorytmem modelowania PWCM pozwala skrócić skompresowane dane obrazowe o dodatkowe kilka procent.

Dwa uniwersalne kodery map bitowych: według standardu JBIG oraz wykorzystujący nieco bardziej dopasowaną do danych obrazowych metodę serializacji BKO pozwoliły również uzyskać wysoką efektywność kompresji obrazów testowych, przy czym średnia bitowa dla BKO jest na poziomie średniej kodera CALIC.

Okazuje się, że modele informacji obrazowej wyższego poziomu (obiektowe, predykcyjne, transformacji liniowych) nie usprawniają procesu redukcji nadmiarowości wejściowej reprezentacji danych, są za mało elastyczne w dopasowaniu modeli źródeł do lokalnych (chwilowych) cech sygnału. Są one zaś szczególnie istotne przy porządkowaniu, tworzeniu hierarchii informacji uwzględniając niekiedy przesłanki semantyczne w kompresji selektywnej.

3.3.6 Kompresja selektywna

W kompresji selektywnej poszukiwana jest efektywna reprezentacja semantycznie definiowanej informacji wyrażonej wstępnie danymi w postaci źródłowej, zaś indeksowanie to tworzenie indeksu (indeksów) obiektów (kolekcji obiektów) danego typu, czyli spisu (zestawu) tych obiektów ze względu na określone ich właściwości (cechy, czyli wartości ustalonego argumentu indeksu - zobacz [27]) mające znaczenie dla użytkownika. Przy wyborze argumentu indeksu (tj. określonych właściwości indeksowanych obiektów) istotna jest rozróżnialność obiektów, a więc wydobycie wyjątkowej treści każdego z obiektów z ogólnego, mało znaczącego tła. Kluczową rolę odgrywa tu dotarcie do realnej informacji związanej z każdym obiektem (w naszym przypadku obrazem), tj. istoty przekazywanej treści. Semantyczne określenie informacji odgrywa tutaj decydującą rolę.

W indeksie występują identyfikatory obiektów wskazujące (pośrednio lub bezpośrednio) ich fizyczną lokalizację. Korzystające z indeksów wyszukiwarki dostarczają na żądanie odbiorcy obiekty o określonych cechach sięgając niekiedy do ich skompresowanej reprezentacji, dekodując i prezentując np. obraz.

Nowe standardy kompresji, np. JPEG2000 przewidują efektywne łączenie kompresji z indeksowaniem, wyznaczanie cech w dziedzinie przekształconych w kompresji danych, co często daje lepszą selektywność wyszukiwania pozwalając porównywać optymalizowane reprezentacje informacji z uwzględnieniem ich znaczenia.

Praca nad standardem kompresji obrazów JPEG2000 jest przykładem szerszego spojrzenia na zagadnienia kompresji selektywnej w ostatnich latach, bogatszego o uwzględnienie wymagań odbiorcy i subiektywnego znaczenia przesyłanych mu danych. Możliwa jest ingerencja w procedury przetwarzania danych oraz organizacji strumienia wyjściowego, a przez to wykorzystanie wiedzy wynikłej z doświadczenia użytkownika, tradycji, kultury itd. Opracowano zagadnienia interakcyjnej transmisji, rekonstrukcji obrazów z ingerencją odbiorcy w jakość i ilość otrzymywanych danych, które stają się realną informacją weryfikowaną przez odbiorcę (jego cele). Wprowadzono elementy zabezpieczania danych, ochrony własności intelektualnych, ukrywania informacji przed postronnymi odbiorcami. W ostatnich miesiącach trwają intensywne prace nad indeksowaniem obrazów w ramach nowego standardu JPSearch, tworzonego przez członków komitetów JPEG i MPEG jako zintegrowane uzupełnienie standardów JPEG. Realizacja takich założeń nie jest prosta i na obecnym etapie, niedoskonała. Przeprowadziliśmy szereg testów porównawczych kodera JPEG2000 z realizacją starszego standardu JPEG. Ocena uzyskanych rezultatów nie jest jednoznaczna - zobacz przykład z rys. 3.38.

O ile w przypadku obrazu testowego Lenna wyraźnie lepsza jakość rekonstrukcji kompensuje kilkukrotnie większą złożoność obliczeniową kodera JPEG2000, to jednak w przypadku obrazu Frog trudno wskazać lepszą wersję w ocenie subiek-

Rysunek 3.38: Porównanie efektywności selektywnej kompresji obrazów według standardów JPEG i JPEG2000: a) obraz testowy Lenna odtworzony po kompresji w stopniu 50:1 (JPEG daje wartość PSNR równą 28,18 dB, a dla JPEG2000 PSNR=30,5 dB); b) obraz testowy Frog zrekonstruowany po kompresji w stopniu 30:1 (JPEG daje PSNR=24,8 dB, a JPEG2000 - PSNR=25,46 dB). W ocenie subiektywnej trudno wskazać wersję Frog o lepszej jakości.

tywnej. Główną zaletą JPEG2000 nie jest więc jego efektywność, ale większa możliwość dostosowania do semantycznej warstwy przekazywanej informacji.

Zastosowania medyczne wymagają jeszcze szerszego spojrzenia na zagadnienie kompresji. Nie bez przyczyny nie ma dotąd powszechnej zgody środowisk lekarskich na stosowanie kompresji stratnej w obrazowaniu medycznym. Obawy budzi wiarygodność diagnostyczna rekonstruowanych obrazów oraz sposób ustalenia optymalnych parametrów procesu kompresji. Nie przekonuje wiele prac dowodzących nawet poprawy jakości obrazów po kompresji [[110],[111], nie ma jasnych reguł prawnych [112]. Sformułowanie obiektywnych kryteriów optymalizacji algorytmu kompresji tych obrazów jest nadal sprawą otwartą, choć jest

przedmiotem licznych badań od wielu lat.

Mechanizmy selekcji informacji

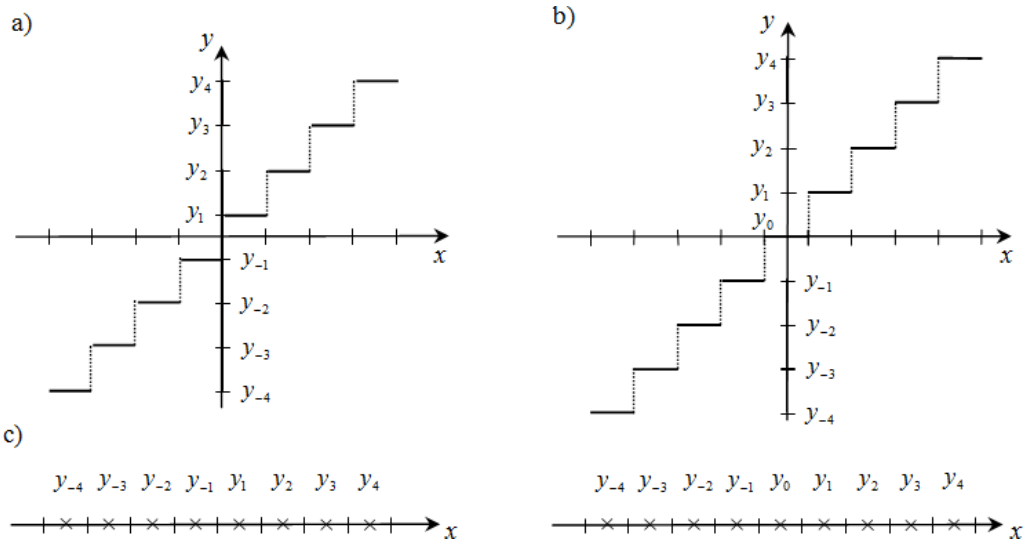
Podstawową metodą usuwania przede wszystkim semantycznej nadmiarowości danych, kosztem nieodwracalnej numerycznie postaci danych rekonstruowanych w procesie kompresji jest kwantyzacja. Równie uniwersalnym rozwiązaniem jest porządkowanie danych (nierosnąco) według ilości przenoszonych przez nie informacji i kontrolowanie zakresu kodowania ciągu źródłowego, zależnie od wymagań zastosowania. Przykładowo, przy zmniejszeniu kanału przesyłowego przekazana zostaje jedynie najważniejsza, najefektywniej kodowana część informacji, a w razie możliwości czy dodatkowych żądań odbiorcy – pozostała. Zależnie od zastosowań, selekcja informacji może przebiegać w schematach zdecydowanie bardziej specyficznych, przy zastosowaniu określonych wymagań semantycznych czy użytkowych właściwych danemu przekazowi informacji. Przykładowo, wstępna detekcja określonych obiektów w obrazie może zróżnicować sposób kodowania wybranych fragmentów obrazu – obszary zawierające istotne elementy przekazu kodowane są w sposób numerycznie odwracalny, podczas gdy pozostałe – jedynie w zakresie koniecznego tła. Przyjrzyjmy się przede wszystkim podstawowemu mechanizmowi kwantyzacji, typowemu dla wielu metod kompresji stratnej.

Kwantyzacja Mechanizm kwantyzacji, w najprostszej, skalarnej i równomiernej wersji przedstawiony na rys. 3.39 rozumiany jest w algorytmach kompresji jako złożenie dwu odwzorowań:

- kodera (kwantyzacja prosta): jako odwzorowanie nieskończonego (skończonego dużego) zbioru wartości rzeczywistych dziedziny, określonego i podzielonego na przedziały (na rysunku wzdłuż osi x), na zbiór indeksów przedziałów kwantyzacji, dzielących dziedzinę w sposób rozłączny i zupełny; indeksy $d_i \in \{\ldots, -4, -3, -2, -1, 1, 2, 3, 4, \ldots\}$ wskazują przedziały, do których wpadają kolejne dane wejściowe x_i; wartości d_i są w dalszej kolejności bezstratnie kodowane;

- dekodera (kwantyzacja odwrotna): jako odwzorowanie zbioru indeksów d_i w zbiór rzeczywistych, rekonstruowanych wartości y_i – reprezentantów przyporządkowanych przedziałów kwantyzacji.

Kwantyzator jest jednoznacznie zdefiniowany przez zbiór przedziałów kwantyzacji (*de facto* zbiór punktów granicznych, dzielących zakres wartości sygnału wejściowego na te przedziały) oraz zbiór wartości rekonstruowanych. Wartość rekonstrukcji powinna przybliżać zbiór wartości kwantowanych, należących do danego przedziału, w sposób możliwie wiarygodny, tj. minimalizujący zniekształcenie (błąd kwantyzacji) według przyjętej miary. Jeśli operator kwantyzacji (tj.



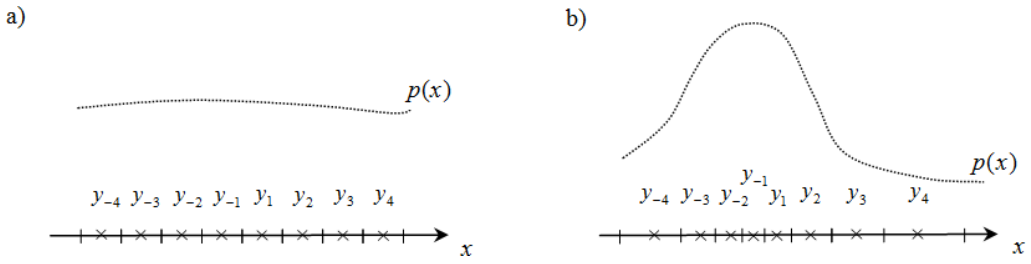
Rysunek 3.39: Opis prostych schematów kwantyzacji skalarnej, równomiernej jako odwzorowania nieskończonego, ciągłego zbioru wartości (reprezentowanego osią x) w skończony zbiór dyskretny (poziomy wzdłuż osi y): a) bez zera, b) z zerem. Poniżej, uproszczony wykres jednej osi x z naniesionymi wartościami rekonstrukcji w kolejnych przedziałach kwantyzacji – odpowiednio dla schematów bez zera i z zerem: c).

połączonych funkcji kodera i dekodera) oznaczymy przez $Q(\cdot)$, a ciąg K wartości wejściowych przez $X = \{x_i\}_{i=1}^{K}$, to w wyniku procesu kwantyzacji tych wartości do M poziomów alfabetu rekonstrukcji $Y = \{y_j\}_{j=1}^{M}$, otrzymujemy ciąg wartości rekonstruowanych $\tilde{X} = \{\tilde{x}_i\}_{i=1}^{K}$ przybliżający sygnał oryginalny. Kwantyzacja jest więc przekształceniem:

$$Q : \mathbb{R} \to \mathbb{R}, \quad \text{tak że } Q(x_i) = \tilde{x}_i \tag{3.100}$$

przy czym $\tilde{x}_i = y_j$, jeśli $x_i \in [\beta_{j-1}, \beta_j)$, a $\{\beta_j\}_{j=0}^{M}$ to granice przedziałów kwantyzacji $B_1, B_2, \ldots, B_M$. Jakość tego przybliżenia określa błąd kwantyzacji $D_Q(X, \tilde{X}) = D_Q$. Jedną z najczęściej stosowanych miar jest średniokwadratowy błąd kwantyzacji postaci $D_Q = \frac{1}{K}\sum_{i=1}^{K}(x_i = \tilde{x}_i)^2$.

Uśrednienie błędu kwantyzacji po wszystkich próbach wymusza minimalizację tego błędu w skali globalnej. Korzystniej jest, przy założeniu danej liczby przedziałów, zagęścić przedziały kwantyzacji w obszarach skali wartości licznie pokrytych przez dane wejściowe (histogramowe maksima), zapewniając wierniejszą rekonstrukcję danych dominujących. Uzależnienie schematu kwantyzacji od estymaty funkcji gęstości prawdopodobieństwa zmiennej X prowadzi do kwantyzacji nierównomiernej (ze zróżnicowanym rozmiarem przedziałów kwantyzacji – zobacz rys. 3.40), realizowany metodą Lloyda-Maxa [99, 100].



Rysunek 3.40: Przykłady projektowania schematu kwantyzacji zależnie od postaci estymowanej funkcji gęstości prawdopodobieństwa $p(x)$ zmiennej opisującej zbiór danych wejściowych: a) kwantyzacja równomierna; b) nierównomierna.

Można wykazać, że optymalny średniokwadratowo model kwantyzacji spełnia następujące dwa warunki centroidu i najbliższego sąsiada:

- mając dane przedziały kwantyzacji (przypisane do funkcji kodera), najlepszym odwzorowaniem liczb całkowitych indeksów kwantyzacji w zbiór wartości rekonstruowanych (funkcja dekodera) są środki masy (centroidy) tych przedziałów kwantyzacji, obliczane według

$$y_j = \frac{\int_{\beta_{j-1}}^{\beta_j} xp(x)dx}{\int_{\beta_{j-1}}^{\beta_j} p(x)dx} \tag{3.101}$$

- mając dany zbiór poziomów rekonstrukcji (dekoder), najlepszym określeniem przedziałów kwantyzacji jest ustalenie położenia punktów granicznych w środku pomiędzy kolejnymi wartościami rekonstrukcji, według reguły

$$\beta_j = \frac{y_j + y_{j+1}}{2} \tag{3.102}$$

  odpowiada to przypisaniu każdej wartości wejściowej do najbliższej odległościowo wartości rekonstrukcji.

Często wykorzystywanym w kompresji rozwiązaniem jest optymalizacja kwantyzatora z kryterium uwzględniającym entropię strumienia indeksów kwantyzacji. Wymaga ono poruszania się w przestrzeni R-D (*Rate-Distortion*), ustalając najlepszą relację pomiędzy uzyskiwaną średnią bitową strumienia (jego entropią) a wartością błędu kwantyzacji. Chodzi o równoczesne zmniejszanie obu tych wielkości do pewnej granicy, zależnej przede wszystkim od rodzaju kwantyzacji, właściowości X, jak również od sposobu kodowania wartości wyjściowych kwantyzatora.

3.3.7 Nowe paradygmaty kompresji

Ciągle powracają pytania o naturę informacji, skuteczny jej opis ilościowy i jakościowy, metody ekstrakcji, klasyfikacji, indeksowania informacji zmieniającej swój

charakter, zasięg, nośnik, znaczenie. Zmianie ulegają paradygmaty, czyli wzorce, standardy stosowanych z sukcesem rozwiązań. Standardy opracowywane przez kilka lat, w przeciągu kilku następnych tracą swoją aktualność, użyteczność (np. standard kompresji obrazów JPEG, coraz mniej użyteczny w wielu zastosowaniach).

Uzupełnieniem przedstawionych rozważań są zebrane na podstawie licznych eksperymentów (w tym własnych) w zakresie kompresji danych (w tym obrazowych) stwierdzenia i wskazówki sugerujące kierunek doskonalenia obowiązujących dziś paradygmatów kompresji

Uniwersalny archiwizer

W przypadku odwracalnej kompresji danych istotne okazują się następujące spostrzeżenia:

- świat 'ujęty' w formie rejestracji określonego stanu rzeczywistości w danej chwili nie ma charakteru gaussowskiego; znaczy to, że:
  - dekorelacja danych nie oznacza ich statystycznej niezależności, a model źródła bez pamięci staje się mało skuteczny; wynika stąd, że wykorzystanie modeli z pamięcią w algorytmach kwantyzacji i binarnego kodowania znacząco zwiększa efektywność opisu źródeł informacji;
- obrazy zazwyczaj nie mają charakteru stacjonarnego, ani nawet ergodycznego więc:
  - ważną rolę odgrywa adaptacyjność algorytmów modelowani źródeł, stosowanie metod segmentacji i lokalnej klasyfikacji, zarówno w skali globalnej (na poziomie obrazów) jak i lokalnej (na poziomie fragmentów obrazów);
  - wymagane są wiarygodne sposoby szacowania prawdopodobieństw symboli na podstawie niewielkiej statystyki danych, skuteczne metody upraszczania, kwantyzacji, wyboru kontekstów;
  - inne niż statystyczne metody opisu źródeł informacji okazują się nieskuteczne (predykcyjne, transformacyjnego kodowania, obiektowe).

Najlepszym rozwiązaniem do celów archiwizacji wydaje się uniwersalny koder danych z możliwością doboru modelu źródła zależnie od charakteru danych.

Elastyczny koder wielu skal

Selektywna kompresja informacji rządzi się innymi prawami. Warto zauważyć, że:

- modelowane obrazy zachowują cechę samopodobieństwa w zmieniającej się skali (po rozciągnięciu obrazu do większej dziedziny i przeskalowaniu zachowane zostają właściwości statystyczne oryginału), a więc:
  - kluczową rolę odgrywają przekształcenia (transformacje) danych ze skalowaniem rozdzielczości, zachowujące informację o położeniu i treści częstotliwościowej współczynników poszczególnych skal;
  - szczególnie ważnym elementem wynikającym z rozkładu zależności danych w przestrzeni ze skalowaniem rozdzielczości jest kodowanie pozycji współczynników transformaty i ich znaczeń (względem ustalonej wartości progowej).

Efektywny koder obrazów powinien zapewnić następujące cechy reprezentowanej informacji:

- hierarchiczność korelującą z semantycznym znaczeniem kolejnych przybliżeń źródła informacji (progresja treści) zapewniająca progresję optymalną w sensie R(D) (tj. od maksymalnego przyrostu informacji na bit transmitowanych danych do minimalnego);
- selektywność rozkładu informacji w poszczególnych obszarach wielu skal dziedziny z możliwością rekonstrukcji informacji zarówno w pełnej, niezakłóconej wersji (nawet w sensie odwracalnej kompresji danych), jak też odpowiednio wybranej, przybliżonej (w możliwie szerokim zakresie) na poziomie pojedynczych pikseli, obiektowym lub semantycznym;
- elastyczność pozwalającą na realizację porządku kodowania i dekodowania informacji zależnie od wymagań użytkownika (interaktywność procesu kompresji/dekompresji).

Trwają poszukiwania optymalnych przekształceń obrazów w wielu skalach. Klasyczne falkowe przekształcenie obrazów zakłada separowane jądro transformacji (sekwencja przekształceń 1W po wierszach i kolumnach), co nie pozwala dobrze opisać obiektów w obrazach (występuje uwypuklenie informacji w kierunku pionowym i poziomym kosztem innych kierunków). Aproksymacja nieliniowa gładkiej krzywej jest często lepsza w przypadku wykorzystania transformacji wielorozdzielczych z jądrem 2W (*curvelets*), filtrów kierunkowych (*contourlets*), przekształceń geometrycznych po podziale obrazu na bloki o różnej wielkości, co daje lokalny opis informacji w różnej skali (*beamlets*, *wedgelets*, *bandlets* i inne). Zastosowanie oszczędniejszej, bardziej upakowanej i uporządkowanej reprezentacji danych w dziedzinie tych przekształceń pozwala również zwiększyć selektywność rozwiązań uwzględniających znaczenie rozkładów symboli na poszczególnych poziomach tak uzyskanej hierarchii informacji obrazowej. Wykorzystywane są tutaj

podstawy teoretyczne kształtowania baz - atomów czas-częstotliwość dostarczane przez analizę harmoniczną i funkcjonalną. Uzyskane już rezultaty poprawy efektywności w stosunku do JPEG2000 są obiecujące i sugerują kierunek doskonalenia metod kompresji sygnałów w ramach planowanego przez komitet JPEG nowego standardu kompresji - AIC (*advanced image coding*).

3.3.8 Podsumowanie

Trudno sobie wyobrazić rozwój nowoczesnej telekomunikacji, telemedycyny, e-biznesu, obieg różnego typu dokumentów w Internecie, doskonalenie nowoczesnych technologii sieciowych typu grid, skuteczne wyszukiwarki czy globalne bazy danych bez optymalizowanych algorytmów kompresji, oszczędnych formatów, użytecznych standardów z wieloaspektową charakterystyką strumieni danych przesyłanych i gromadzonych, sprzętowych realizacji kodeków etc.

Kompresja danych jako efektywna i uniwersalna koncepcja reprezentowania informacji poparta efektywnymi realizacjami, przedmiot działalności naukowej i sztuki inżynierskiej jest obiektem zainteresowań w wielu obszarach zastosowań współczesnych technologii teleinformatycznych. Rozumiana jako poszukiwanie, tworzenie i badanie oszczędnej, a zarazem funkcjonalnej reprezentacji danych związanych z informacją także w sensie znaczeniowym pozwala przyspieszyć i wzbogacić przekaz wiadomości, zwiększyć niezawodność wymiany informacji pomiędzy nadawcą i odbiorcą, lepiej archiwizować posiadane zasoby.

W przypadku obrazów szczególnie istotny okazuje się aspekt semantyczny, interpretacyjny, domagający się uwzględnienia charakterystyki (profilu) odbiorcy w procesie wyznaczania użytecznej reprezentacji informacji. Praca nad metodami kompresji danych obrazowych w zakresie określania znaczeń przesyłanych pakietów danych zyskuje przez to wymiar bardziej humanistyczny i służebny wobec wielu oczekiwań, zamierzeń i ograniczeń współczesnego człowieka.

Zwrócono uwagę na wiele otwartych kwestii przede wszystkim w zakresie rozumienia pojęcia informacji, sposobów selekcji informacji zależnie od subiektywnych wymagań odbiorcy, charakterystyki percepcyjnych zdolności odbiorcy, integracji z rozwiązaniami stosowanymi w indeksowaniu i innych metodach przetwarzania informacji. Nieustające, zintensyfikowane w ostatnich latach prace nad multimedialnymi standardami komitetów JPEG i MPEG stanowią dowód dużego zaangażowania badaczy oraz rosnących potrzeb i niesłabnącego zainteresowania rynku. Aby im sprostać konieczna jest silniejsza integracja zarówno w obrębie różnych dyscyplin i specjalności technicznych, jak i obszarów wiedzy związanych z zastosowaniem technik informacyjnych.

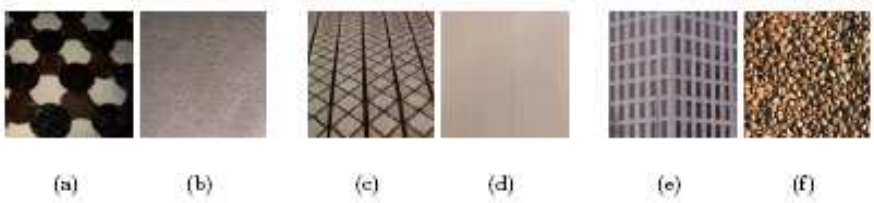
Ponieważ dziś coraz cenniejszy staje się czas i selekcja realnej informacji, a każdy dzień przynosi nowe technologie gromadzenia i przekazu informacji, poszukiwanie optymalnych metod reprezentowania informacji staje się palącą potrzebą chwili. Sugerowane kierunki badań wydają się być użyteczne.

3.4 Metody indeksowania

Deskryptory stanowią serce systemów indeksowania, bo bezpośrednio dotyczą opisu treści znaczonej i przeszukiwanej. Od ich skuteczności zależą możliwości różnicowania treści przy zachowaniu niezmienniczości określonej ich specyfiki w różnych warunkach akwizycji i wymiany danych. Poniżej opisano przykładowe deskryptory jako rozszerzenie prostych pomysłów zamieszczonych w p. 2.3.4.

Cechy teksturowe Tamury

W [118] zaproponowano zestaw 6 cech teksturowych, korespondujących z percepcją człowieka: skrośność (*coarsness*) , kontrast (*contrast*) , kierunkowość (*directionality*), liniowość (*line-likeness*), regularność (*regularity*) i zgrubność (*roughness*). Przeprowadzone przez samych autorów koncepcji testy [118, 119] wykazały szczególną użyteczność 3 pierwszych miar.



Rysunek 3.41: Przykłady właściwości teksturowych według wybranych cech Tamury. a) duża skrośność b) mała skrośność c) duży kontrast d) mały kontrast e) ukierunkowana f) nieukierunkowana. Obrazki zaczerpnięte z [120]

Cenną zaletą cech teksturowych Tamury jest ich znacząca korelacja z percepcją człowieka. Aspekt tej korelacji był jednym z głównych założeń autorów, które przedstawione zostały w [118]. Można tam znaleźć także interesujące i rzadko spotykane opisy testów z użytkownikami, oceniającymi korelację miar obliczeniowych z ich subiektywnym wrażeniem. Przykładowe obrazy, pokazujące tekstury o skrajnych wartościach wykorzystanych cech, przedstawiona są na rysunku 3.41. Cechy te definiujemy następująco:

1. Skrośność - daje informacje o wielkości ziarna w teksturze. Im wartości skrośności jest większa, tym większe mamy ziarno. Idea wyznaczania skrośności w mierze Tamury polega na użyciu operatorów o różnym rozmiarze. Dokładnie procedura jej wyznaczania wygląda następująco:

   a) dla każdego punktu (n_0, n_1) wyznacz średnią wartość w jego sąsiedztwie. Wielkość tego sąsiedztwa to potęgi dwójki, czyli np. $1 \times 1, 2 \times$

$2, 4 \times 4, \ldots, 32 \times 32$:

$$A_k(n_0, n_1) = \frac{1}{2^{2k}} \sum_{i=1}^{2^{2k}} \sum_{j=1}^{2^{2k}} X(n_0 - 2^{k-1} + i, n_1 - 2^{k-1} + j) \quad (3.103)$$

b) dla każdego punktu (n_0, n_1) wyznacz różnice pomiędzy nie nachodzącymi na siebie obszarami po przeciwległych stronach punktu w kierunku poziomym i pionowym:

$$E_k^h(n_0, n_1) = |A_k(n_0 + 2^{k-1}, n_1) - A_k(n_0 - 2^{k-1}, n_1)| \quad (3.104)$$

oraz:

$$E_k^v(n_0, n_1) = |A_k(n_0, n_1 + 2^{k-1}) - A_k(n_0, n_1 - 2^{k-1})| \quad (3.105)$$

c) w każdym punkcie (n_0, n_1) wybierz rozmiar sąsiedztwa, który prowadzi do największej wartości różnicy:

$$S(n_0, n_1) = \arg \max_{k=1\ldots5} \max_{d=h,v} E_k^d(n_0, n_1) \quad (3.106)$$

d) weź średnią z 2^S jako miarę skrośności dla obrazu:

$$F_{crs} = \frac{1}{N_0 N1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} 2^{S(n_0,n_1)} \quad (3.107)$$

2. Kontrast - w szerszym sensie kontrast stanowi o jakości obrazu. Można wyróżnić 4 czynniki, które mają wpływ na kontrast obrazu w skali szarości:

   - dynamika zakresu poziomów jasności,
   - polaryzacja rozłożenia czerni i bieli na histogramie poziomów szarości,
   - ostrość krawędzi,
   - okres powtarzalności wzorców tekstury.

   Kontrast obrazu jest wyznaczany jako:

$$F_{con} = \frac{\sigma}{\alpha_4^z} \quad (3.108)$$

   gdzie $\alpha_4 = \frac{\mu_4}{\sigma_4}$, $\mu_4 = \frac{1}{N_0 N1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} (X(n_0, n_1) - \mu)^4$ jest czwartym momentem średniej μ, σ^2 jest wariancją poziomów szarości obrazu, a z zostało eksperymentalnie dobrane jako $\frac{1}{4}$.

3. Kierunkowość - kierunkowość jest cechą mówiącą o występowaniu w teksturze kierunku. Nie chodzi tu jednak o to, jaki ten kierunek jest, a jedynie

−1	0	1
−1	0	1
−1	0	1

−1	−1	−1
0	0	0
1	1	1

o określenie czy on występuje, tak więc dwie tekstury różniące się jedynie orientacją będą posiadały identyczną kierunkowość.

Do wyznaczenia kierunkowości wyznaczane jest różnicowe przybliżenie pochodnej poziomej $\triangle_H$ i pionowej $\triangle_V$ poprzez splot obrazu $X(n_0, n_1)$ z maskami, odpowiednio (filtr Prewitta):

i wtedy dla każdego punktu (n_0, n_1) wyznaczana jest zależność:

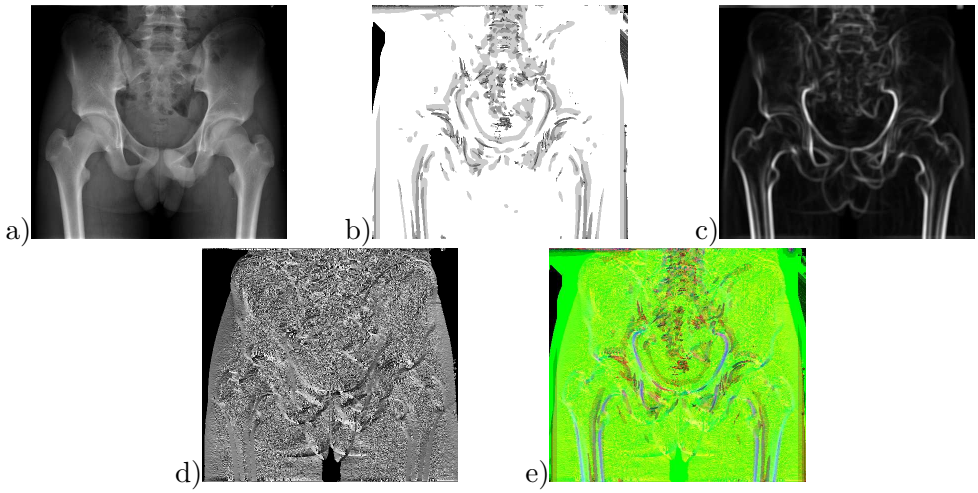
$$\theta = \frac{\pi}{2} + \tan^{-1} \frac{\triangle_V(n_0, n_1)}{\triangle_H(n_0, n_1)} \tag{3.109}$$

Z tych wartości jest następnie wyznaczany 16–przedziałowy histogram H_D, stanowiący opis kierunkowości.

Aby przedstawione powyżej cechy wykorzystać w indeksowaniu obrazów, trzeba je odpowiednio dostosować. W swojej pierwotnej formie każda z cech Tamury daje skalarny wynik dla całego obrazu. W zastosowaniu do indeksowania wskazana byłaby reprezentacja bardziej szczegółowa, w skrajnym przypadku dająca wartość cechy dla każdego piksela w obrazie. Takie podejście pozwala na stworzenie histogramu cechy i jego łatwe porównywanie z histogramami w bazie referencyjnej. W przypadku skrośności, aby otrzymać wartość cechy dla każdego piksela, wykonywane są kroki a) do c) algorytmu, dając w efekcie miarę skrośności dla każdego piksela [28]. Kontrast jest wyznaczany w sąsiedztwie 15×15 dla każdego piksela. Kierunkowość dla każdego piksela wyznaczana jest w ten sposób, że zamiast filtru Prewitta zastosowany jest filtr Sobela, natomiast kierunkowość θ wyznaczana jest dla każdego piksela, określając kierunkowość w jego sąsiedztwie. W przypadku kierunkowości warto też zwrócić uwagę na fakt, że dla tej miary autorzy w [118] nie opisują dostatecznie jasno sposobu wyznaczenia globalnej miary kierunkowości – te same problemy mieli zresztą autorzy [28, 121], którzy w obu przypadkach wspominali o konieczności adaptacji algorytmu wyznaczania skośności.

Mając trzy wartości dla każdego piksela: skośność, kontrast i kierunkowość, można wyznaczyć trójwymiarowy histogram.

Rysunek 3.42 przedstawia przykładowe rysunki obrazujące, odpowiednio, obraz oryginalny (a), skrośność (b), kontrast (c) i kierunkowość (d), oraz wszystkie te trzy wielkości, przedstawione na obrazie barwnym jako składowe RGB (e).



Rysunek 3.42: Obraz cech Tamury dla przykładowego obrazu: a) obraz oryginalny, b) obraz skrośności, c) obraz kontrastu, d) obraz kierunkowości, e) obraz skrośności, kontrastu i kierunkowości jako kolorowy obraz RGB

Globalny Deskryptor Teksturowy

W [122, 28] opisano deskryptor teksturowy, próbujący opisać całościową charakterystykę teksturową obrazu. Jest to wektor danych, który obejmuje:

- wymiar fraktalny, jako miara 'chropowatości' tekstury; wymiar fraktalny jest narzędziem matematycznym, wykorzystanym w analizie tekstur [123, 122];
- cechy wyznaczone na podstawie macierzy powinowactwa; elementami tej macierzy są estymowane prawdopodobieństwa $s(i, j)$ wystąpienia par punktów o jasnościach i oraz j, dla określonej odległości pomiędzy punktami i przyjętym kierunku analizy; macierz powinowactwa opisuje więc teksturę poprzez informację o rozkładzie jasności punktów w otoczeniu dla określonej odległości i kierunku;
- entropia, jako miara nieuporządkowania danych obrazowych;
- skrośność, jako wielkość charakteryzująca wielkość ziarna tekstury.

W [122] celem autora było opracowanie deskryptora, który w możliwie szerokim zakresie będzie opisywał właściwości tekstury w kontekście ich korelacji z percepcją człowieka. Jest to więc założenie podobne do tego, które postawili autorzy z grupy Tamury w [118] i z racji swej potencjalnie dużej użyteczności zostało zaimplementowane i wykorzystane w opracowanym systemie indeksowania.

Tak więc, deskryptor ten stanowi 35–wymiarowy wektor, w skład którego wchodzi 1 wartość na wymiar fraktalny, 1 na entropię, 1 na skrośność i 32 wartości wyznaczone z macierzy powinowactwa. Z macierzy powinowactwa, zgodnie z

wynikami prac w [122], wyznaczana jest średnia jasność, kontrast, drugi moment różnicowy oraz entropia. Każda z tych wielkości wyznaczana jest dla odległości 1 oraz 2 oraz dla kątów 0°, 45°, 90° i 135°, co w efekcie daje 32 wartości. Bliższe szczegóły dotyczące tego deskryptora można znaleźć w [122].

3.5 Komputerowa inteligencja

Jest to zagadnienie frapujące badaczy od dziesiątków lat, tak jak od tysięcy lat badana jest ta fascynująca cecha ludzkiego umysłu[3]. Inteligencja oznacza przede wszystkim sprawność w zakresie zasadniczych zdolności umysłowych, takich jak: myślenie i rozwiązywanie problemów, postrzeganie, rozpoznawanie, rozumienie, zapamiętywanie, posługiwanie się liczbami, symbolami, wyobraźnię przestrzenną, rozumowanie przez indukcję i dedukcję (ogólnie procesy poznawcze, które generują nowe treści na podstawie rozumowania), płynne używanie języka, odczuwanie i kontrola emocji, koncentracja uwagi, planowanie, wyciąganie wniosków z doświadczeń, kreatywność (twórczość, oryginalność), zdolność oceny i interpretacji, podejmowanie decyzji, itp. Zdolności konkretnych osób są zwykle ukierunkowane na szczególne typy inteligencji, np. językową, emocjonalną czy twórczą.

W jakich aspektach potencjał nauk komputerowych sięga ludzkiej inteligencji, których zdolności dotyczy, czy stanowi docelowo jej substytut, replikę, a może innowację? Niełatwo znaleźć odpowiedź na te pytania, chociaż niewątpliwie warto je rozważyć w kontekście tak trudnego zadania, jak użytkowanie przekazu informacji obrazowej.

3.5.1 Inteligencja ludzka

Fransis Galton stwierdził mało wnikliwie w 1883 roku [**?**], że inteligencja to podstawowa zdolność umysłu, decydująca o sukcesie jednostki w „walce o byt", zależna przede wszystkim od energii działania oraz wrażliwości zmysłów. Trafniej zdefiniował to pojęcie Alfred Binet (1905) jako mniej więcej zdolność do wydawania trafnych sądów, bazująca na adekwatnym rozumieniu sytuacji oraz logicznym wnioskowaniu, zwłaszcza w odniesieniu do problemów i sytuacji dnia codziennego [**?**]. Warte podkreślenia jest bardzo pragmatyczne w tych definicjach rozumienie inteligencji.

Od poziomu inteligencji zależy poprawność rozumienia złożonych, abstrakcyjnych problemów i skuteczność poszukiwania trafnych rozwiązań, a także sprawność działania w sytuacjach trudnych, niecodziennych, w warunkach licznych ograniczeń, niepewności, w stresie. To umiejętność stawiania celów, planowania, podjęcia decyzji w warunkach ekstremalnych, nagłych, sytuacjach "życiowych", to możliwość twórczego myślenia o własnym myśleniu.

Szczególną rolę w zachowaniach inteligentnych zajmuje rozum i intuicja w kontekście ludzkiej świadomości. Inteligencja oznacza sprawne uczenie się, myślenie abstrakcyjne, adaptacja do bieżącej sytuacji, podejmowanie skutecznych decyzji, ale także świadomość określonych celów, praw, wartości, wyzwań, ist-

[3] Umysł to ogół aktywności mózgu ludzkiego, przede wszystkim takich, których posiadania człowiek jest świadomy (Wikipedia)

niejących uwarunkowań. Świadomość pozwala formułować zadania, stawiać cele, określać dążenia – stanowi więc warstwę podmiotową zachowań inteligentnych.

Rozum kojarzony jest przede wszystkim z myśleniem logicznym. To zdolność do operowania pojęciami abstrakcyjnymi, do analitycznego myślenia i wyciągania wniosków na podstawie dostępnych danych, informacji, wiedzy, eksperymentów. To umiejętność uczenia się, używania zdobytych doświadczeń i posiadanej wiedzy do rozwiązywania konkretnych problemów życiowych. Znacznie trudniejszym w definicji pojęciem jest intuicja.

Intuicja natomiast oznacza zdolność do nagłego przebłysku myślowego, w którym dostrzega się rozwiązanie problemu lub znajduje odpowiedź na nurtujące pytanie. To możność szybkiego dopasowania rozwiązania problemu do zaistniałych uwarunkowań. Jest to proces bardziej kreatywny i działający na wyższym poziomie abstrakcji w porównaniu do myślenia logicznego.

Słowo *intuitio* (łac.) oznacza wejrzenie, wewnętrzne przekonanie, że mam rację. Chodzi więc o zdolność bezpośredniego pojmowania, szybkiego dotarcia do bezpośredniej wiedzy bez udziału obserwacji, rozumu, świadomej myśli – myślenie intuicyjne jest podobne do percepcji, postrzegania, czyli błyskawiczne i bez wysiłku. Intuicja to inaczej

- wgląd, olśnienie pojawiające się w trakcie rozwiązywania problemu,
- przeczucie, zdolność przewidywania, szybkie rozpoznanie,
- przekonanie, którego nie można w pełni uzasadnić,
- poznanie docierające do istoty rzeczy,

Intuicja to nie są emocje, a proces podświadomy, którego nie można kontrolować – przewidywanie, domyślanie się niebazujące na przejrzystym wnioskowaniu. Zasadniczo można jedynie dopuszczać lub odrzucać podawane przez intuicję rozwiązania.

Próbując nieco bardziej zrozumieć mechanizm tego zjawiska, wydaje się, że na intuicyjne myślenie mają wpływ takie czynniki jak:

- mimowolne uczenie się, automatyczne, nieuświadomione zdobywanie wiedzy,
- kreatywne działanie poprzez automatyzm zachowania według ustalonych, aczkolwiek w większości nieznanych reguł reagowania na podstawie własnego doświadczenia życiowego,
- markery somatyczne jako automatyczny sygnał przewidywanych skutków podjęcia decyzji – są to specjalne rodzaje uczuć, emocji, odnoszące się do wcześniejszych doświadczeń, o charakterze ostrzegającym (negatywne) lub zachęcającym (pozytywne).

Świadomość określana jest jako stan psychiczny, w którym jednostka zdaje sobie sprawę ze zjawisk wewnętrznych (własne procesy myślowe, psychika) oraz zjawisk zachodzących w środowisku zewnętrznym (świadomość otoczenia, możliwość reakcji). Jako samoświadomość jest specyficzną, gatunkową cechą człowieka, stanowi podstawę tworzenia wiedzy i zapamiętywania, a więc działań inteligentnych. Jest zawsze intencjonalna, nakierowana na jakiś przedmiot materialny lub abstrakcyjny i powiązana z odczuciem własnego "ja". W szczególności jest to:

- stan przytomności, czuwania, odbierania bodźców;
- zdolność do celowej orientacji i odczuwania, tj. przeżywania doznań i stanów emocjonalnych.

3.5.2 Mechanizmy i schematy poszukiwania rozwiązań

W sprawnym i stającym na odpowiednio wysokim poziomie abstrakcji, tj. inteligentnym, działaniu odgrywają istotną rolę takie elementy jak pamięć robocza (jej pojemność, trwałość, strategie gospodarowania), strategie poznawcze (m.in, wyobrażeniowa, analityczna, globalna – ważny jest wybór właściwej strategii), kontrola poznawcza (względem czynności ważnych dla rozwiązania zadania), zasoby uwagi (wielkość "mocy przetwarzania" systemu poznawczego, odnosząca się do pojęcia ludzkiej świadomości i woli).

Dwa podstawowe sposoby rozumowania (wnioskowania) to metoda dedukcyjna i indukcyjna. Można też mówić o wnioskowaniu:

- do przodu, które rozpoczyna się od analizy faktów, a następnie na podstawie dostępnych reguł i faktów generowane są nowe fakty tak długo, aż wśród nich znajdzie się poszukiwane przez użytkownika rozwiązanie (cel) lub zabraknie reguł;
- wstecz – wnioskowanie sprowadza się do wersyfikacji postawionej na początku hipotezy poprzez poszukiwanie argumentów (dowodów), które ją potwierdzą lub obalą.

Możliwe jest także łączenie obu strategii, stosowanie ich naprzemiennie zależnie od bieżącej sytuacji w zbiorze wygenerowanych stanów – jeśli zdefiniowany cel wydaje się mało osiągalny, brakuje skutecznych reguł (operatorów) przybliżających, można zdefiniować pomocniczy podcel, nadający kierunek lokalnym poszukiwaniom.

Rozumowanie dedukcyjne jest prostym wyciąganiem wniosków, nie wymagającym tworzenia nowych twierdzeń czy pojęć. Wykorzystuje się jednie formułę logicznej konsekwencji: na podstawie istniejących przesłanej (faktów, wiarygodnych danych treningowych) oraz dostępnej wiedzy tzw. zastanej formułuje się wnioski (tj. konkluzje, nowe fakty), które stają się wiedzą nabytą. Zakładając

poprawność wiedzy (reguł logicznych), do której się odwołujemy, z prawdziwości przesłanek wynika logiczna prawdziwość konkluzji. Taki rodzaj wnioskowania stanowi podstawę przede wszystkim systemów eksperckich (inaczej ekspertowych), ważnego obszaru skutecznych rozwiązań sztucznej inteligencji.

Indukcja logiczna to sposób rozumowania polegający na wyprowadzaniu nowych twierdzeń, weryfikacji hipotez, sugerowaniu możliwości zaistnienia nowych faktów na podstawie intuicyjnej analizy wejściowych przesłanek.

W rozumowaniu indukcyjnym obserwując określone fakty (a więc postrzegając prawdę) szukamy ich wyjaśnienia w postaci możliwie wiarygodnych przyczyn, bazując na dostępnej wiedzy, zastanej lub nabytej we wspomagającym rozumowaniu dedukcyjnym. Weryfikując domniemane przyczyny zadajemy pytanie o ich prawdziwość i na tej podstawie wyprowadzamy wniosek ogólny, nową teorię czy prawo.

Prawdziwość przyczyn potwierdzonych faktów wynika w takim rozumowaniu z prawdziwości stosowanych reguł, stąd szczególna rola algorytmu wnioskującego.

Poszukiwanie inteligentnych rozwiązań realnych, życiowych problemów często nie jest proste. Trudno jest z góry określić ciąg czynności prowadzących do ich rozwiązania – jednym ze sposobów jest systematyczna analiza kolejnych alternatyw. Zaletą takiego rozwiązania jest łatwość formułowania kolejnych zadań. Wymagane jest jedynie określenie zbioru stanów przestrzeni rozwiązywanego problemu (w tym stanu początkowego i zbioru możliwych stanów końcowych) oraz zbioru operatorów przekształcających stany tej przestrzeni (zastosowane do określonych stanów generują nowe stany). Rozwiązanie polega na wyznaczeniu ciągu operatorów przekształcających stan początkowy w stan końcowy.

Korzysta się przy tym z dobranych strategii realizacji procesu przeszukiwań przestrzeni stanów, od metod ślepych po zmyślne heurystyki. Poszukiwanie żądanego stanu – najlepszego rozwiązania odbywa się nierzadko w sposób względny, subiektywny, zależny od reguł wypracowanych doświadczalnie, bazujących na opiniach ekspertów. Metody analityczne są zwykle niepraktyczne,

Metody ślepe nie wykorzystują żadnej informacji o zadaniu – specyfice rozwiązywanego problemu, dzięki czemu mają one charakter uniwersalny. Polegające na kolejnym przeszukiwaniu (niekiedy wszystkich) możliwych dróg prowadzących do rozwiązania według z góry ustalonego porządku, w szczególności:

- deterministycznie, np.
  - w głąb, wszerz, dwukierunkowo, z jednolitym kosztem – pomocne są tutaj grafowe czy drzewiaste struktury stanów – węzłów;
  - zachłannie – po ekspansji stanów badane są nowe węzły i najbardziej obiecujący z nich jest wybierany do dalszej ekspansji; taka lokalna optymalizacja uniemożliwia powrót do żadnego przodka aktualnie badanego węzła;

  - siłowe, *brute force*, polegające na sukcesywnym sprawdzeniu wszystkich możliwych kombinacji w poszukiwaniu rozwiązania problemu, bez jakiejkolwiek szczegółowej analizy, co prowadzi do prostych realizacji i dużej złożoności obliczeniowej, zapewniając pełen sukces wyznaczenia optymalnego rozwiązania.

- niedeterministycznie, np. metodą Monte Carlo, z losowym wyborem punktów w przestrzeni dopuszczalnych rozwiązań i obliczaniem dlań funkcji celu – najlepsze rozwiązanie jest uznawane za rozwiązanie problemu.

W zadaniach przeszukiwania mianem "heurystyczne" określa się wszelkie reguły, zasady, prawa, kryteria i intuicje (również takie, których konieczność ani skuteczność nie jest całkowicie pewna), które umożliwiają wybranie najbardziej efektownych kierunków działania zmierzających do celu. Istotne jest dobre dopasowanie do problemu, wykorzystanie lokalnej czy czasowej charakterystyki stanów. Heurystyka oznacza metodę znajdowania przyzwoitego (dobrego?) rozwiązania przy akceptowalnych nakładach obliczeniowych. Nie daje gwarancji uzyskania optymalnych czy choćby prawidłowych rozwiązań, a nierzadko nawet bez oszacowania, jak blisko optymalnego jest otrzymane rozwiązanie. Jej stosowanie zawsze grozi pominięciem najlepszego ruchu.

W heurystyce ważne są sensowne przypuszczenia odnośnie kierunku poszukiwania rozwiązań, choćby przez wykluczanie rozwiązań nie rokujących sukcesów, zawężanie przestrzeni przeszukiwań, wybór możliwie najkrótszej, najbardziej prawdopodobnej drogi. Stosowanie heurystyki powinno umożliwiać uniknięcie badania tzw. ślepych ścieżek i skuteczne wykorzystanie zdobytych w trakcje badania obserwacji. Spodziewanym efektem jest średnia poprawa efektywności, przy braku poprawy w przypadku pesymistycznym.

Heurystyka jest więc rozumiana jako praktyczna strategia poprawiająca efektywność znajdowania rozwiązania złożonych problemów. Ogólniej, jest to nauka o metodach i regułach rządzących dokonywaniem odkryć i tworzeniem wynalazków. Inaczej, heurystyka nazywamy metodologię twórczego rozwiązywania zadań czy problemów złożonych (tj. wymagających w wyczerpujących rozwiązaniach olbrzymich ilości obliczeń) poprzez eksperyment, metodę prób i błędów, analogie, odwołanie do doświadczenia. Wykorzystywane jest podejście logiczne, matematyczne, ale też komputerowe (numeryczne), przez eksperyment, często za pomocą metody prób i błędów, odwoływania się do analogii, uogólnień. Heurystyka dąży do rozwiązania najkrótszą drogą, omijając mniej obiecujące ścieżki, wykorzystuje proste kryterium wyboru kierunku, a jej działania nie daje się analizować.

3.5.3 Sztuczna inteligencja

Sztuczna inteligencja kojarzona jest z maszyną myślącą, czyli są to metody, algorytmy, urządzenia posiadające (skutecznie naśladujące) funkcje ludzkiego umy-

słu. Inaczej, sztuczna inteligencja jest nauką o maszynach wykonujących zadania, które wymagają inteligencji, gdy są rozwiązywane przez człowieka.

Sztuczna inteligencja (*artificial intelligence* – AI) to dział informatyki, którego przedmiotem jest badanie reguł rządzących inteligentnymi zachowaniami człowieka, tworzenie formalnych modeli tych zachowań i — w konsekwencji — realizacja narzędzi komputerowych (za pomocą algorytmów, procedur, oprogramowania, rozwiązań sprzętowych) symulujących, naśladujących lub wspomagających te zachowania.

SI ma powiązania z psychologią, medycyną, fizjologią, bioniką, cybernetyką, teorią gier, modelowaniem matematycznym i in., czerpiąc pojęcia, metody i rezultaty oraz ubogacając poprzez oferowanie własnych pojęć i aparatu badawczego, metod obiektywizacji i formalizacji wiedzy. Przykładowe, wykorzystywane metody i narzędzia to – w warstwie konceptualnej – języki programowania (głównie Lisp i Prolog), języki systemów eksperckich (CLIPS, jego rozszerzenia – np. Jess, Flops, OPS5, Smalltalk), środowiskowe programy ułatwiające implementacje systemów (Pro Genesis, KEE, Loops, Level 5 Object, Aion Execution System), systemy szkieletowe (ExSys, DecisionPro, PC-Shell, G2, XpertRule), zintegrowany pakiet oprogramowania narzędziowego (np. SPHINX), algorytmy wyszukiwania informacji, poszukiwania rozwiązań problemów złożonych, dopasowywania wzorców i in., systemy wspomagania rozpoznania, decyzji, schematy i procedury działań zaradczych, przewidywania, ostrzegania, planowania, sterowania, zaś w warstwie materialnej – komputery o specjalnej architekturze, urządzenia komunikowania się z komputerem (interfejsy, różne technologie interakcji), inteligentne roboty itp.

Komputerowe realizacje SI wykorzystuje się konkretniej do rozpoznawania kształtów (np. pisma, elementów obrazów), dźwięków (np. mowy), do gry (np. w szachy, strategicznych), dowodzenia twierdzeń, wyszukiwania informacji, analizy wyników eksperymentalnych, w celu komponowania muzyki, tłumaczenia testów, formułowania ekspertyz (ocen, opinii), sterowania, monitorowania, śledzenia obiektów, robotów i procesów technologicznych, do sugerowania prostych diagnoz lekarskich, w koncepcji inteligentnego domu, zarządzania sprzętem, aktualizacji informacji i zasobów wiedzy, klasyfikacji zagadnień, problemów itp.

Jednoznaczna definicja systemów SI jest trudna, a granice płynne. Ogólnie są to systemy bazujące na wiedzy i doświadczeniu, wykorzystujące bardziej jakościową wiedzę niż modele matematyczne, charakteryzujące się symbolicznym wnioskowaniem, chociaż z biegiem lat w coraz większym stopniu wykorzystujące metody numeryczne. Algorytmy odwołują się coraz częściej do sprawdzonych heurystyk, poszukując wręcz najbardziej korzystnych uproszczeń w sytuacji zastanej. Tworzone systemy informacyjne (maszyny) mają zdolność uczenia się, pozyskiwania wiedzy, głębokiej adaptacyjności oraz autonomiczności.

Wśród fundamentalnych obszarów SI wyróżnia się przede wszystkim repre-

zentację wiedzy, systemy ekspertowe, sieci neuronowe, algorytmy genetyczne, szerzej – ewolucyjne, kognitywistykę, teorię gier, przetwarzanie języka, rozumienie mowy, widzenie komputerowe, uczenie maszyn, logikę rozmytą, programowanie automatyczne i robotykę.

Zasadniczy podział zagadnień sztucznej inteligencji ogranicza się do:

- silnej SI, dotyczącej systemów myślących, odwołującej się do osiągnięć kognitywistyki (tj. nauki zajmującej się badaniem, wyjaśnianiem i modelowaniem umysłu oraz procesów poznawczych, w tym percepcją, reprezentacją, emocjami, świadomością, pamięcią, rozumowaniem, mową, komunikacją itp.);

- słabej SI, nazywanej inteligencją obliczeniową, *soft computing*, zajmującej się problemami szczegółowymi, o jasno zdefiniowanym celu oraz kryteriach wykorzystując logikę, teorie zbiorów, teorię automatów, probabilistykę i statystykę itp.; podstawowe metody to automatyczne wnioskowanie, transmutacje wiedzy, stosowanie heurystyk, algorytmy mrówkowe, maszynowe uczenie się (z nadzorem, bez nadzoru, odkrywanie asocjacji i wzorców sekwencji, boty (tj. narzędzia do przeszukiwania i pozyskiwania wiedzy, z mechanizmem decyzyjnym na bazie wcześniej zdobytej wiedzy) itd.

Aproksymacja i optymalizacja

Zadania rozwiązywane w ramach SI można przyporządkować dwóm postawowym kategoriom: zadania aproksymacji oraz zadania optymalizacji.

Aproksymacją nazywamy znajdowanie ciągłego modelu zjawiska, cechy czy sygnału za pomocą funkcji czy krzywej przechodzącej w pobliżu zadanego, ziarnistego (dyskretnego) zbioru punktów. Zwykle aproksymację rozumie się więc w kontekście funkcjonalnego opisu danej dziedziny. Jest to problem dobrania nieznanej funkcji (zespołu funkcji określonej klasy, kawałków funkcji, krzywej itp.) na podstawie ograniczonych informacji o danym procesie, tj. skończonej liczby wartości (często zmierzonych, a więc obarczonych błędem pomiaru) danej dziedziny. Inaczej, mamy więc do czynienia z zagadnieniem przybliżania (dopasowywania) danych (*data fitting*).

Zadania aproksymacji (inaczej ekstrapolacji, rozpoznania) obejmują przede wszystkim

- znajdowanie ukrytych zależności pomiędzy danymi,

- przewidywanie (predykcja) zachowań obiektów, przebiegu zdarzeń, wyników własnych działań, prognozowanie trendów;

- uogólnienia wiedzy (”skoro wszyscy sportowcy – biegacze trenują przynajmniej pięć dni w tygodniu, to brak takiego treningu może oznaczać niebycie biegaczem”);

- uzupełnianie fragmentów zniszczonych zdjęć, zwiększanie rozdzielczości obrazów cyfrowych w celu wyświetlenia na wysokiej jakości monitorze;
- rozpoznanie (efekt klasyfikacji) nieznanych obiektów na podstawie znanych przykładów, np. rozpoznawanie obrazów, mowy, OCR;
- sterowanie obiektami, modelowanie zachowań, naśladowanie, animacja.

  Podstawowy schemat postępowania obejmuje wybór uzasadnionej klasy funkcji dobrze aproksymujących właściwości opisywanego zjawiska (jest to zasadniczy przedmiot rozważań teorii aproksymacji), a następnie procedurę (algorytm) dobierania konkretnej postaci funkcji odpowiadającej pomierzonym wartościom (to przede wszystkim obszar analizy i metod numerycznych). Użyteczna klasa funkcji aproksymujących to przede wszystkim funkcje względnie gładkie (gładkość to cecha sygnałów naturalnych), stanowiące zbiór możliwie zupełny (tj. kompletny, obejmujący opisem wszystkie możliwe przypadki) względem rozważanej grupy problemów. Korzystne jest przy tym, gdy są to funkcje łatwe w komputerowej obróbce (liczenie wartości, określanie pochodnych i całek itp.). Przykładem stosowanych klas aproksymacji opisów naturalnych zjawisk są funkcje wielomianowe, przedziałami wielomianowe, funkcje sklejane.

  Formalnie, mając dane niektóre wartości nieznanej funkcji f na X, tj. ciąg przykładów (próbek) – zbioru treningowego $(x_1, y_1), \ldots, (x_n, y_n)$, chcemy zgadnąć wartości $f : X \to Y$ w innych punktach dziedziny, tj. w dowolnym $x_0 \in X$. Szukana f, dająca zgadywane $y_0 = f(x_0)$, wyznaczana jest przy określonych ograniczeniach, wynikających z przyjętego kryterium aproksymacji.f jest możliwie wiarygodnym uogólnieniem charakteryzującym dany problem, rodzajem modelu określonej cechy dziedziny.

  Kryterium aproksymacji może mieć różną postać, zwykle jedną z poniższych

  – $$\forall_{i=1,\ldots,n} f(x_i) = y_i \quad \text{(interpolacja)} \tag{3.110}$$

  – $$\forall_{i=1,\ldots,n} |f(x_i) - y_i| < \epsilon \tag{3.111}$$

    gdzie ϵ jest maksymalnym dopuszczalnym błędem w punkcie

  – $$\min_f \sum_{i=1}^{n} |f(x_i) - y_i| \tag{3.112}$$

  – $$\max_f \Pr(f(x_1), \ldots, f(x_n)) \tag{3.113}$$

    gdzie f rozumiana jest jako zmienna losowa, proces losowy lub pole losowe.

W przypadku obrazów, obok dwuwymiarowych funkcji przybliżających rozkład wartości jasności danej grupy pikseli, stosowany jest też opis konturów obiektów za pomocą krzywych S konstruowanych w płaszczyźnie obrazów. Do najbardziej użytecznych klas krzywych aproksymujących należy zaliczyć parametryczne krzywe wielomianowe typu Beziera, Hermite'a czy też krzywe sklejane. W kryteriach aproksymacji, analogicznych do (**??**) występuje euklidesowa metryka $\|S - (x_i, y_i)\|$.

Znajdowanie optymalnych rozwiązań dla bardziej złożonych zagadnień aproksymacji, a takimi niewątpliwie jest większość przybliżeń realnych obiektów i cech obrazowych, nie jest zadaniem prostym. Wydaje się, że do dziś aktualne jest ogólne stwierdzenie P.J. Daviesa z 1965 roku [**?**], że jednym z najbardziej zaskakujących faktów teorii aproksymacji jest nieskuteczność najprostszych i najbardziej naturalnych rozwiązań. Nierzadko w użytecznych realiach ocieramy się o problemy NP-trudne.

Optymalizacja jest metodą poszukiwania wśród wielu alternatyw rozwiązania najlepszego (optymalnego), ocenianego według wiarygodnego, liczbowego kryterium jakości. Przykładowo, są to problemy minimalizacji funkcji kosztu poszukiwań, minimalizacji funkcji błędu względem rozwiązania wzorcowego, maksymalizacja wygranej (w grach logicznych) itp. W złożonych problemach często nie wystarcza jedno kryterium – mówimy wtedy o optymalizacji wielokryterialnej, np. odwieczny problem maksymalizacji zysków przy minimalizacji kosztów czy też maksymalizacji jakości kompresowanego stratnie obrazu przy minimalizacji średniej bitowej jego reprezentacji. Dobór metody przetwarzania obrazów źródłowych, kiedy chcemy jednocześnie zredukować szum, wyostrzyć krawędzie i zachować oryginalne cechy tekstury jest też dobrym przykładem takiego zagadnienia.

Przez X oznaczmy dowolny, skończony zbiór wszystkich możliwych rozwiązań problemu (tzw. przestrzeń stanów). Do oceny rozwiązań – stanów wykorzystajmy rzeczywistą funkcję celu $f: X \to R$. Zadaniem jest znaleźć takie $x_0 \in X$ według jednego z kryteriów

– maksimum

$$x_0 = \arg\max_{x \in X}\{f(x)\} \tag{3.114}$$

– minimum

$$x_0 = \arg\min_{x \in X}\{f(x)\} \tag{3.115}$$

Klasycznym przykładem zadań optymalizacji jest sortowanie, poszukiwanie wyjścia z labiryntu, poszukiwanie najkrótszej drogi dojścia do celu (np. problem komiwojażera), posunięcia giełdowe maksymalizujące zysk, zarządzanie pamięcią operacyjną czy dostępem do zasobów itp.

Realne problemy optymalizacji mają zwykle olbrzymią przestrzeń stanów, niepozwalającą zastosować trywialnych metod typu *brute force* jako praktycznego rozwiązania. W takich przypadkach poszukiwane są skuteczne heurystyki, pozwalające przeglądać przestrzeń stanów w rozsądnych wymiarach czasowych. Jeśli problem da się opisać analitycznie, co nie jest częste w realnych zjawiskach, zadanie optymalizacji sprowadza się do policzenia odpowiedniego ekstremum zadanej funkcji celu (lub funkcjonału). Wykorzystuje się wtedy zwykle układ równań uzyskany poprzez przyrównanie składowych gradientu funkcji celu do zera. Można też poszukiwać lokalnych ekstremów kierując się kierunkiem gradientu. Stosowany jest też losowy wybór rozwiązań według określonego scenariusza, np. w algorytmach ewolucyjnych.

Łatwo zauważyć, że zadanie optymalizacji może znaleźć zastosowanie w problemie aproksymacji, np. do wyszukania ze skończonego zbioru funkcji danej klasy rozwiązania dającego minimalny błąd przybliżenia danych według kryterium (3.112). Odwrotnie, aproksymację uproszczonych rozwiązań, jako pewnego rodzaju redukcję przestrzeni stanów można wykorzystać do rozwiązania problemu optymalizacji (np. w metodzie Newtona).

3.5.4 Komputer (nie)może być inteligentny

Ponieważ termin "inteligencja" trudno precyzyjnie zdefiniować, zmierzyć, jednoznacznie ocenić, Turing zaproponował test oceny subiektywnej z kryterium weryfikacji: nie sposób odróżnić komputera od człowieka na podstawie udzielanych odpowiedzi. Celem jest stwierdzenie, czy maszyna jest inteligentna. Osoba oceniająca zadaje pytania, na które odpowiadają w sposób anonimowy maszyna oraz człowiek. Jeśli maszyna zostanie uznana za człowieka, to można o niej powiedzieć, że jest inteligentna.

Toczy się spór o świadomość, tj. możliwość wytworzenia świadomych systemów sztucznej inteligencji. Wielu przedstawicieli nauki o procesach poznawczych (*cognitive science*), m.in. M. Minsky, widzi taką możliwość, zaś sceptycy (m.in. J.R. Searle, R. Penrose) twierdzą, że świadomość jest jedyną w swoim rodzaju właściwością ludzkiego mózgu, który nie przypomina w działaniu komputera.

3.5.5 Obliczeniowa mądrość

Pojęcie komputerowej (ew. obliczeniowej) mądrości, *computational sapience (wisdom)* (rozważane m.in. w [14]) obejmuje przede wszystkim:

- szeroki i selektywny dostęp do informacji oraz wiedzy (niezastąpiona rola Internetu, efektywnych metod indeksowania zawartością, nowoczesnych technologii semantycznych, gridowych, chmur obliczeniowych, itp.);

- dominującą role obrazowego przekazu informacji, w tym

  - efektywne reprezentacje treści,
  - wiarygodna charakterystyka odbiorcy – użytkownika,
  - automatyczne rozumienie obrazów lub komputerowe wspomaganie rozumienia obrazów,

- maksymalne wykorzystanie ludzkich zdolności z wykorzystaniem konwencji komputerowego asystenta, dzięki m.in.
  - integracji dostępnych środków, zasobów oraz metod (m.in. komputerowego wspomagania),
  - inteligentnemu interfejsowi człowiek – komputer.

3.5.6 Formalizacja wiedzy

Jak wspomniano, wiedza ekspertów staje się często źródłem sformalizowanych, logicznych reguł opisujących wzajemne relacje pomiędzy obiektami i ich cechami, niejako kształtuje przydatne w danym obszarze wiedzy metody wnioskowania, opisuje dane uczące czy weryfikuje rezultaty "inteligentnych" zachowań systemu komputerowego. Przyjrzyjmy się nieco bliżej zasadom i istniejącym możliwościom na styku ludzkiej wiedzy eksperckiej oraz komputerowych możliwości reprezentowania i opisu informacji, uzupełnionych olbrzymią mocą obliczeniową.

W użytecznych zastosowaniach komputerowej inteligencji coraz większą rolę odgrywa formalizacja wiedzy, czyli tworzenie maszynowej czy komputerowej reprezentacji wiedzy danej dziedziny (czyli wiedzy dziedzinowej), właściwej danemu zastosowaniu. Chodzi z grubsza o to, by wiedza i związane z nią umiejętności danej dziedziny wyrazić na sposób "zrozumiały" dla maszyny, by mogła stać się przedmiotem analiz, odwołań, wnioskowań, służących automatycznemu rozwiązywaniu zagadnień optymalizacji i aproksymacji w konkretnych realiach zastosowań.

Można to inaczej opisać jako tworzenie precyzyjnego modelu danego zakresu wiedzy, obejmującego możliwie kompletny jej zapis, z zachowaniem odpowiedniego poziomu abstrakcji hierarchii pojęć, zależności, reguł wnioskowania czy odniesień do realistycznej warstwy decyzyjnej, będącej skutkiem właściwej interpretacji wiedzy odniesionej do uwarunkowań danego przypadku. Służy temu m.in. coraz częściej stosowane narzędzie ontologii.

Filozoficzne pojęcie ontologii

Ontologia to teoria bytu, istoty, istnienia i jego sposobów, przedmiotu i jego własności, przyczynowości, czasu, przestrzeni, konieczności i możliwości. Chociaż termin "ontologia" pojawił się w literaturze filozoficznej dopiero w XVII wieku, to jej źródła sięgają IV w p.n.e, kiedy to Platon sformułował nową kategorię transcendencji, a Arystoteles zaproponował system uniwersalnych kategorii pod

nazwą "metafizyka", służący klasyfikacji wszystkich istniejących bytów. Termin "ontologia" wywodzi się z greckich słów: *ontos* – byt i *logos* – słowo). Ontologię spopularyzowali w swoich pracach J. Clauberg i Ch. Wolf (XVII), gdzie oznaczał zamiennie z "metafizyką" arystotelesowską teorię bytu [134]. Rozważania na temat ontologii kontynuowali tak sławni filozofowie, jak G. Leibniz, I. Kant czy B. Bolzano , definiując ją jako naukę o rodzajach i strukturach obiektów, ich właściwości, a także zdarzeń, procesów, relacji i dziedzin opisywanej rzeczywistości [135]. Ontologia stawia pytania typu: co stanowi prazasadę i praprzyczynę rzeczywistości? – jak klasyfikować byty, – jakie klasy pojęć są niezbędne do opisu i wnioskowania na temat danego procesu? i inne.

Informatyczne pojęcie ontologii

Wykorzystanie ontologii w informatyce wymuszone zostało rozwojem tzw. technologii semantycznych i koniecznością coraz większej integracji, albo porozumienia na linii człowiek-komputer. Jako opis wybranego wycinka rzeczywistości stało się pojęciowym narzędziem służącym formalnym opisom praktycznej wiedzy i doświadczeń ekspertów, rozumianych przede wszystkim jako najbardziej wiarygodny wykładnik znaczeń i ocen, pozwalający formułować kryteria optymalizacji i szacować dopuszczalne błędy aproksymacji. Rozumienie ontologii w zupełnie nowym, teleinformatycznym kontekście wymagało oczywiście doprecyzowania podstawowych definicji oraz kształtowania zupełnie inaczej rozumianych modeli i metod. Ten proces rozpoczął się w na początku lat dziewięćdziesiątych. Jednak sam "duch" filozoficznej ontologii niewątpliwie przetrwał, w innym kształcie pozwala nam po nowemu opisywać stary świat, szczególnie w takich zastosowaniach jak medycyna.

Na początku, kiedy powstawały pierwsze ontologie [147, 148, 149, 150], wśród przyczyn uzasadniających ich tworzenie wymieniano:

1. konieczność systematyzacji i objaśnienia struktury wiedzy w danej dziedzinie [144];

2. umożliwienie i ułatwienie współdzielenia struktury wiedzy i informacji w danej dziedzinie, zarówno przez ludzi, jak i systemy komputerowe [136, 144, 151];

3. umożliwienie i ułatwienie ponownego użycia wiedzy (*knowledge reuse*) zarówno przez ludzi, jak i systemy komputerowe [136, 144, 151].

Przedstawione poniżej rozważania, dotyczące samego rozumienia pojęcia ontologii w odniesieniu do zmieniających się potrzeb zaczerpnięto przede wszystkim z pracy [201].

Podstawowe definicje Według Neches et al [202] ontologia definiuje podstawową terminologię i relacje opisujące daną dziedzinę, jak również reguły określające jej rozszerzenia. Definicja ta oddaje intuicyjny sens konstruowania ontologii w celu formalizacji wiedzy dziedzinowej, nie podaje jednak żadnych wyróżników służących jej praktycznej realizacji. Według rozważań autorów, ontologię stanowi przede wszystkim słownik opisujący dziedzinę oraz zasady jego konstrukcji. Warto podkreślić, że ontologia obejmuje tutaj nie tylko terminologię *explicite* zawartą w przyjętym modelu wiedzy, ale również to wszystko, co można z niej wydobyć poprzez wnioskowanie.

Klasyczna, najczęściej wykorzystywana definicja ontologii została podana przez T. Grubera w 1993 roku [136, 138]. Stwierdził on, że **ontologia jest jawną specyfikacją warstwy pojęciowej**. Zakładał, że warstwa pojęciowa (*conceptualization*) to abstrakcyjny model zjawisk w ograniczonym wycinku rzeczywistości, uzyskany poprzez identyfikację istotnych pojęć (obiekty, zdarzenia, stany itp.) z nim związanych i relacje pomiędzy nimi. Słowo specyfikacja zaś oznacza, że definicje istotnych w danej dziedzinie pojęć i relacji muszą być precyzyjnie i jednoznacznie sformułowane, przy czym opis ten powinien w pierwszej kolejności uwzględniać ich znaczenie. Oznacza to, że ontologiczny model dziedziny określa strukturę wiedzy w danej dziedzinie, ograniczając możliwe interpretacje zdefiniowanych tam pojęć i relacji. Budowa ontologii jest zawsze związana z konstrukcją słownika zawierającego zbiór formalnych definicji pojęć będących opisem modelowanej dziedziny. Interpretację ontologii jako słownika reprezentującego wiedzę o danej dziedzinie opisano w [144].

Przykładowo, wynikiem bardzo pobieżnej ontologicznej analizy warstwy pojęciowej wybranego obszaru medycyny są takie pojęcia jak: choroba, symptom, diagnoza, rozpoznanie, terapia i relacje pomiędzy nimi, takie jak "choroba wywołuje (określone) symptomy", "terapia leczy (tą) chorobę".

Przymiotnik formalna oznacza, że model musi być czytelny dla maszyny, specyfikacja to wymóg jednoznacznego sformułowania definicji pojęć i relacji, określenie wspólna odnosi się do faktu, że wiedza zawarta w ontologii powinna być akceptowana przez ogół użytkowników.

Konkretnej, Gruber [137] uściślił konceptualizację dziedziny jako (C, I, R, F, A), gdzie: C – zbiór wszystkich pojęć opisujących dziedzinę, I – zbiór obiektów istniejących w dziedzinie, nazywanych też instancjami klas (pojęć), R – zbiór wszystkich relacji definiowanych na C, F – zbiór funkcji zdefiniowanych na C, zwracających jako wynik działania jedno z pojęć należących do modelowanej dziedziny, A — zbiór aksjomatów nakładających ograniczenia na możliwe w modelu znaczenia pojęć, relacji i funkcji.

Borst [139] poszerza definicję Grubera w kierunku jeszcze bardziej użytecznym w zastosowaniach informatyki, określając ontologię jako **formalną specyfikację wspólnej warstwy pojęciowej**. Specyfikacja formalna znaczy tutaj –

czytelna dla maszyny, wykluczająca więc raczej bezpośrednie użycie języka naturalnego. Z kolei wspólna warstwa pojęciowa to taka, która jest akceptowana przez ogół użytkowników, możliwie ustandaryzowana, stanowiąca *consensus* zespołów czy ośrodków, odgrywających dominującą rolę w kształtowaniu wiedzy danego obszaru.

Obok oczywistych zalet takiego rozumienia ontologii, pojawiły się także pewne wątpliwości [141, 142, 143]. Obawy dotyczyły definiowania ontologii z wykorzystaniem pojęcia warstwy pojęciowej, który wywodzi się z epistemologii (inaczej teorii poznania) i dotyczy sposobu spostrzegania świata przez obserwatora, a więc subiektywnie. Modelując dziedzinę należy natomiast dążyć do maksymalnego obiektywizmu. W [143] wskazano także na problem wymogu współdzielenia wiedzy – czy model zbudowany na potrzeby tylko jednej aplikacji, a więc mogący wykorzystywać niekoniecznie powszechnie przyjętą wiedzę, nie ma prawa do nazywania się ontologią?

Pojawiły się więc określania bardziej precyzyjnie i mniej kontrowersyjne, choć wydaje się – mniej użyteczne w opisie złożonych abstrakcyjnie i nie do końca jednoznacznych realnych pojęć i relacji, które odwołują się do formalizmów logicznych. Według nich ontologia to:

- teoria logiczna, która podaje jawną, częściową warstwę pojęciową [141];
- zbiór logicznych aksjomatów, zaprojektowanych w celu wyjaśnienia zamierzonego znaczenia słownika [142];
- hierarchiczna struktura terminów opisujących daną dziedzinę, która może być użyta do budowy bazy wiedzy ją opisującej [145].

Współczesna definicja ontologii podana przez Maedche w 2002 roku [146] także nie korzysta z kontrowersyjnego pojęcia warstwy pojęciowej oddzielając strukturę samej ontologii od struktury opisującego ją leksykonu. Ontologię definiują tutaj dwa zbiory, zbiór O określający strukturę ontologii, oraz zbiór L zawierający strukturę opisującego ją leksykonu. Struktura ontologii definiująca pojęcia i występujące między nimi relacje ma postać $O = \{C, R, Hc, Rel, A\}$ gdzie kolejno: C stanowi zbiór wszystkich pojęć zdefiniowanych w modelu, R jest zbiorem nietaksonomicznych relacji (zwanych właściwościami, slotami lub rolami), definiowanych jako nazwane połączenie między pojęciami, Hc stanowi zbiór taksonomicznych relacji pomiędzy pojęciami, Rel to zdefiniowane nietaksonomiczne relacje pomiędzy pojęciami, a A jest zbiorem aksjomatów.

Struktura leksykonu ma postać $L = \{Lc, Lr, F, G\}$ gdzie: Lc to definicje leksykonu dla zbioru pojęć, Lr oznacza definicje leksykonu dla zbioru relacji, F – referencje dla pojęć, a G to referencje dla relacji.

W definicji Maedche'a ontologię tworzą taksonomia pojęć i semantyczna interpretacja terminów użytych do ich opisu. Dlatego tłumaczenie terminów wystę-

pujących w ontologii z jednego języka narodowego na drugi nie zmienia struktury pojęciowej samej ontologii.

Podsumowując, podstawowym powodem konstrukcji modeli ontologicznych są dziś zastosowania internetowe (sieciowe). Ontologii nie należy utożsamiać z katalogiem czy taksonomią (usystematyzowaniem) obiektów w danej dziedzinie. Ontologia dostarcza przesłanek pozwalających je budować [135, 142]. Ontologia związana jest z obiektem, a nie z jego subiektywnym odbiorem [135, 142]. W ontologii relacje (zależności) między obiektami nie są opisywane funkcyjnie [135]. Istnieje wiele ontologii — nie jest możliwe stworzenie jednej ogólnej ontologii [135]

3.6 Podsumowanie

Warto zwrócić uwagę przede wszystkim na rolę tworzenia dobrego, czyli wiarygodnego i upakowanego (reprezentowanego przez stosunkowo małą liczbę parametrów) modelu danych. Omawiane metody kompresji, indeksowania, a także analizy danych bazują na mniej lub bardziej abstrakcyjnym opisie zasadniczych cech treści, która stanowi najbardziej istotne dla odbiorcy przesłanie przekazu multimedialnego. Połączenie dobrej aproksymacji zasadniczej treści przekazu z doborem odpowiedniej formy zwartej reprezentacji i skutecznym opisem jej kluczowych właściwości stanowi o sukcesie nowych technologii.

Zasygnalizowane metody przetwarzania, segmentacji, ekstrakcji cech czy klasyfikacji dają bogaty arsenał w doskonaleniu przekazu multimedialnego. Ważne jest tutaj zarówno uzyskanie dużej wyrazistości w prezentacji dostarczanej informacji, jak też zautomatyzowanie metod rozumienia treści przekazu w celu selektywnego doboru danych, stanowiących informacje dla konkretnego odbiorcy.

Uzupełnieniem tych metod w coraz większym stopniu stają się narzędzia komputerowej inteligencji. Zagadnienie inteligencji, działanie ludzkiego mózgu, przejawy komputerowej "świadomości", formy optymalizacyjne sztucznej inteligencji, coraz częściej znajdują odzwierciedlenie w użytecznych formach systemów ekspertowych, ewolucyjnych algorytmów optymalizacji czy tez konstrukcjach formalnego opisu wiedzy dziedzinowej. Odwoływanie się do różnych form wnioskowania, wyszukiwania rozwiązań czy dopasowania przybliżeń pełni przede wszystkim rolę wspierającą ambitne ludzkie zadania, dotyczące interpretacji danych, odkrywania treści i jej oceny, czy wreszcie podejmowania ważnych decyzji (np. w diagnostyce medycznej). Dzięki temu multimedia mogą w coraz większym stopniu służyć człowiekowi.

Zadania do tego rozdziału podano na stronie 366.

Ćwiczenie pozwalające na eksperymentalną weryfikację podstawowych metod komputerowego przetwarzania informacji zamieszczono na stronach, odpowiednio: kompresji – str. 382, indeksowania – str. 385, przetwarzania multimediów – str. 388.

Rozdział 4

Użytkowanie informacji, czyli narzędzia

Rozdział ten stanowi opis pomysłów, eksperymentów, narzędzi, systemów, a nawet urządzeń stanowiących konkretną realizację metod użytkowania multimediów. Omawiane są bardziej złożone algorytmy czy koncepcje, jak też odwołania do konkretnych rozwiązań ukazując efekty ich działania. Podane przykłady są w dużym stopniu pragmatycznym rozwinięciem metod sygnalizowanych w rozdziale trzecim, stanowiąc ich aktualne uzupełnienie.

4.1 Kompresja danych

Przedstawiono kilka wybranych metod kodowania zarówno odwracalnego, jak też z selekcją informacji, zwracając szczególną uwagę na różnorodność pomysłów i różne postacie nadmiarowości opisywane w fazie modelowania danych źródłowych, uzupełnione o precyzyjne reguły tworzenia ciągu bitowego reprezentacji kodowej. Bitowy ciąg kodowy, zachowując jednoznaczną dekodowalność, spełnia także, zależnie od zastosowań, szereg dodatkowych kryteriów użyteczności.

4.1.1 Kodowanie Huffmana

Opis opracowanej przez D.A. Huffmana w 1952 roku metody kodowania jest najczęściej cytowaną publikacją z teorii informacji [203]. Przez lata metodę Huffmana optymalizowano ze względu na rosnący stopień różnych form adaptacji [204, 205, 206, 207, 208, 209]. Wykorzystywano ją także w szeregu standardów: dotyczących kodowania faksów (a więc obrazów dwupoziomowych) Grupy 3 [210] i Grupy 4 [211], kompresji obrazów wielopoziomowych JPEG [214], czy też sekwencji obrazów – MPEG-1 [212] i MPEG-2 [213], wykorzystywanych do dziś.

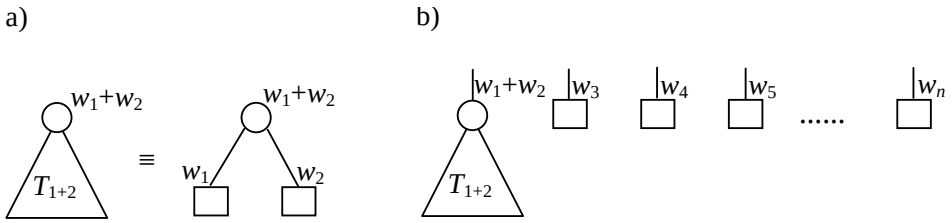
Algorytm

Kod Huffmana jest optymalny w kategorii kodów symboli realizując podstawową zasadę różnicowania długości słów kodowych symboli ze względu na przewidywana częstość ich występowania w kodowanym strumieniu danych. Konstrukcja słów kodowych realizowana jest za pomocą drzewa binarnego, budowanego od liści do korzenia na podstawie rozkładu wag łączonych kolejno liści.

Inicjalizacja algorytmu zakłada ustalenie zbioru wolnych węzłów zewnętrznych – liści z przypisanymi ich symbolami alfabetu oraz wagami. Początkowo, dwa węzły o najmniejszych wagach łączone są ze sobą w elementarne poddrzewo, z wierzchołkiem rodzica o wadze równej sumie wag węzłów łączonych. Następnie wyszukiwane i łączone ze sobą są kolejne wierzchołki o najniższych wagach (wśród pozostałych liści i wierzchołków poddrzew), tworząc kolejne poziomy drzewa, aż do podłączenia wszystkich wolnych węzłów. Algorytm ten zobrazowano na rys. 4.1. Dwa liście symboli o najmniejszych wagach w_1 i w_2 połączono tworząc elementarne poddrzewo T_{1+2}, ustalając przy tym odpowiednią wagę rodzica, reprezentowanego teraz na liście wolnych węzłów. W kolejnych krokach maleje liczba wolnych węzłów, aż do utworzenia pełnej struktury drzewa.

Taka procedura tworzenia drzewa zapewnia, że:

- liść symbolu o najmniejszej wadze ma najdłuższe słowo, leżąc najgłębiej w drzewie na poziomie $m_{\max}$;
- liść symbolu o drugiej w kolejności najmniejszej wadze należy do tego samego, elementarnego poddrzewa, a więc leży również na poziomie $m_{\max}$,



Rysunek 4.1: Przykładowa inicjalizacja metody Huffmana: a) utworzenie elementarnego poddrzewa poprzez łączenie dwóch węzłów o najmniejszych wagach; b) zestaw wolnych węzłów podlegający analizie w kolejnym kroku algorytmu.

gdyż jedynie drzewo lokalnie pełne jest efektywne w kodowaniu.

W ten sposób symbole o wagach mniejszych znajdą się na niższych poziomach (lub najwyżej równych) w stosunku do liści symboli o większych wagach. Konsekwencją będą niekrótsze słowa kodowe, co dowodzi optymalności konstruowanego iteracyjnie według powyższej zasady drzewa kodowego. Kryterium optymalizacji określa postać wyrażenia: $\sum_{a_i \in A_S} w_i |\varsigma_i|$ (oznaczenia jak w p. 2.3.2) , minimalizowana po wszystkich możliwych postaciach drzewa binarnego z ustalonym alfabetem symboli przypisanych liściom zewnętrznym drzewa.

Podsumowując, algorytm kodu Huffmana przedstawia się następująco:

Algorytm 4.1 *Kod Huffmana*

1. Określ wagi wszystkich symboli alfabetu (na podstawie liczby wystąpień w wejściowym ciągu danych); symbole te wraz z wagami przypisz liściom konstruowanej struktury drzewa binarnego, stanowiącym początkowy zbiór tzw. wierzchołków wolnych (tj. wierzchołków, które mogą być łączone w elementarne poddrzewa z węzłem rodzica umieszczonym na wyższym poziomie drzewa).

2. Sortuj listę wierzchołków wolnych w porządku nierosnącym wartości ich wag.

3. Odszukaj dwa wolne wierzchołki z najmniejszymi wagami i połącz je z nowotworzonym węzłem rodzica w elementarne poddrzewo; wagę nowego wierzchołka ustal jako sumę wag dzieci.

4. Usuń z listy wierzchołków wolnych dwa węzły o najmniejszych wagach; wpisz na tę listę nowy wierzchołek rodzica.

5. Przypisz gałęziom prowadzącym od rodzica do węzłów-dzieci etykiety: `0` i `1` (np. lewa gałąź - `0`, prawa - `1`).

6. Powtarzaj kroki 3-5 aż do momentu, gdy na liście wierzchołków wolnych pozostanie tylko jeden węzeł - korzeń drzewa.

7. Odczytaj ze struktury drzewa słowa kodowe kolejnych liści (czyli przypisanych im symboli); słowo stanowią łączone binarne etykiety kolejnych gałęzi odczytane przy przejściu od korzenia do danego liścia (od najbardziej znaczącego bitu gałęzi korzenia do najmniej znaczącego bitu gałęzi dochodzącej do liścia).

□

Proces dekodowania przebiega jak w metodzie S-F (algorytm **??**), jedynie w p. 2 budowane jest drzewo według kodu Huffmana (algorytm 4.1).

Dekodowanie sekwencji kodu symboli S-F odbywa się następująco:

Algorytm 4.2 *Dekodowanie metodą Huffmana*

1. Pobierz ze zbioru danych zakodowanych wartości prawdopodobieństw występowania (wagi) poszczególnych symboli alfabetu.
2. Zbuduj drzewo binarne jak w kodzie Huffmana (według algorytmu 4.1).
3. Dekodowanie kolejnych symboli (prostsza realizacja algorytmu 2.3 dzięki wykorzystaniu struktury drzewa):
   a) ustaw korzeń drzewa jako aktualny węzeł;
   b) pobierz bit z wejścia: jeśli wczytano `0` to przejdź do lewego syna aktualnego węzła, w przeciwnym razie (wpr) przejdź do jego prawego syna;
   c) jeśli aktualny węzeł to węzeł wewnętrzny – kontynuuj (b), wpr odczytaj symbol przypisany liściowi, prześlij go na wyjście i powtarzaj od (a) aż do wyczerpania zbioru danych wejściowych.

□

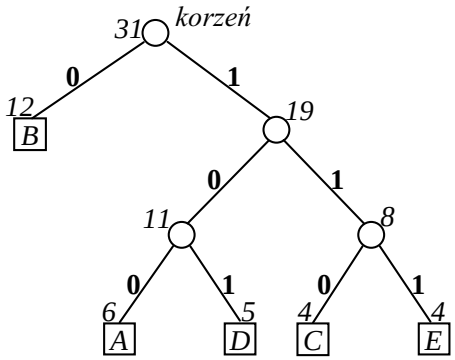
Prześledźmy działanie algorytmu Huffmana na następującym przykładzie.

Przykład 4.1 *Kodowanie metodą Huffmana*
Dany jest źródłowy ciąg symboli o alfabecie: $A_S = \{A, B, C, D, E\}$, przy czym częstość występowania poszczególnych symboli przedstawia tabela 4.1.

Tabela 4.1: Symbole alfabetu przykładowego (przykład 4.1) zbioru danych wraz z częstością wystąpień.

Symbol	A	B	C	D	E
Częstość wystąpień	6	12	4	5	4

Sposób konstrukcji binarnego drzewa kodowego za pomocą algorytmu 4.1 został



Rysunek 4.2: Binarne drzewo kodowe Huffmana z przykładu **??**. Pochyloną czcionką oznaczono wagi poszczególnych wierzchołków.

przedstawiony na rys. 4.2.
Słowa kodowe przyporządkowane poszczególnym symbolom alfabetu przedstawiają się następująco: $A \rightarrow \texttt{100}$, $B \rightarrow \texttt{0}$, $C \rightarrow \texttt{110}$, $D \rightarrow \texttt{101}$, $E \rightarrow \texttt{111}$. Najczęściej występującemu symbolowi B przypisano jednobitowe słowo kodowe, podczas gdy pozostałym – trójbitowe. Długości słów kodowych przypisanych poszczególnym symbolom nie oddają w pełni rozkładu ich wag, co jest spowodowane podstawowym ograniczeniem kategorii kodów symboli. Całkowita liczba bitów każdego ze słów zmusza do zaokrągleń wartości entropii własnej, będącej miarą informacji występującej przy pojawieniu się poszczególnych symboli – zobacz tabelę 4.2. Obliczono w niej efektywność wyznaczonego drzewa Huffmana. Średnia długość słowa kodowego wynosi $\bar{L}_{\text{Huff}} = 2,226$, co jest wartością bliską entropii rozważanego źródła informacji, zakładając model bez pamięci: $H(S_{\text{DMS}}) = 2,176$. □

Tabela 4.2: Efektywność kodu Huffmana dla danych z przykładu 4.1. Zestawiono ilość informacji związaną z wystąpieniem poszczególnych symboli alfabetu z tabeli 4.1 z uzyskaną efektywnością reprezentacji kodowej. W tabeli znajdują się wartości oszacowanych prawdopodobieństw poszczególnych symboli $P(a_i) = N(a_i)/\sum_j N(a_j)$, informacji własnej $I(a_i)$, całkowitej informacji związanej z występowaniem danego symbolu oraz długości słów kodowych i zakodowanej reprezentacji.

Symbol a_i	$P(a_i)$	$I(a_i) = -\lg P(a_i)$ [bity/symbol]	$N(a_i) \cdot I(a_i)$ [bity]	$\lvert\varsigma_i\rvert$ [bity]	$N(a_i) \cdot \lvert\varsigma_i\rvert$ [bity]
A	6/31	2,369	14,215	3	18
B	12/31	1,369	16,431	1	12
C	4/31	2,954	11,817	3	12
D	5/31	2,632	13,161	3	15
E	4/31	2,954	11,817	3	12
Suma	-	-	67,441	-	69

Sposób konstrukcji słów kodowych metodą Huffmana według algorytmu 4.1 nie zależy od kolejności wystąpienia symboli w strumieniu wejściowym, gdyż bazuje na posortowanym zbiorze symboli alfabetu. Jeśli symbole pojawią się seriami: $6 \times A,\ 12 \times B,\ 4 \times C,\ 5 \times D$ i $4 \times E$, to wystarczy zakodować liczbę wystąpień kolejnych symboli na czterech bitach, a sam symbol na trzech (zasada RLE), by uzyskać 35 bitową sekwencję kodową jednoznacznie dekodowalną, znacznie krótszą od 69-bitowej reprezentacji Huffmana. Wskazuje to na poważne ograniczenie efektywności kodów symboli budowanych na bazie źródeł bez pamięci, które nie uwzględniają zależności w występujących bezpośrednio po sobie ciągach wartości. Wadę tę częściowo eliminuje np. adaptacyjna metoda Huffmana, jak też stosowanie modeli źródeł z pamięcią i kodów strumieniowych.

Adaptacyjna wersja kodu Huffmana pozwala na bieżące kształtowanie rozkładów prawdopodobieństw występowania symboli, zależnie od lokalnych statystyk w modelu przyczynowym (z uwzględnieniem jedynie danych już zakodowanych). Drzewo Huffmana jest wtedy stale modyfikowane, zależnie od naliczanych statystyk i aktualizowanych wag poszczególnych symboli, przy czym węzły na poszczególnych poziomach drzewa rozmieszczone są według ustalonego porządku wag (porządek niemalejący od liści do korzenia, zaś na poszczególnych poziomach – np. od lewej do prawej). Po zakodowaniu kolejnego symbolu waga odpowiedniego liścia jest inkrementowana, a drzewo korygowane, jeśli zaburzony został przyjęty porządek (tj. gdy węzeł o wyższej wadze znalazł się za węzłem o wadze niższej – następuje wtedy zamiana). Taka sama procedura adaptacyjnej modyfikacji struktury drzewa Huffmana jest precyzyjnie powtarzana w dekoderze celem wiernej rekonstrukcji sekwencji źródłowej. Adaptacyjna postać algorytmu kodowania jest więc związana z większą złożonością obliczeniową. w typowych implementacjach wzrost ten wynosi 50-100%.

4.1.2 Kod Golomba

Kod Golomba należy do szerszej rodziny kodów przedziałowych, które znalazły zastosowanie w standardach multimedialnych, m.in. w JPEG-LS [89] i H.264 [91]. Wyróżnikiem kodów przedziałowych jest potencjalnie nieograniczony alfabet źródła informacji dzielony jest na przedziały, a słowa kodowe symboli konstruowane są jako konkatenacja (zlepienie) przedrostka wskazującego na przedział z wyróżnikiem konkretnego elementu danego przedziału. W przypadku kodu Golomba jest to zlepienie kodu unarnego z kodem dwójkowym prawie stałej długości.

Kod dwójkowy prawie stałej długości

Aby uzyskać oszczędniejszą reprezentację kodu dwójkowego (niż za pomocą kodu z p.2.3.2) przy liczbie symboli alfabetu nie będącej potęgą dwójki, można zróż-

nicować długości słów poszczególnych symboli alfabetu. Częściej występującym symbolom można przyporządkować jedynie $\lfloor \log_2 n \rfloor$ bitów kodu binarnego prawie stałej długości, pozostałym zaś – $\lceil \log_2 n \rceil$ bitów, zachowując jednoznaczną dekodowalność kodu.

Dokładniej, w kodzie dwójkowym prawie stałej długości $\hat{B}_n$ (dla n-elementowego alfabetu $A_S = \{a_0, \ldots, a_{n-1}\}$) pierwszym r symbolom (o indeksie $0 \leqslant i < r$) przypisywane są słowa $k = \lfloor \log_2 n \rfloor$ bitowe postaci $\hat{B}_n(a_i) = B_k(i)$, a pozostałym - słowa o długości $k+1$ bitów postaci $\hat{B}_n(a_i) = B_{k+1}(r+i)$, gdzie $r = 2^{\lceil \log_2 n \rceil} - n$.

Oczywiście, dla źródeł o alfabecie zawierającym $n = 2^i,\ i = 1, 2, \ldots$ symboli, otrzymujemy
$\hat{B}_n = B_{\log_2 n}$, podczas gdy $r = 0$. Oznacza to, że kod dwójkowy prawie stałej długości przyjmuje postać kodu stałej długości.

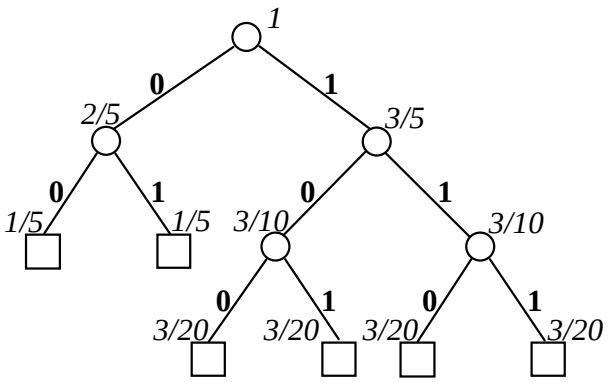
Przykład 4.2 *Kod dwójkowy prawie stałej długości*
Ustalmy postać słów kodu $\hat{B}_n$ dla symboli dwóch alfabetów o odpowiednio $n = 3$ oraz $n = 5$, sprawdzając jednoznaczną dekodowalność takiego kodu.
W pierwszym przypadku mamy: $r = 2^{\lceil \log_2 3 \rceil} - 3 = 1$ oraz długość krótszego słowa $k = \lfloor \log_2 3 \rfloor = 1$. Wtedy pierwszy symbol alfabetu otrzyma jednobitowe słowo kodowe $\varsigma_0 = \hat{B}_3(a_0) = B_1(0) = \texttt{0}$, dwa pozostałe zaś słowa dwubitowe, odpowiednio $\varsigma_1 = \hat{B}_3(a_1) = B_2(1+1) = \texttt{10}$ i $\varsigma_2 = \hat{B}_3(a_2) = B_2(3) = \texttt{11}$. $\hat{B}_3$ jest kodem przedrostkowym, jest więc jednoznacznie dekodowalny.
Analogicznie, dla $n = 5$ otrzymujemy: $r = 2^{\lceil \log_2 5 \rceil} - 5 = 3$ oraz $k = \lfloor \log_2 5 \rfloor = 2$.
Trzy krótsze, dwubitowe słowa kodowe to $\varsigma_0 = \hat{B}_5(a_0) = B_2(0) = \texttt{00}$,
$\varsigma_1 = \hat{B}_5(a_1) = B_2(1) = \texttt{01}$ i $\varsigma_2 = \hat{B}_5(a_2) = B_2(2) = \texttt{10}$, a pozostałe:
$\varsigma_3 = \hat{B}_5(a_3) = B_3(3+3) = \texttt{110}$ i $\varsigma_4 = \hat{B}_5(a_4) = B_3(7) = \texttt{111}$. $\hat{B}_5$ jest także kodem przedrostkowym. □

W kodzie $\hat{B}_n$ słowa kodowe tworzone są jedynie na podstawie pozycji (indeksu) danego symbolu w alfabecie, bez analizy rozkładu wag poszczególnych symboli. Długość tych słów jest minimalnie zróżnicowana, a więc ich efektywne wykorzystanie w algorytmach kodowania jest możliwe szczególnie w przypadku źródeł o stosunkowo równomiernym (tj. zrównoważonym) rozkładzie prawdopodobieństw symboli. Pokazuje to następujący przykład.

Przykład 4.3 *Efektywność kodu dwójkowego*
Wyznaczmy słowa kodu Huffmana dla źródła opisanego zbiorem prawdopodobieństw: $P_S = \{p_0, \ldots, p_5\} = \{1/5,\ 1/5,\ 3/20,\ 3/20,\ 3/20,\ 3/20\}$.
Stosując algorytm 4.1 można skonstruować drzewo Huffmana jak na rys. 4.3. Postać tego drzewa jest *znormalizowana*, tj. liście są uszeregowane w porządku od lewej do prawej według niemalejącej głębokości (każde drzewo Huffmana można znormalizować zamieniając odpowiednio miejscami węzły tego samego poziomu, jeśli relacja maksymalnej głębokości liści ich poddrzew nie jest zgodna z

porządkiem uszeregowania – mechanizm wykorzystywany w algorytmach adaptacyjnych).
Odczytane z drzewa słowa kodowe kolejnych symboli są następujące: $\varsigma_0 = \texttt{00}$, $\varsigma_1 = \texttt{01}$, $\varsigma_2 = \texttt{100}$, $\varsigma_3 = \texttt{101}$, $\varsigma_4 = \texttt{110}$ i $\varsigma_5 = \texttt{111}$. Są one identyczne ze słowami kodu $\hat{B}_6$ pod warunkiem uporządkowania symboli alfabetu zgodnie z nierosnącym ich prawdopodobieństwem. Oznacza to, że w tym przypadku kod dwójkowy prawie stałej długości jest równoważny znormalizowanemu kodowi Huffmana, optymalnemu w kategorii kodów symboli.

□



Rysunek 4.3: Znormalizowane drzewo Huffmana konstruowane dla zrównoważonego rozkładu prawdopodobieństwa symboli alfabetu z przykładu 4.3, z typowym etykietowaniem gałęzi; drzewo to daje słowa kodowe jak w kodzie dwójkowym prawie stałej długości.

Okazuje się, że rozkład prawdopodobieństw symboli źródła z przykładu 4.3 jest zrównoważony, tj. suma dwóch najmniejszych prawdopodobieństw rozkładu jest większa od prawdopodobieństwa największego. Dla takich źródeł, po uporządkowaniu symboli alfabetu zgodnie z nierosnącym ich prawdopodobieństwem, kod dwójkowy prawie stałej długości jest optymalnym kodem symboli.

Kod unarny

Kod unarny wykorzystuje prostą regułą tworzenia kolejnych słów kodowych symboli alfabetu o potencjalnie nieograniczonej długości. Każde ze słów ustalane jest niezależnie, jedynie na podstawie wskazania pozycji symbolu (lub całego przedziału symboli) w hipotetycznym alfabecie.

Według przedrostkowego kodu unarnego, kolejnym symbolom alfabetu przypisywane są słowa postaci: `1`, `01`, `001`, `0001`, ... lub w logice odwróconej: `0`, `10`, `110`, `1110`, ... Definicja kodu unarnego $U(i)$ dla dowolnej, całkowitej $i \geqslant 0$,

będącej elementem alfabetu liczb lub indeksem symbolu alfabetu niesprecyzowanej postaci, jest następująca: $U(i) \triangleq \underbrace{\texttt{0...0}}_{i \text{ razy}}\texttt{1}$ lub $\bar{U}(i) \triangleq \underbrace{\texttt{1...1}}_{i \text{ razy}}\texttt{0}$.

Za pomocą $U(i)$ można szybko określić słowa dowolnie rozszerzanego alfabetu. Kod ten można wykorzystać do efektywnej kompresji strumienia danych z liczbami naturalnymi (z zerem), przy założeniu: $\forall_i$ i jest nie mniej prawdopodobna niż $i-1$. Długość kolejnych słów kodowych ς_i różni się o jeden: $L_i = |\varsigma_i| = L_{i-1} + 1$. Ponieważ w kodzie efektywnym długość słowa powinna być równa (proporcjonalna do) informacji własnej danego symbolu $L_i = -\log_2 p_i$, to kod unarny jest najbardziej efektywny w przypadku, gdy wartości prawdopodobieństw symboli p_i maleją z potęgą dwójki.

Wyobraźmy sobie, że kodowany strumień danych pochodzi ze źródła informacji o alfabecie $A_S = \{0, 1, 2, \ldots\}$, czyli $a_i = i$, zaś rozkład prawdopodobieństw symboli alfabetu wynosi $P_S = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\}$, czyli $p_i = 2^{-(i+1)}$. Słowa kodu unarnego przypisane kolejnym symbolom tego źródła mają długości odpowiednio $L_0 = 1$, $L_1 = 2$, $L_2 = 3, \ldots$, czyli $L_i = i + 1$. Obliczając średnią bitową takiego kodu mamy:

$$R = \sum_{a_i \in A_S} p_i \cdot L_i = \sum_{a_i \in A_S} p_i \cdot (i+1) = \sum_{a_i \in A_S} p_i \cdot (-\log_2 p_i) = H(S)$$

czyli rozkład długości słów dokładnie odpowiada rozkładowi entropii źródła informacji. Kod unarny jest więc w tym przypadku kodem optymalnym.

Zależność na $p_i = 2^{-(i+1)}$, $i = 0, 1, 2, \ldots$ jest szczególnym przypadkiem rozkładu geometrycznego $p_i = (1-\rho)\rho^i$ dla $\rho = \frac{1}{2}$. Kod unarny jest optymalny dla wszystkich rozkładów geometrycznych z $\rho \leqslant \frac{1}{2}$, tj. pozwala uzyskać w tym przypadku rozkład długości słów jak w kodzie Huffmana. Ponieważ przy $\rho < \frac{1}{2}$ przyrost informacji własnej kolejnych symboli jest większy od 1 bit/symbol, to zróżnicowane o jeden bit słowa są również optymalne. Warunkiem koniecznym jest jednak uporządkowanie symboli alfabetu według nierosnących prawdopodobieństw ich występowania.

Jednak dla rozkładów geometrycznych z $\rho > \frac{1}{2}$ informacja własna kolejnych symboli alfabetu narasta wolniej niż 1 bit/symbol, czyli wolniej od wydłużanego o 1 bit słowa kolejnych symboli. Znaczy to, że kod unarny przestaje być wówczas rozwiązaniem optymalnym. Średni przyrost długości słowa kodowego na symbol mniejszy od 1 bita można uzyskać za pomocą kodu Golomba, łączącego kod unarny (z szybszym przyrostem długości słowa o nieograniczonej precyzji) z kodem dwójkowym prawie stałej długości (o znacznie wolniejszym przyroście długości słowa o ograniczonej precyzji). Okazuje się, że dla rozkładów geometrycznych z $\rho > \frac{1}{2}$ kod Golomba z odpowiednio dobranym parametrem (rzędem) jest optymalnym kodem symboli.

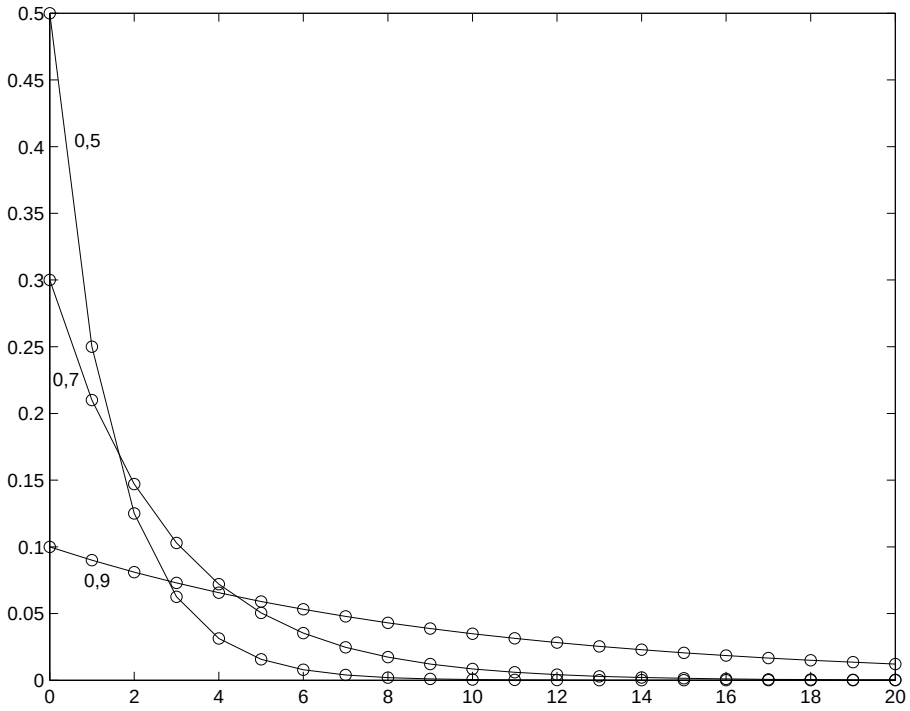
Konstrukcja kodu Golomba

Rząd kodu Golomba $m \in \mathbb{Z}^+$ określa stały rozmiar przedziału przy podziale potencjalnie nieskończonego alfabetu według zasady:

$$A_S = \{\underbrace{a_0, a_1, \ldots, a_{m-1}}_{\text{przedział } 0}, \underbrace{a_m, a_{m+1}, \ldots, a_{2m-1}}_{\text{przedział } 1}, \ldots\} = \{\pi_0^{(m)}, \pi_1^{(m)}, \ldots\} \qquad (4.1)$$

Słowo kodowe kodu Golomba $G_m(a_i)$ symbolu $a_i \in \pi_u^{(m)}$ składa się z wskaźnika tego przedziału $u = \lfloor i/m \rfloor$, zapisanego w kodzie unarnym, oraz względnego miejsca symbolu w przedziale $k = i \mod m$, $k \in \{0, 1, \ldots, m-1\}$. Mamy więc $G_m(a_i) = U(u)\hat{B}_m(a_k)$ lub alternatywnie $\bar{G}_m(a_i) = \bar{U}(u)\hat{B}_m(a_k)$.

Efektywność kodu Golomba warunkowana jest uporządkowaniem realnego alfabetu kodowanych symboli, tak że $p_i \geqslant p_{i+1}$ oraz właściwym doborem wartości m. Ustalenie tego parametru związane jest bezpośrednio z wartością ρ, charakteryzującą rozkład geometryczny (zobacz rys. 4.4), możliwie wiernie aproksymujący rozkład prawdopodobieństw uporządkowanego alfabetu źródła. Szybsze opadanie funkcji rozkładu (mniejsze ρ) wymaga krótszych przedziałów, wolniejsze - dłuższych.



Rysunek 4.4: Rozkład geometryczny (jednostronny) dla wartości ρ: 0,5, 0,7, 0,9.

Wyznaczanie poszczególnych słów kodu Golomba jest proste obliczeniowo. Postać drzew binarnych kodu różnego rzędu ukazuje ich zasadnicze właściwości. Widać to w przykładzie 4.4.

Tabela 4.3: Przykładowe słowa kodu Golomba różnych rzędów dla kolejnych liczb całkowitych $i \geqslant 0$; przy $m = 1$ występują tylko słowa kodu unarnego; | oznacza miejsce sklejenia słowów kodu unarnego oraz dwójkowego prawie stałej długości.

	Kolejne słowa kodu Golomba określonego rzędu						
i	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$
0	0	0|0	0|0	0|00	0|00	0|00	0|00
1	10	0|1	0|10	0|01	0|01	0|01	0|010
2	110	10|0	0|11	0|10	0|10	0|100	0|011
3	1110	10|1	10|0	0|11	0|110	0|101	0|100
4	11110	110|0	10|10	10|00	0|111	0|110	0|101
5	111110	110|1	10|11	10|01	10|00	0|111	0|110
6	1111110	1110|0	110|0	10|10	10|01	10|00	0|111
7	11111110	1110|1	110|10	10|11	10|10	10|01	10|00
8	111111110	11110|0	110|11	110|00	10|110	10|100	10|010
9	1111111110	11110|1	1110|0	110|01	10|111	10|101	10|011
10	11111111110	111110|0	1110|10	110|10	110|00	10|110	10|100
⋮	...	...	...	...	...	...	...

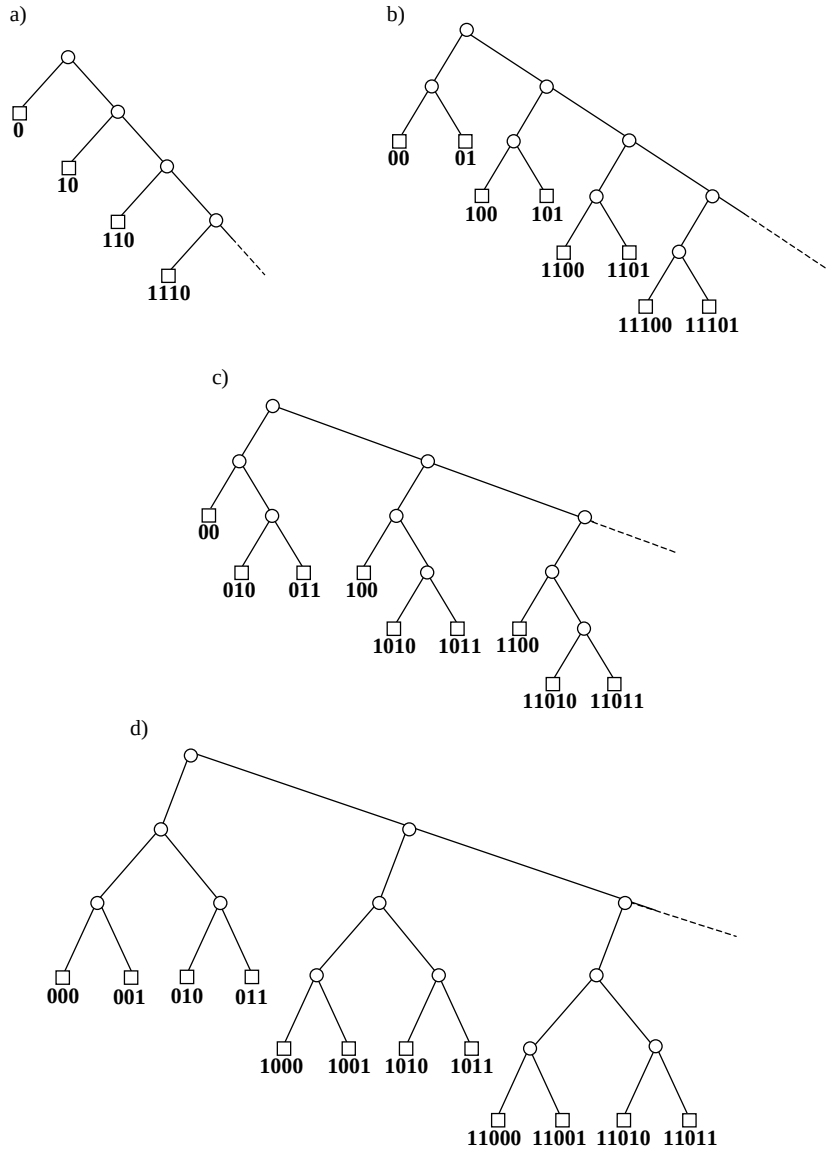
Przykład 4.4 *Kod Golomba*
Wyznaczmy słowo kodowe Golomba rzędu $m = 3$ dla $i = 14$. Ustalając wartość $u = \lfloor 14/3 \rfloor = 4$ uzyskujemy przedrostek $U(4) = \texttt{00001}$, zaś przy
$k = 14 \mod 3 = 2$ musimy wyznaczyć słowo kodu dwójkowego $\hat{B}_3(2)$. Obliczmy więc $r = 2^{\lceil \log_2 3 \rceil} - 3 = 1$, czyli $k \geqslant r$, stąd $\hat{B}_3(2) = B_{\lceil \log_2 3 \rceil}(2+1) = B_2(3) = \texttt{11}$. Poprzez konkatenację otrzymujemy $G_3(a_{13}) = U(4)\hat{B}_3(1) = \texttt{00001 11}$. □

Zestawienie słów kodu Golomba dla kolejnych liczb całkowitych nieujemnych przedstawiono w tab. 4.3. Z kolei binarne drzewa pierwszych słów kodów G_1, G_2, G_3 i G_4 zamieszczono na rys. 4.5. W pierwszym drzewie (dla $m = 1$) od ścieżki *prawych synów* odchodzą tylko liście (jako *lewi synowie* kolejnych węzłów wewnętrznych).

W kolejnych węzłach wewnętrznych ścieżki *prawych synów* podpięte są poddrzewa, początkowo składające się z pojedynczych liści (dla $m = 1$), aż do pełnych poddrzew przy $m = 4$. Przy rosnącym m daje to efekt stopniowego równoważenia drzewa, czyli lepszego dopasowanie do rozkładów geometrycznych o mniejszych nachyleniach (większym ρ).

Kod wykładniczy

Przykładem kodu przedziałowego o zmiennym rozmiarze przedziału jest wykładniczy kod Golomba (*exp_Golomb*), zastosowany w koderze CAVLC (*Context based Adaptive Variable Length Coding*) standardu H.264.



Rysunek 4.5: Wybrane drzewa Golomba dla kodów rzędu: a) $m = 1$, b) $m = 2$, c) $m = 3$, d) $m = 4$.

Słowa kodowe kodu *exp_Golomb* konstruowane są według reguły:

$$G_{exp}(i) = \underbrace{0 \ldots 0}_{m} 1 (i + 1 - 2^m)_{2,m} \tag{4.2}$$

gdzie $m = \lfloor \log_2(i+1) \rfloor$. Przykładowe słowa kodu zebrano w tabeli 4.4.

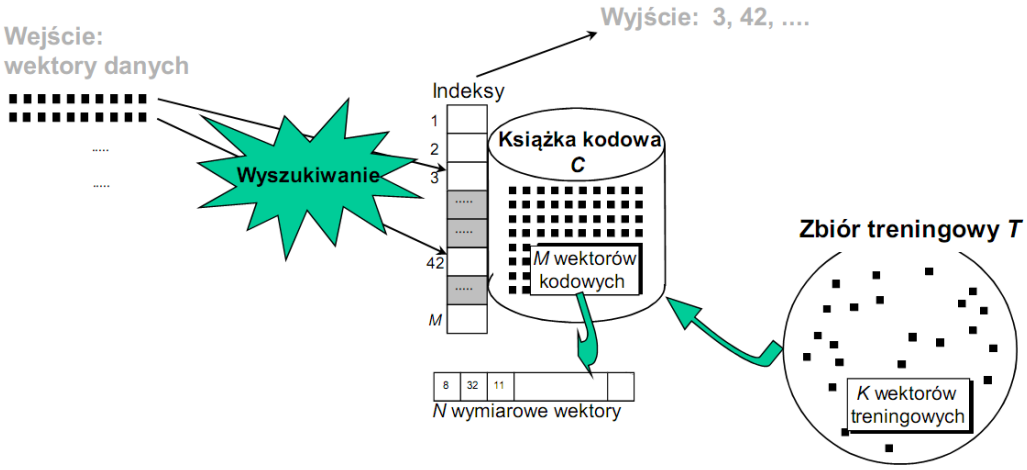
Charakterystyczną cechą tej wersji kodu Golomba jest wykładniczo rosnąca długość przedziału, równa 2^m.

Tabela 4.4: Pierwsze, kolejne słowa wykładniczego kodu przedziałowego.

i	0	1	2	3	4	5	6	7	8	$\cdots$
ς_i	1	010	011	00100	00101	00110	00111	0001000	0001001	$\cdots$

4.1.3 Kwantyzacja wektorowa

Wektorowa kwantyzacja jest uogólnieniem kwantyzacji skalarnej (zobacz p. 3.3.6) na przypadek wielowymiarowych wektorów. Podstawowe zasady konstrukcji kwantyzatora są bardzo podobne, z tym że w miejsce przedziałów kwantyzacji pojawiają się regiony decyzyjne określone w wielowymiarowej przestrzeni wektorów, a zbiór wartości rekonstruowanych zastąpiony jest książką kodów zawierającą wektory rekonstruowane, zwane wektorami kodu. Każdemu wektorowi kodu przyporządkowany jest indeks - kolejne słowo kodowe. Indeksy wskazujące na reprezentantów z książki kodów, przybliżających kolejne wejściowe wektory danych stanowią bitowy strumień nowej, stratnej reprezentacji oryginalnego zbioru danych. Podstawowy schemat metody kompresji bazujący na wektorowej kwantyzacji przedstawiono na rys. 4.6.



Rysunek 4.6: Istota metody kompresji wektorowej kwantyzacji, zamieniającej wektory danych źródłowych na indeksy książki kodów, wskazujące odpowiednie wektory przybliżeń.

Według teorii informacji łączna kwantyzacja sekwencji zmiennych losowych, pomiędzy którymi występuje zależność, jest zawsze lepsza od niezależnej kwantyzacji każdej zmiennej. Tak więc wektorowa kwantyzacja (*vector quantization* - VQ), eliminująca zależności pomiędzy kolejnymi zmiennymi losowymi, jest skuteczniejsza od kwantyzacji skalarnej (*scalar quantization* - SQ). Jedynie w przypadku, gdy te zmienne losowe są niezależne (jeśli można je np. opisać wielo-

wymiarowym rozkładem Gaussa), zastosowanie SQ daje wyniki porównywalne z VQ.

W praktyce zastosowanie VQ w algorytmach kompresji napotyka jednak na szereg problemów. Mała wymiarowość wektorów nie daje satysfakcjonującej skuteczności kompresji. Przykładowo, bloki o rozmiarach 2×2 czy 4×4 formowane z obrazów, stanowią wektory odpowiednio 4-ro i 16-to elementowe o ograniczonej podatności na kwantyzację. Jeśli natomiast użyjemy bloków o większych rozmiarach, np. 8×8 czy 16×16, zwiększając wymiarowość wektorów, a co za tym idzie potencjalną skuteczność algorytmu kwantyzacji, wówczas okazuje się, że zbyt duże rozmiary wektorów stają się z kolei mało przydatne w praktycznych aplikacjach VQ ze względu na trudności realizacyjne. Gubione są bowiem szczegóły informacji wejściowej ze względu na zbyt zgrubne przybliżenia wektorami rekonstrukcji. Wynika to z konieczności ograniczenia rozmiarów książki kodowej, a co za tym idzie liczby wektorów kodu. Zbyt duże książki kodowe wymagają bowiem bardzo czasochłonnych algorytmów przeszukiwań oraz olbrzymich pamięci do ich przechowywania. Powodem ograniczenia wymiarowości wektorów w VQ jest także konieczność określenia wielowymiarowej funkcji gęstości prawdopodobieństwa zmiennej losowej, modelującej wartości wektorów w przestrzeni danych, przy konstrukcji optymalnych wektorów kodu. Wiarygodna estymata takiego rozkładu wymaga dużego zbioru treningowego o własnościach zbliżonych do kompresowanych zbiorów danych. Jeśli zwiększamy wymiarowość przestrzeni wektorów, wówczas przy tej samej ilości zbiorów treningowych maleje rzetelność wyznaczanej coraz bardziej złożonej wielowymiarowej funkcji gęstości (zjawisko analogiczne do rozrzedzania kontekstu bezstratnych koderów z modelami statystycznymi wyższych rzędów). Brak jest poza tym praktycznych rozwiązań pozwalających wyznaczyć optymalny zbiór wektorów kodu w skali globalnej całej przestrzeni wektorów danych (uogólniony na wiele wymiarów algorytm Lloyda-Maxa znajduje minima lokalne). Stosowane algorytmy wektorowej kwantyzacji ograniczają więc zazwyczaj wymiarowość kwantowanych wektorów oryginalnego zbioru danych do wartości nie przekraczających 16, do 20.

Dekompozycja danych oryginalnych

Duża złożoność algorytmów wektorowej kwantyzacji, duże koszty obliczeniowe i sprzętowe (konieczność dużej pamięci operacyjnej do realizacji algorytmów) oraz inne ograniczenia realizacyjne powodują problemy z uzyskaniem zadawalającej skuteczności kompresji w dużym obszarze zastosowań. Podejmowano próby wykorzystania różnych schematów wstępnej dekompozycji danych w celu uzyskania możliwe największej redukcji nadmiarowości przed fazą kwantyzacji, przy stosunkowo niewielkich kosztach obliczeniowych i sprzętowych. Do najbardziej popularnych i skutecznych rozwiązań problemu skutecznej dekompozycji danych należy zaliczyć metody wykorzystujące:

- transformaty częstotliwościowe o nieskończonym nośniku,
- przekształcenia fraktalne,
- przekształcenia falkowe,

Bardzo efektywnym rozwiązaniem okazało się wykonanie wstępnej dekompozycji danych przy pomocy różnego typu transformat, a następnie kwantyzacja współczynników tych transformat przy pomocy prostszych rodzajów kwantyzatorów. Dobre metody dekompozycji danych, redukujące znacznie nadmiarowości, w tym przede wszystkim korelacje pomiędzy wartościami współczynników nowej przestrzeni, pozwoliły uzyskać wyższą efektywność całego schematu kompresji przy znacznie mniejszej złożoności i małych kosztach realizacyjnych. Okazuje się, że również w tym przypadku najlepsza postać transformaty w schemacie kodowania transformacyjnego, pozwalająca całkowicie zdekorelować wejściowe źródło danych jest bardzo trudna w praktycznej realizacji ze względu na olbrzymie koszty obliczeniowe. Jądro tego przekształcenia jest bowiem zależne od sygnału transformowanego. Można jednak osiągnąć zbliżone poziom dekorelacji danych przy pomocy prostszej, niezależnej od sygnału transformaty, która pozwala zrealizować bardzo szybkie algorytmy kompresji.

Nieco inne rozwiązanie, bliższe koncepcji wektorowej kwantyzacji zastosowano w koderach fraktalnych, używanych zasadniczo do stratnej kompresji obrazów. Wykorzystując aparat przekształceń afinicznych do generacji operatorów zwężających z atraktorami w postaci fraktali przybliżających kolejne porcje obrazu, zbudowano algorytm kompresji oparty na idei samopodobieństwa obrazu. Wyznaczanie dużej księgi kodowej na podstawie licznych wektorów treningowych zastępowane jest procesem definiowania słownika na bazie wzajemnego podobieństwa małych fragmentów kompresowanego obrazu - tak jakby obraz był przybliżany na podstawie samego siebie. Po zdefiniowaniu wielu operatorów zwężających opisujących obraz można go następnie wygenerować w kilku lub co najwyżej kilkunastu iteracjach, zaczynając od dowolnej postaci pola obrazu.

Jeszcze inna koncepcja wywodzi się z teorii aproksymacji. Wiadomo, że dla oszczędniejszego opisu sygnału za pomocą jądra liniowej transformaty potrzeba możliwie dużego podobieństwa sygnału i funkcji bazowych tego przekształcenia. Skoro kompresowane dane reprezentują w zdecydowanej większości przypadków źródła silnie niestacjonarne, należy zastosować dobre narzędzie do opisu sygnałów niestacjonarnych. Transformacja Fouriera lub jej podobne (sinusowa, kosinusowa) wykorzystują funkcje sinusoidalne i kosinusoidalne o nieskończonym nośniku przy założeniach stacjonarności (lub pseudostacjonarności) sygnału, a więc nie są dobrym narzędziem w przypadkach sygnałów niestacjonarnych. Wykorzystano więc narzędzie przekształcenia liniowego falek o funkcjach bazowych określonych na skończonym nośniku, o skończonej energii, świetnie nadające się do analizy sygnałów niestacjonarnych tworząc bardzo oszczędną ich reprezentację.

Kodowanie odwracalne

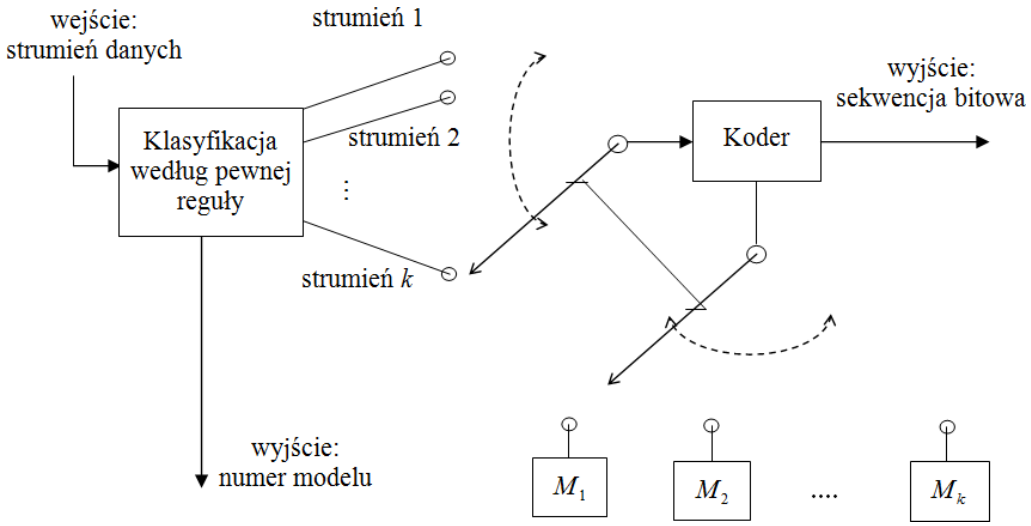
Sposób bezstratnego kodowania powstałych po kwantyzacji zbiorów wartości o zredukowanej dynamice (alfabecie) bardzo silnie zależy od realizowanych wcześniej technik dekompozycji i samej kwantyzacji. Trudno więc tutaj mówić o rozwiązaniach optymalnych w każdym przypadku. Zagadnienie to jest analogiczne do rozważanych w pierwszej części tej pozycji problemów bezstratnej kompresji danych. Wyróżnikiem może być jednak duża wiedza a priori o strumieniach danych wchodzących na wejście kodera ze względu na ukształtowanie tego strumienia przez poprzednie etapy schematu kompresji.

Można wyróżnić dwa podstawowe rodzaje tychże schematów. W pierwszym optymalizuje się maksymalnie funkcje dekorelacyjne w procesach dekompozycji i kwantyzacji, wyrzucając do kodowania zbiory danych prawie zupełnie zdekorelowanych czy wręcz niezależnych, o silnie zróżnicowanym alfabecie danych, zmiennej liczbie bitów, itd. W takich przypadkach rola odwracalnego kodowania jest znikoma lub wręcz żadna. Zadanie redukcji nadmiarowości jest w zadawalającym stopniu wykonywane wcześniej, np. w metodzie kwantyzacji, która jest optymalizowana pod kątem minimalnej entropii danych wyjściowych w relacji do poziomu błędu kwantyzacji, a faza kodowania może być pominięta.

Innym rozwiązaniem jest zmniejszenie czasochłonności i stopnia złożoności algorytmów kwantyzacji poprzez zastąpienie ich rozwiązaniami znacznie prostszymi, bardziej efektywnymi w stosunku do stopnia złożoności. Dają one co prawda mniejszą redukcję nadmiarowości, ale dodatkowa dekorelacja danych może być wykonana skutecznie w koderze odwracalnym przy znacznie niższych ostatecznych kosztach czasowych. Przykładowo, znając statystyką strumienia danych kwantowanych można zaprojektować efektywny statyczny koder arytmetyczny czy Huffmana o mniej rozbudowanych strukturach i lepszych warunkach początkowych.

Często spotykaną praktyką jest podział zbioru kodowanego na kilka podzbiorów (strumieni) o wyraźnie różnej statystyce i niezależne kodowanie każdego z nich koderem o innych parametrach modelu źródła informacji czy wręcz strukturze, bądź też jednym koderem o modelach przełączanych (rys. 4.7).

Dodatkowo w zależności od zastosowań generowany jest w koderze strumień o cechach kodu sekwencyjnego lub progresywnego, ustalający pewną hierarchię przekazywanej informacji, w niektórych rozwiązaniach również kod zagnieżdżony, kiedy to koder jest zatrzymywany w momencie osiągnięcia założonej długości kodu wyjściowego. Czasami też fazy kwantyzacji i kodowania przeplatają się emitując równolegle ze złożoną procedurą kwantyzacji kolejne partie kwantowanych wartości np. w szybkich zastosowaniach transmisyjnych.



Rysunek 4.7: Schemat kodera VQ o kilku modelach książek $M_1, M_2, \ldots, M_k$, przełączanych zależnie od charakterystyki strumienia wejściowego.

4.2 Wyszukiwanie informacji

Zamieszczone tutaj rozważania odnoszą się do podstaw teorii indeksowania i wyszukiwania informacji zakreślonych w punkcie 2.3.4.

4.2.1 Przegląd narzędzi służących wyszukiwaniu

Konstruowane na podstawie indeksów narzędzia do wyszukiwania treści obrazowej wykorzystują różnorakie struktury słownikowe do konstrukcji indeksu, list obiektowych czy też mechanizmów dostępu do danych fizycznych przeszukiwanych obiektów multimedialnych. Poniżej krótko scharakteryzowano wybrane wyszukiwarki (zebrane przez P. Bonińskiego w [33]).

W pierwszej kolejności warto wymienić rozwiązania komercyjne, o szerokiej skali zastosowań.

QBIC

QBIC (*Query By Image Content*) [124] jest pierwszym komercyjnym systemem indeksowania danych obrazowych ich zawartością, wprowadzonym na rynek przez firmę IBM. Mimo swojej prostoty, jest to jeden z bardziej znanych systemów typu CBIR, który z pewnością wpłynął na kształt późniejszych rozwiązań w tej dziedzinie.

QBIC obsługuje dużą gamę form zapytań: przykładowego obrazu, zakreślonego konturu, rozkładu kolorów bądź wybranego fragmentu tekstury. Wykorzystywane przez system indeksy są relatywnie proste i obejmują podstawowe cechy histogramowe, zmodyfikowane teksturowe cechy Tamury [118] jako kombinacja skrośności, kontrastu i kierunkowości [125] oraz cechy kształtu, w tym cyrkularność, mimośród, orientacja głównych osi i grupa momentów inwariantnych[126].

Efektywność systemu *QBIC* nie jest wysoka, tym niemniej znalazł on zastosowania komercyjne, np. był wykorzystywany przez słynne rosyjskie muzeum Ermitaż w Sankt Petersburgu. Funkcjonalność systemu *QBIC* jest obecnie dostępna w komercyjnym systemie zarządzania bazą danych IBM DB2 w ramach rozszerzenia oferowanego pod nazwą *DB2 Image Extender*.

Virage

System *Virage* [127] jest kolejnym, szeroko znanym systemem typu CBIR, z zastosowaniami komercyjnymi. Obsługuje różne formy zapytań: przykładowy obraz, kontur, tekstura, rozkład kolorystyczny. *Virage* umożliwia modyfikację parametrów indeksowania poprzez przypisanie wag każdej wyznaczanej cesze. Koncepcja ta została po raz pierwszy zastosowana w praktycznym systemie właśnie w *Virage* i była często wykorzystywana w rozwiązaniach późniejszych (np. w [128]).

System *Virage* znalazł zastosowanie w wielu instytucjach i firmach, choćby w stacji telewizyjnej CNN, w dużej mierze dzięki integracji z bazą danych firmy Sybase™oraz Oracle™. Swego czasu *Virage* był też stosowany jako silnik wyszukujący w usłudze *Altavista PhotoFinder*, co jest dowodem skuteczności tego rozwiązania.

Natomiast zdecydowana większość prac nad systemami indeksowania zawartością prowadzona jest na uniwersytetach i innych środowiskach badawczych, dlatego tutaj właśnie można znaleźć największe bogactwo koncepcji i rozwiązań w obszarze systemów wyszukiwania. Poniżej wymieniono kilka przykładowych rozwiązań.

SCORE

SCORE (*System for COntent based REtrieval of pictures*) [129] proponuje specjalny model reprezentacji zawartości obrazów. Każdy z obrazów w bazie danych jest opisany przez zmodyfikowany diagram encji reprezentujących obiekt oraz związków. Encje nie oznaczają tu jednak typów, ale konkretne obiekty. Podobnie symbol związku dotyczy jednego konkretnego powiązania, a nie zbioru powiązań. Wyróżnione są dwa typy związków: akcje opisują pewne sytuacje rozpoznawane w obrazie (np. pies goni kota), zaś relacje przestrzenne określają względne pozycje występujących obiektów (na lewo, pod spodem, z przodu).

Tworzenie zapytania polega na graficznym wyborze kilku obiektów z palety ikon. Następnie użytkownik określa dodatkowe parametry i atrybuty obiektów (kolor, rozmiar, liczba) oraz definiuje żądane związki między obiektami. Wykonanie zapytania uwzględnia przybliżone dopasowywanie wartości atrybutów oraz akcji (np. sosna jest drzewem) i reguły dedukcji dla związków przestrzennych (np. przechodniość relacji na lewo).

Photobook

Jednym z najbardziej znanych systemów jest opracowany w MIT Media Lab system *Photobook* [130]. Rozwiązanie to jest interesujące ze względu na próbę połączenia koncepcji w pełni automatycznego indeksowania obrazów zawartością z adnotacjami dodawanymi przez użytkownika, wspierającymi proces wyszukiwania. Wykorzystywane są cechy kształtu i tekstury, posiada też zestaw cech dedykowany rozpoznawaniu twarzy (jest to efekt współpracy z firmą Viisage Technology, która zaowocowała systemem FaceID, wykorzystywanym przez policję w Stanach Zjednoczonych).

SMDS

SMDS [131] jest jedną z pierwszych prób opracowania podstaw technologii multimedialnych. Formalnie zdefiniowano instancję medium, która reprezentuje jeden konkretny typ mediów, np. audio, wideo, obrazy, dokumenty. Instancja medium

zawiera w sobie poszczególne egzemplarze danego typu (np. ścieżka audio) oraz cechy opisujące zawartość tych egzemplarzy. Określono również formalnie strukturę bazy danych, która oprócz instancji mediów obejmuje również elementy pozwalające na osłabianie treści zapytań, np. hierarchię generalizacji cech (mustang jest przykładem forda) lub dopuszczenie substytutów wartości atrybutów (kolor żółty można zastąpić pomarańczowym). Zaproponowane definicje są na tyle ścisłe, że pozwalają utworzyć język zapytań w formie operacji logicznych (podobny do języka PROLOG). Wykorzystano także język bazujący na składni SQL, który na niższym poziomie wykorzystuje odpowiednie formuły logiczne. Poniższe zapytanie w języku SMDS-SQL znajduje obrazy, na których dostrzegalny jest biały ford:

```
SELECT M
FROM smds source1 M
WHERE FindType(M)=image
AND FindObjWithFeature(ford)
AND Color(ford, white, S)
```

SEMCOG

Język CSQL [132] jest częścią systemu zarządzania bazą danych *SEMCOG*, który służy do przechowywania obrazów statycznych. System wprowadza hierarchiczną strukturę modelowania obrazów, która wspiera zarówno zapytania na poziomie całych obrazów, jak też poszczególnych obiektów. Dodatkowo uwzględniono dwojaki charakter przechowywanych obrazów – ich cechy wizualne i semantyczne. Uwzględnienie cech wizualnych pozwala np. na tworzenie zapytań o podobieństwo dwóch obrazów (dotyczące kształtów, kolorów, rozmiarów), natomiast cechy semantyczne umożliwiają wyszukiwanie obrazów na podstawie opisu treści obrazowej, który jest definiowany ręcznie przez użytkownika lub półautomatycznie przez algorytmy przetwarzania obrazów.

Model zakłada, że obraz jest obiektem złożonym z wielu obiektów składowych, które mają określoną semantykę i cechy wizualne (np. człowiek, samochód). Struktura każdego obiektu obejmuje więc jego obraz (zbiór pikseli), cechy semantyczne i relacje przestrzenne. Obiekty mogą również zawierać kolejne podobiekty i relacje przestrzenne między nimi. W zapytaniach można specyfikować kryteria selekcji, odwołujące się do semantyki obrazów, wizualnego podobieństwa obrazów oraz relacji przestrzennych, określonych w modelu obrazu. Poniższe, przykładowe zapytanie wyszukuje wszystkie obrazy, na których widoczna jest osoba na prawo od obiektu podobnego do zadanego obrazu obiektu.

```
SELECT image P
WHERE P contains X
AND P contains Y
AND X is_a człowiek
AND Y i_like
AND X to_the_right_of Y
```

Język *MOQL* (Multimedia Object Query Language)

MOQL [133] jest rozszerzeniem języka OQL opracowanego dla obiektowych baz danych. Celem twórców była reprezentacja dowolnych danych multimedialnych. Sposób przechowywania nie został konkretnie określony, natomiast założono, że oprócz samych mediów, dostępne będą również informacje semantyczne dotyczące interesujących obiektów wchodzących w ich skład.

Rozszerzenia OQL dotyczą przede wszystkim nowych wyrażeń, jakie można stosować w ramach klauzuli WHERE. Są to m.in. predykaty i funkcje przestrzenne — np. *intersect*, predykaty i funkcje temporalne – np. *overlap*, predykat CONTAINS. Składnia zapytań została również poszerzona o specjalną klauzulę PRESENT, która daje szerokie możliwości definiowania sposobu prezentacji wyników.

Przedstawione poniżej, przykładowe zapytanie wyszukuje pary (obraz, wideo), gdzie wszystkie samochody widoczne na obrazku mają być wyszukane z wideoklipu. Obraz i wideo są prezentowane w oknach o określonej pozycji i rozmiarze. Pokaz obrazu trwa 20 sekund i rozpoczyna się 10 sekund przed początkiem klipu wideo, który jest odtwarzany przez 30 minut.

```
SELECT m, v
FROM Images m, Videos v
WHERE FOR ALL c IN (SELECT r FROM Cars r WHERE m CONTAINS r)
v CONTAINS c
PRESENT atWindow(m, (0,0), (300, 400))
AND atWindow(v, (301, 401), (500, 700))
AND play(v, 10, normal, 30*60) parStart display(m, 0, 20)
```

ASSERT

ASSERT [64] jest jednym z nielicznych przykładów systemu CBIR do zastosowań medycznych. Jest on dedykowany indeksowaniu obrazów CT (*Computed Tomography*) płuc ze wsparciem detekcji symptomów rozedmy płuc. Do oceny charakteru i rodzaju zmian wykorzystano deskryptory cech teksturowych.

GIFT/medGIFT

Bardzo interesującym i dostępnym w ramach fundacji GNU jest GNU Image Finding Tool (GIFT), opracowany na uniwersytecie w Genewie system typu CBIR. Do komunikacji z klientem wykorzystuje on bazujący na XML protokół MRML (*Multimedia Retrieval Markup Language*)[1], dający szereg możliwości, takich jak: uzyskanie informacji o kolekcjach obrazów na serwerze, dostępnych algorytmach, ich parametrach, itp. Ustandaryzowany sposób komunikacji teoretycznie umożliwia uniezależnienie klienta od konkretnego rodzaju serwera.

Z pewnością GIFT jest interesującą, rozwojową platformą do budowy własnych rozwiązań typu CBIR, szczególnie, że projekt dostępny jest w ramach licencji GNU GPL daje dostęp do pełnych kodów źródłowych.

FIRE

Flexible Image Retrieval Engine (*FIRE*) [120] jest projektem związanym z projektem IRMA, mającym na celu realizację systemu typu CBIR o uniwersalnym przeznaczeniu, z architekturą umożliwiającą łatwą rozbudowę systemu o nowe cechy, metryki i struktury danych.

[1] http:/www.mrml.org

4.3 Podsumowanie

Rozdział ten stanowi pewnego rodzaju amalgamat doświadczeń w obszarze multimediów. Jest wybraną, zmieniającą się z założenia kolekcją metod, przykładowych rozwiązań, opisów eksperymentów, wniosków z obserwacji itp. Jest to najbardziej otwarta część tego podręcznika, aktualizowana w jego cyfrowej wersji.

Zadania do tego rozdziału podano na stronie 370.

Rozdział 5

Pragmatyzm multimediów

Wokół rozwijanych standardów multimedialnych koncentrowały się wysiłki wielu badaczy, teoretyków i praktyków, próbujący sprostać bieżącym wyzwaniom współczesności. Silny aspekt pragmatyczny, a jednocześnie otwarty i twórczy tej pracy, narodziny nowych technologii przekładanych prawie natychmiast na usługi użyteczne w szerokiej, niemal globalnej skali odbiorców, zasługuje na szczególną uwagę.

Najstarszymi (1980 rok) i szeroko stosowanymi obecnie standardami kompresji obrazów cyfrowych są międzynarodowe standardy kodowania cyfrowych faksów, odpisów (*facsimile*) Grupy 3 i Grupy 4 opracowane przez grupę konsultacyjną CCITT (*Consultative Committee of the International Telephone and Telegraph*). Standardy te dotyczą jedynie binarnych obrazów zawierających teksty i dokumenty. Wzrost liczby zastosowań wielopoziomowych obrazów cyfrowych w szerokiej gamie zastosowań nieuchronnie prowadzi do opracowywania nowych standardów kompresji. Udogodnienia związane z wprowadzeniem tych standardów dotyczą nie tylko łatwiejszej wymiany obrazów pomiędzy różnymi systemami i aplikacjami, lecz także pozwalają na znaczące ograniczenie kosztu budowy wyspecjalizowanych urządzeń cyfrowych niezbędnych w wielu systemach kompresji obrazów w czasie rzeczywistym. W ostatnich latach prace nad nowymi standardami kompresji obrazów prowadzone były w trzech zasadniczych kierunkach:

- obrazy binarne
- pojedyncze obrazy wielopoziomowe, monochromatyczne i kolorowe
- obrazy ruchome (sekwencje wizyjne, wideo)

W 1988 roku został uformowany komitet znany jako JBIG (*Joint Bilevel Imaging Group*) pod auspicjami ISO-IEC/JTC1/SC2/WG8 i CCITT SG VIII NIC,

w celu opracowania standardu kompresji i dekompresji obrazów binarnych. Grupa skoncentrowała swoje wysiłki na poszukiwaniu efektywniejszego algorytmu od opracowanych wcześniej przez CCITT, w zastosowaniu do klasycznych aplikacji (n.p. ośmiu binarnych obrazów odniesienia zaproponowanych przez CCITT), a także rozszerzenia ich stosowalności do nowych aplikacji. Chodziło głównie o opracowanie algorytmów progresywnych i adaptacyjnych.

Należy też wspomnieć o opracowanej przez naukowców z IBM w 1988 roku binarnej wersji kodera arytmetycznego o nazwie Q-koder i zastosowaniu go do kompresji obrazów binarnych w następujących opracowaniach:

- technika ABIC (*Arithmetic Binary Image Compression*).
- CCITT Group 3 i 4 - kodowanie długości sekwencji do kompresji obrazów binarnych
- JBIG/JBIG2 (*Joint Bilevel Imaging Group*) - kodowanie arytmetyczne do kompresji danych binarnych

Komitet znany powszechnie pod nazwą JPEG (*Joint Photographic Experts Group*), działający jako ISO-IEC/JTC1/SC2/WG10 przy bliskiej nieformalnej współpracy z CCITT SG VIII NIC, uformowany został pod koniec 1986 roku w celu opracowania międzynarodowego standardu dla pojedynczych, wielopoziomowych, monochromatycznych i kolorowych obrazów. Zadaniem zespołu było zdefiniowanie standardu dla tak różnych zastosowań jak foto- i telegazeta, grafika komputerowa, skład komputerowy, mała poligrafia, kolorowe faksy, systemy medyczne i wiele innych. Pomimo tego, iż w tej dziedzinie nie istniały wcześniejsze standardy, członkowie JPEG byli silnie przekonani, że wymagania zdecydowanej większości tych zastosowań winny być uwzględnione w standardzie. Ostateczna propozycja standardu, która została opublikowana w 1992/1993 r. jako standard międzynarodowy ISO/IEC zawiera trzy główne składniki: 1) system podstawowy, który zawiera prosty i efektywny algorytm, adekwatny w stosunku do większości zastosowań kompresji obrazów, 2) zbiór rozszerzeń systemu zawierający przede wszystkim algorytm progresywnego kodowania rozszerzający pole zastosowań, 3) niezależna bezstratna metoda kodowania dla zastosowań wymagających tego typu kompresji. Bardziej dokładne omówienie tego standardu, ze względu na przewidywane przez twórców jego zastosowanie także w systemach obrazowania medycznego, zostanie przedstawione w następnej części tego podrozdziału.

Rozwój nowych metod i technik multimedialnych w dużym stopniu odbywa się przy okazji prac nad nowymi standardami. Tak było z kompresją obrazów doskonaloną w ramach standardów JPEG[1] (normalizacyjna grupa robocza ISO/IEC SC 29/WG 1 - *Joint Picture Expert Group*) oraz kompresją sekwencji wizyjnych

[1] http://www.jpeg.org.

i dźwięku, czy też opisem multimediów w rodzinie standardów MPEG[2] (MPEG - normalizacyjna grupa robocza ISO/IEC SC 29/WG 11 - *Moving Picture Expert Group*). Nie tylko sankcjonowano standardem uznane powszechnie rozwiązania istotnych technik multimedialnych, ale ogłaszano konkursy na nowe algorytmy, metody, opracowania, by rozwiązać bieżące i przyszłe problemy świata multimediów w określonych uwarunkowaniach (np. kolejne części JPEG2000, czy rozwój technik promowanych przez MPEG), indeksowania treści multimedialnej z wykorzystaniem doskonalszych deskryptorów (MPEG-7), stworzyć zintegrowaną platformę multimedialną (MPEG-21), itd. Wręcz aranżowano nowe prace badawcze w dużych zespołach, by sprostać wyzwaniom dynamicznego rozwoju technologicznego według perspektyw umiejętnie zarysowanych w kolejnych wezwaniach (*call*) grup normalizacyjnych. Standardy multimedialne mają więc charakter silnie innowacyjny.

Normy JPEG i MPEG mają charakter ogólny (generyczny). Wynika to ze specyfiki multimediów, które mają charakter powszechny, uniwersalny, o szerokim spektrum zastosowań w różnorodnych produktach komercyjnych, przemysłowych, teleinformatycznych, elektronicznych. Takie podejście wymaga definiowania profili aplikacyjnych i poziomów zgodności, a niekiedy zestawów narzędziowych dostosowujących ogólne ramy standardu do wiodących, konkretnych zastosowań.

[2] http://www.mpeg.org.

5.1 Standardy rodziny JPEG

Chociaż zestawem standardów w pełni zasługujących na miano multimedialnych jest efekt wieloletniej pracy grupy MPEG, to jednak nie sposób nie wspomnieć o równoległych działaniach grupy JPEG. Prace te w jakimś sensie były komplementarne do MPEG, multimedialne w nieco węższym sensie, przede wszystkim w różnorodnym reprezentowaniu i opisie treści obrazowej, będącej niewątpliwie najistotniejszym składnikiem przekazywanych strumieni informacji. Zasadniczym przedmiotem prac standaryzacyjnych grupy JPEG jest obraz pojedynczy, tzw. statyczny, ale także sekwencja obrazów. Przekaz informacji obrazowej rozpatrywany jest w różnorodnych kontekstach aplikacyjnych.

Zasadniczy nurt opracowań tej rodziny standardów dotyczy zagadnienia efektywnej kompresji obrazów pojedynczych o trzech i większej liczbie komponentów, ograniczonej w coraz mniejszym stopniu dynamice wartości komponentów oraz stale rosnącej rozdzielczości kodowanych zobrazowań. Zakres możliwych do uzyskania stopni kompresji jest bardzo szeroki – od odwracalnych kodeków (z rekonstrukcją obrazów źródłowych z dokładnością do pojedynczego bitu) w niewielkim stopniu (typowo od 2:1 do 3:1) redukujących rozmiary plików, aż po możliwość dowolnej redukcji rozmiaru kodowanego obrazu z możliwością kontroli poziomu zachowanej jakości danych źródłowych, z selekcją informacji zachowanej w skompresowanej reprezentacji danych.

5.1.1 JPEG, czyli najpopularniejszy kodek obrazów

Specyfikacja normy standardu JPEG [214] zawiera:

- opis procesu przetwarzania źródłowych danych obrazowych w dane obrazowe skompresowane;
- opis procesu przetwarzania skompresowanych danych obrazowych w zrekonstruowane dane obrazu;
- wskazania dotyczące praktycznych implementacji standardu;
- opis zakodowanej reprezentacji skompresowanych danych obrazowych.

Specyfikacja nie opisuje kompletnej zakodowanej reprezentacji obrazu, może ona zawierać pewne parametry zależne od aplikacji. W normie wyszczególniono cztery procedury (tryby) kompresji, a mianowicie:

1. podstawowa (*baseline process*).
2. rozszerzona na bazie DCT(*extended DCT- based process*).
3. bezstratna (*lossless process*).

4. hierarchiczna (*hierarchical process*).

Początkowo źródłowe obrazy kolorowe z przestrzeni barw RGB, zwykle stosowanej przy rejestracji obrazów, są konwertowane do przestrzeni barw YCrCb w celu efektywniejszej kompresji, zgodnie z zależnością:

$$\begin{aligned} Y &= 0,2989 \cdot R + 0,5866 \cdot G + 0,1145 \cdot B \\ Cr &= 0,5 \cdot R - 0,4183 \cdot G - 0,0816 \cdot B \\ Cb &= -0,1687 \cdot R - 0,3312 \cdot G + 0,5 \cdot B \end{aligned} \tag{5.1}$$

Wydziela się w ten sposób składową luminancji (Y), na którą przede wszystkim wrażliwe jest ludzkie oko podczas percepcji treści obrazowej. Dalej każdy komponent kodowany jest niezależnie, przy uwzględnieniu pewnych różnic związanych z mniejszej czułością percepcji składowych chrominancji.

Kodowanie każdego składnika obrazu (luminancja, składowe chrominancji) przebiega analogicznie. W podstawowym procesie kodowania dane wejściowe są ośmiobitowe, w rozszerzonym - 12 bitowe. Obraz jest dzielony na bloki 8×8 i każdy blok jest transformowany za pomocą DCT (*discrete cosine transform*), przy czym kodowanie obrazu przebiega sekwencyjnie, tzn. z lewej strony na prawą zaczynając od góry obrazu i przemieszczając się na dół.

W standardzie JPEG wykorzystano zbiór funkcji bazowych transformacji DCT – 2W dyskretna transformacja kosinusowa funkcji obrazu $f(x, y)$ w bloku o rozmiarach $N \times N$ zdefiniowana jest w sposób następujący:

$$k(u,v) = \frac{1}{\sqrt{2N}} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{\pi(2x+1)u}{2N} \cos \frac{\pi(2y+1)v}{2N} \tag{5.2}$$

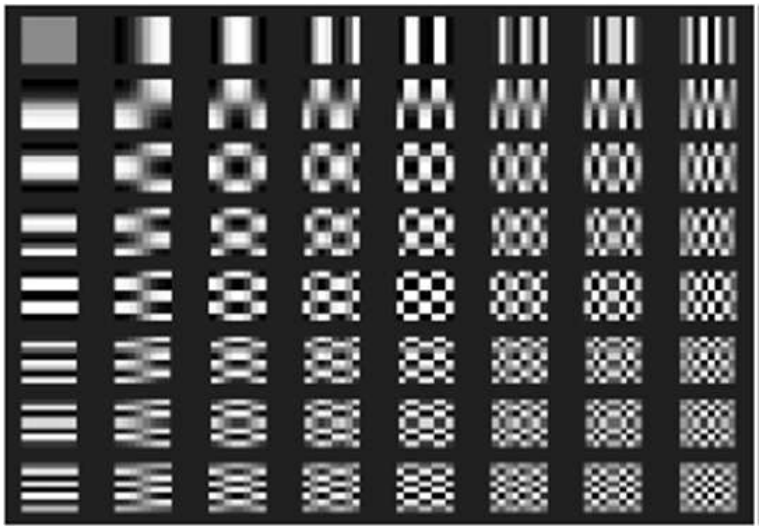
jako przekształcenie proste, a odwrotne:

$$f(x,y) = \frac{1}{\sqrt{2N}} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)k(u,v) \cos \frac{\pi(2x+1)u}{2N} \cos \frac{\pi(2y+1)v}{2N} \tag{5.3}$$

gdzie $k(u, v)$ są wartościami współczynników w dziedzinie DCT, $C(u), C(v) = \frac{1}{\sqrt{2}}$ dla $u, v = 0$ oraz $C(u), C(v) = 1$ w p.p.

W JPEG przyjęto blokową postać DCT, dla $N = 8$, ze względu na redukcję efektów Gibbsa (pierścieniowe zniekształcenia powodowane obcinaniem górnej części pasma) oraz dostosowanie do lokalnej charakterystyki widmowej obrazu. Istotną okazała się także możliwość przyspieszenia procesu transformacji – przy podziale blokowym redukowana jest liczba obliczeń w stosunku do wersji pełnokadrowej, można opracować szybkie algorytmy ze stablicowanymi wartościami kosinusów, możliwe jest też zrównoleglenie obliczeń niezależnych przekształceń w blokach. Dwuwymiarowe funkcje bazowe przedstawiono na rys. 5.1.

Norma określa $N = 8$, a obliczone 64 współczynniki DCT z każdego bloku podlegają **kwantyzacji** skalarnej z przedziałem kwantyzacji dobranym dla każdej harmonicznej. Kolejne wartości $k(u, v)$ są dzielone przez odpowiadające im



Rysunek 5.1: Baza 64 dwuwymiarowych funkcji kosinusowych zastosowana w JPEG; według ogólnie przyjętej konwencji w lewym górnym rogu bloku ustalana jest wartość średnia pikseli w bloku, która odpowiada współczynnikowi określonemu przez splot danych z funkcją stałą); kolejne wartości współczynników DCT są efektem splotu danych z kolejnymi harmonicznymi bazy funkcji kosinusowych.

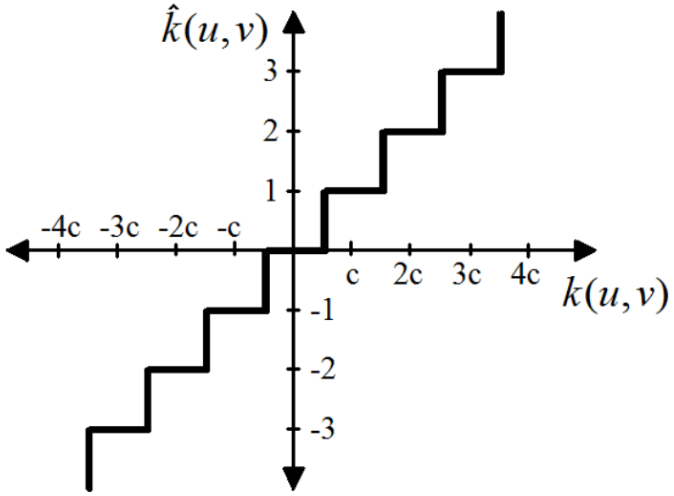
wartości w tablicy kwantyzacji $\mathbf{Z} = \begin{bmatrix} z(0,0) & \cdots & z(7,0) \\ \vdots & \vdots & \vdots \\ z(0,7) & \cdots & z(7,7) \end{bmatrix}$ i zaokrąglone do najbliższej liczby całkowitej: $\hat{k}(u,v) = \left[\frac{k(u,v)}{z(u,v)}\right]$. Tablicę tę można dobrać w zależności od aplikacji, przykładowe postacie oddzielnych tablic dla luminancji i chrominancji pokazano na rys. 5.2.

Uzyskano schemat kwantyzacji równomiernej (ze stałym przedziałem) z zerem, jak na rys. 5.2. Dla przykładowej wartości $z(u,v) = c$ wartość $\hat{k}(u,v) = p$ wtedy i tylko wtedy, jeśli $c(p - \frac{1}{2}) \leqslant k(u,v) < c(p + \frac{1}{2})$. Dla $|k(u,v)| < \frac{c}{2}$ mamy $p = 0$, czyli następuje wyzerowanie współczynnika, tj. usunięcie odpowiedniej harmonicznej z widma sygnału odtwarzanego w procesie dekompresji według (5.3).

Powyższe rozwiązanie pozwala przy stosunkowo niewielkim błędzie kwantyzacji zachować współczynniki o wartościach powyżej progu $\frac{c}{2}$. Poprzez dobranie odpowiednich wartości tablicy kwantyzacji $z(u,v)$, zgodnie z wagą percepcji ustalonej dla poszczególnych współczynników-harmonicznych można zachować wysoką psychowizualną jakość obrazu w subiektywnej ocenie obserwatora. Standard JPEG dopuszcza dowolną postać tablicy kwantyzacji, jednak w normie podano rekomendowane postacie tablic kwantyzacji dla luminancji i chrominancji (rys. 5.2). Tablice te zawierają doświadczalnie dobrane wartości, które odzwierciedlają subiektywne wrażenie odbioru poszczególnych składowych harmonicznych obrazów. Są one efektem kilkuletnich badań nad psychowizualnym odbiorem obrazów naturalnych.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99



Rysunek 5.2: Charakterystyka procesu kwantyzacji w JPEG: u góry – tablice kwantyzacji dla luminancji (po lewej) i chrominancji, optymalizujące jakość rekonstruowanych obrazów pod kątem oceny psychowizualnej (rekomendowane w normie JPEG, jednak nie obligatoryjne); u dołu – krzywa równomiernej kwantyzacji skalarnej z przedziałem zerowym.

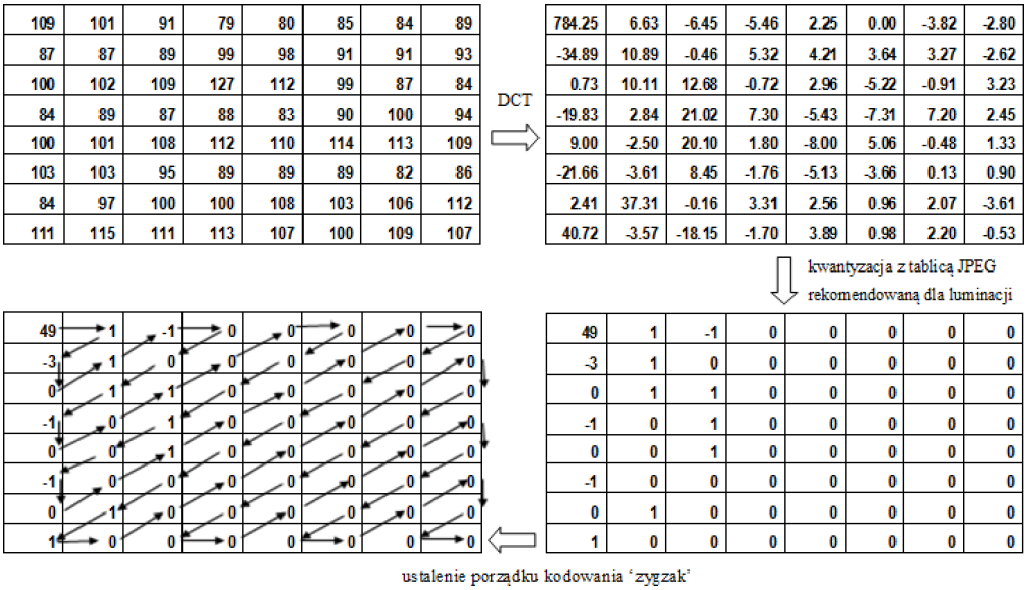
Silniejsza kwantyzacja współczynników w blokach powoduje charakterystyczny dla JPEG efekt blokowy, który jest konsekwencją uzyskanej nieciągłości funkcji jasności obrazu na granicach bloków – rys. 5.3.

Kolejnym etapem schematu kompresji jest kodowanie kwantowanych współczynników DCT. Wartości współczynników każdego bloku są ustawiane w jednowymiarowy ciąg danych według sekwencji zygzak, a następnie kodowane z wykorzystaniem metody Huffmana, kodu binarnego oraz RLE. Charakterystyczny rozkład wartości współczynników to składowa stała oraz dominujące zwykle, co do wartości bezwzględnej, wartości współczynników niskich harmonicznych, malejące przy przechodzeniu do składowych o wyższej częstotliwości, z dużą liczbą zer. Ustalenie ciągu kodowanych danych według porządku zygzak pozwala na

Rysunek 5.3: Efekt blokowy uzyskany dla kompresji JPEG w stopniu 43:1 (para obrazów u góry – z lewej oryginał Lena) oraz 75:1 (dolna para obrazów Barbara).

ustawienie w najbliższym sąsiedztwie danych o zbliżonych wartościach. Uzyskuje się w ten sposób jednowymiarowy ciąg wartości zakończony dużą ilością zer, podatny na efektywne kodowanie - zobacz rys. 5.4. Składową stałą bloku koduje się różnicowo, tzn. koduje się jedynie różnicę pomiędzy wartością składowej stałej obecnego bloku i poprzedniego. Do kodowania kategorii wartości współczynnika używa się kodu Huffmana, przy czym rekomendowana tablica słów kodowych nie jest obligatoryjna. Doprecyzowanie wartości współczynnika wewnątrz kategorii odbywa się za pomocą kodu dwójkowego. W przypadku składowej stałej (DC) ustalono słowa Huffmana dla 12 takich kategorii (rys. 5.5), zaś dla składowych zmiennych (AC) utworzonych zostało 10 kategorii wartości (z wykluczeniem wartości 0), przy czym każda z kategorii otrzymała różne słowa kodowe w zależności od liczby zer poprzedzających wartość niezerową - rys. 5.5. Norma w wersji roz-

szerzonej dopuszcza także stosowanie kodera arytmetycznego.



Rysunek 5.4: Przykład przekształceń w procesie kodowania JPEG: kolejno przykładowy blok obrazu z wartościami funkcji jasności, obliczone współczynniki DCT, te same współczynniki po kwantyzacji z użyciem tablic luminancji z rys. 5.2 oraz ustalenie kolejności kodowania skwantowanych współczynników według porządku zygzak.

Proces dekompresji przebiega dokładnie odwrotnie, przy czym dekoder musi posługiwać się dokładnie tymi samymi tablicami specyfikacji (tablica kwantyzacji, Huffmana). Umożliwia to format zapisu (JFIF), w którym tablice specyfikacji poprzedzone odpowiednimi markerami umieszczone są w pliku razem z danymi skompresowanymi.

Ogólny schemat algorytmu przedstawiono na rys. 5.6.

Tryb rozszerzony i hierarchiczny

Rozszerzony proces kompresji umożliwia kompresję zarówno 8- mio jak i 12- to bitowych danych, przy czym kodowanie może być nie tylko sekwencyjne, ale i progresywne. W modzie progresywnym poszczególne bloki współczynników przeglądane są w tej samej kolejności, jednak wartości ich współczynników kodowane są częściowo w wielu skanach, zgodnie z podziałem na poszczególne podpasma lub mapy bitowe. Rozwiązanie to wymaga jednak zapewnienia dodatkowego bufora pamięci do przechowania skwantowanych wartości współczynników całego obrazu, odpowiednio porządkowanych na etapie ich finalnego kodowania oraz formowania strumienia wyjściowego.

Kategoria	Zakres wartości różnicowej dla współczynnika DC	Długość słowa kodowego	Słowo kodowe
0	0	2	00
1	–1,1	3	010
2	–3,–2,2,3	3	011
3	–7,...,–4,4,...,7	3	100
4	–15,...,–8,8,...,15	3	101
5	–31,...,–16,16,...,31	3	110
6	–63,...,–32,32,...,63	4	1110
7	–127,...,–64,64,...,127	5	11110
8	–255,...,–128,128,...,255	6	111110
9	–511,...,–256,256,...,511	7	1111110
10	–1023,...,–512,512,...,1023	8	11111110
11	–2047,...,–1024,1024,...,2047	9	111111110

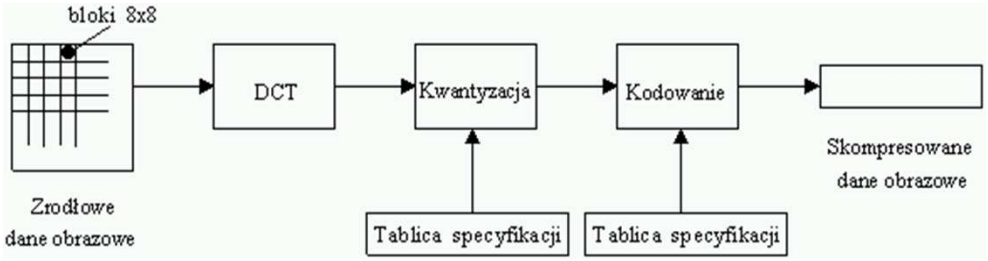
tablica składowej stałej

Kategoria(liczba bitów uzupełniającej części słowa kodowego)	Zakres wartości współczynnika AC
1	–1,1
2	–3,–2,2,3
3	–7,...,–4,4,...,7
4	–15,...,–8,8,...,15
5	–31,...,–16,16,...,31
6	–63,...,–32,32,...,63
7	–127,...,–64,64,...,127
8	–255,...,–128,128,...,255
9	–511,...,–256,256,...,511
10	–1023,...,–512,512,...,1023

Słowo RS (liczba poprzedzających zer/kategoria)	Długość słowa kodowego	Słowo kodowe
0/0 (EOB)	4	1010
0/1	2	00
0/2	2	01
0/3	3	100
...	...	...
0/8	10	1111110110
0/9	16	1111111110000010
0/A	16	1111111110000011
1/1	4	1100
1/2	6	111001
...	...	...
2/A	16	1111111110001110
3/1	6	111010
...	...	...
C/9	16	1111111111100000
C/A	16	1111111111100001
D/1	11	11111111000
...	...	...
F/0 (same zerowe współczynniki)	11	11111111001
F/1	16	1111111111110101
...	...	...
F/9	16	1111111111111101
F/A	16	1111111111111110

tablice składowych zmiennych

Rysunek 5.5: Tablice kodowania rekomendowane normą JPEG, choć nieobligatoryjne: dla składowej stałej DC (po lewej u góry) oraz składowych zmiennych AC - tablica z podziałem na kategorie wartości oraz tablica słów kodowych poszczególnych kategorii poprzedzonych określoną liczbą współczynników zerowych.



Rysunek 5.6: Ogólny schemat blokowy algorytmu kompresji ze standardu JPEG, bazujący na blokowej DCT.

W normie występują dwa rodzaje procedur progresywnego kodowania. Pierwsza, nazywana selekcją widma, dzieli współczynniki ustawione według sekwencji zygzak na kolejne pasma, które zawierają poszczególne części częstotliwościowego spektrum każdego z bloków. Druga procedura związana jest z precyzją, z jaką kodowane są współczynniki w każdym z pasm i nazywana jest sukcesywną aproksymacją. Najpierw kodowana jest pewna liczba bardziej znaczących bitów wartości tych współczynników, a następnie mniej znaczące bity. Można progresywnie kodować współczynniki jedynie przy pomocy procedury selekcji widma, jak też z wykorzystaniem obu procedur. Wówczas mamy do czynienia z tzw. peł-

ną progresją. Ze stwierdzeń zawartych w opisie normy wynika, że zastosowanie selekcji widmowej, jakkolwiek wygodne dla wielu zastosowań, daje porównywalne bądź nieco gorsze wyniki kompresji niż sekwencyjna metoda kodowania, podczas gdy przy pełnej progresji skuteczność kompresji może się okazać nieco większa.

W rozszerzonym procesie kodowania możliwe jest także arytmetyczne kodowanie, a tablica warunkowych zależności danych (*conditioning table*) jest wówczas zapamiętywana jako tablica specyfikacji.

Hierarchiczny proces kodowania macierzy obrazu polega na tworzeniu jego reprezentacji w postaci sekwencji kadrów o różnej rozdzielczości za pomocą rozszerzonego procesu kodowania (wykorzystującego DCT) lub metody bezstratnej (bazującej na predykcji). Można także połączyć te dwie metody i w schemacie kodowania z DCT zastosować kodowanie bezstratne współczynników jedynie zaokrąglonych do najbliższej liczby całkowitej (a więc bez opisanej wyżej metody kwantyzacji z wyspecyfikowaną tablicą), co daje efekt kompresji prawie bezstratnej (zmiany źródłowych wartości pikseli poprzez błędy przybliżeń są praktycznie niezauważalne).

Tworzone są zrekonstruowane składniki odniesienia obrazu różnej rozdzielczości i koduje się różnice pomiędzy obrazem oryginalnym a tymi składnikami odniesienia. Stosowanie filtrów próbkujących obraz oryginalny z różną rozdzielczością tworzy charakterystyczną piramidę przestrzennej rozdzielczości.

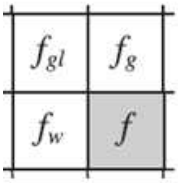
Tryb hierarchiczny może być stosowany alternatywnie, aby zwiększyć jakość rekonstruowanych składników obrazu o danej rozdzielczości w stosunku do procedury z progresją z trybu rozszerzonego, bardziej kontrolując poszczególne poziomy rozdzielczości. W tej procedurze kodowania możliwe jest porządkowanie treści przekazu według ustalonej progresji skali, co jest szczególnie użyteczne w systemach posługujących się wielorozdzielczymi wersjami danych obrazów (zależnymi np. od parametrów urządzeń do wizualizacji, drukarek czy też wymagań do przetwarzania obrazów).

Bezstratny JPEG

Standard opisujący odwracalny algorytm kodowania obrazów wykorzystuje predykcyjne kodowanie wartości pikseli z kilkoma wariantowymi modelami o różnych kontekstach. Można wybrać najbliższe sąsiedztwo piksela rzędu 1, 2 lub 3 jak na rys. 5.7a). Dostępnych jest kilka najprostszych podstawień lub liniowych predykcji ponumerowanych od 1 do 7 jak w tabeli na rys. 5.7b). Opcja 0 oznacza rezygnację z jakiejkolwiek formy predykcji, co jest szczególnie przydatne w różnicowym kodowaniu według trybu hierarchicznego [215].

Opcje 1,2 i 3 to najprostsze predykcje rzędu 1 dobierane ze względu na dominujący kierunek ułożenia informacji w obrazie. Predykcja z numerem 7 wykorzystuje dwóch najbliższych sąsiadów w pionie i w poziomie, podczas gdy opcje 4-6 są najbardziej złożonymi modelami rzędu o bardzo uproszczonym rozkładzie

a)



b)

Numer	Funkcja przewidywania
0	$\hat{f} = 0$
1	$\hat{f} = f_w$
2	$\hat{f} = f_g$
3	$\hat{f} = f_{gl}$
4	$\hat{f} = f_w + f_g - f_{gl}$
5	$\hat{f} = f_w + \lfloor (f_g - f_{gl})/2 \rfloor$
6	$\hat{f} = f_g + \lfloor (f_w - f_{gl})/2 \rfloor$
7	$\hat{f} = \lfloor (f_w + f_g)/2 \rfloor$

Rysunek 5.7: Modele predykcji w bezstratnym JPEG: a) przestrzenny kontekst sąsiednich pikseli (f_w, f_{gl}, f_g) wykorzystywanych przy kodowaniu f; b) równania określające wartość przewidywaną $\hat{f}$ w poszczególnych trybach kodowania predykcyjnego.

wartości wag.

Proces kodowania bezstratnego ujęty w standardzie zamiast koncepcji kodowania transformacyjnego z DCT realizowaną w kwadratowych blokach 8×8 pikseli, proponuje liniową predykcję kodowanych wartości dobieraną dla danego skanu. Różnica pomiędzy wartością przewidywaną $\hat{f}$, a źródłową wartością piksela f kodowana jest dostosowaną metodą Huffmana lub arytmetycznie. W metodzie Huffmana stosuje się 17 kategorii różnicowych wartości błędu predykcji, a w kodowaniu arytmetycznym konstruowany jest dwuwymiarowy model statystyczny.

Odwracalny koder JPEG cechuje stosunkowo mała efektywność kompresji obrazów ze względu na zbyt proste w wielu zastosowaniach modele predykcji oraz wykorzystanie metod entropijnych o ograniczonej efektywności (statyczny koder Huffmana) bądź stosowalności (ze względu na ograniczenia powszechnego dostępu do binarnego kodera arytmetycznego QM). Zaletą jest jednak to, że według specyfikacji JPEG metodą tą można odwracalnie kodować obrazy o dynamice od 2 do 16 bitów, co oznacza potencjalnie bardzo szeroki zakres zastosowań. W przypadku medycznych danych obrazowych, gdzie często stosowane są wyłącznie bezstratne metody kompresji, szczególnie cenna jest możliwość kodowania danych obrazowych dwubajtowych (wartość funkcji jasności piksela zapisana jest na 2 bajtach). Taka sytuacja występuje np. w tomografii komputerowej, tomografii rezonansu megnetycznego czy w radiografii cyfrowej i mammografii.

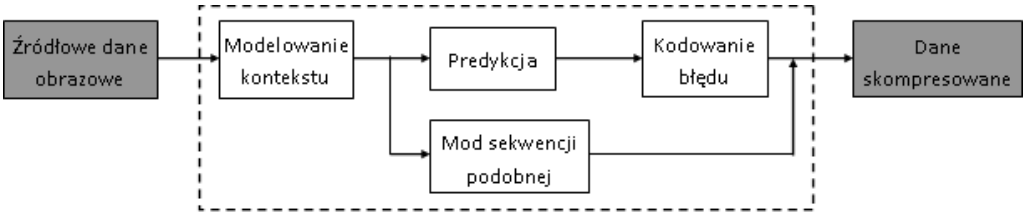
Przykładowo, bezstratną procedurę JPEG wykorzystano na Uniwersytecie Południowej Florydy do archiwizacji znaczącego zestawu obrazów mammograficznych DDSM [226]. Baza ta zawiera wiele referencyjnych, zweryfikowanych

diagnostycznie rezultatów badań, które są wykorzystywane przez naukowców na całym świecie od wielu lat. Te ogólnodostępne zasoby są przydatne w porównawczej ocenie radiologicznej, jak też w testowaniu nowych metod przetwarzania, analizy i poprawy jakości obrazów, wykorzystywanych w komputerowym wspomaganiu diagnostyki, a także w optymalizacji metod kompresji obrazów.

5.1.2 JPEG-LS, czyli mało popularny kodek o dużych możliwościach

Ograniczenia efektywności bezstratnego JPEG doprowadziły do opracowania i przyjęcia w 1996 roku nowego standardu kompresji obrazów rodziny JPEG – JPEG-LS (*lossless and near lossless*) [94, 89].

Główne etapy bezstratnej/prawie bezstratnej kompresji obrazów według standardu JPEG-LS, tj. modelowanie kontekstu, predykcja, kodowanie błędu predykcji oraz alternatywny mod sekwencji podobnej (w szczególnym przypadku kompresji bezstratnej - identycznej), przedstawiono na rys. 5.8.



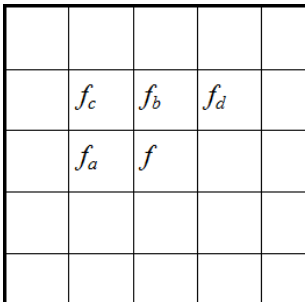
Rysunek 5.8: Schemat blokowy kodera JPEG-LS.

Modelowanie kontekstu polega na w pierwszej kolejności na ustaleniu rozmiaru i kształtu najbliższego sąsiedztwa każdego z kodowanych pikseli. Na podstawie rozkładu wartości pikseli sąsiedztwa: a) wybierany jest tryb kodowania; b) obliczana jest przewidywana wartość piksela, c) konstruowany jest statystyczny model błędów predykcji (tj. różnicy pomiędzy wartością piksela i wartością przewidywaną). Lokalny kontekst z JPEG-LS pokazano na rys. 5.9.

Kontekst określony na podstawie czterech danych o położeniu a, b, c i d pozwala po pierwsze określić, czy informacja zawarta w wartości f ma być kodowana w trybie regularnym czy w trybie sekwencji próbek identycznych:

- tryb **sekwencji** pikseli podobnych jest wybierany wówczas, gdy estymacja na podstawie kontekstu wskazuje, że wartości sąsiednich pikseli z dużym prawdopodobieństwem są prawie identyczne, z tolerancją dopuszczalną w kodowaniu prawie bezstratnym (identyczne w kodowaniu bezstratnym);

- tryb **regularny** jest wybierany wówczas, gdy z estymacji kontekstowej wynika, że nie istnieje duże prawdopodobieństwo wystąpienia kolejnych pikseli prawie identycznych, z tolerancją dopuszczalną w kodowaniu prawie bez-



Rysunek 5.9: Lokalny, przyczynowy kontekst wystąpienia kodowanego piksela f w przestrzeni obrazu, wykorzystany w standardzie JPEG-LS do wyboru trybu kodowania, w nieliniowej predykcji oraz przy modelowaniu błędu predykcji na etapie binarnego kodowania; sąsiednie piksele o wskaźnikach pozycji a, b, c stanowią elementy kontekstu predykcji rzędu 3, natomiast poszerzony zbiór pikseli f_a, f_b, f_c, f_d został użyty w ocenie lokalnych zależności danych do selekcji trybu kodowania oraz zwiększenia efektywność modelu źródła stosowanego w kodowaniu binarnym.

stratnym (identycznych w bezstratnym); wówczas stosowana jest predykcja wartości kodowanego piksela na podstawie kontekstu jego wystąpienia.

Kodowanie sekwencji próbek podobnych

Charakter kontekstu ustalany jest za pomocą prostej estymacji gradientów lokalnych pikseli sąsiednich według zależności: $d_1 = f_d - f_b$, $d_2 = f_b - f_c$, $d_3 = f_c - f_a$. Warunek kontekstu pikseli podobnych w przypadku kompresji odwracalnej wygląda następująco: $d_1 = d_2 = d_3 = 0$, co oczywiście oznacza identyczność pikseli sąsiedztwa: $f_a = f_b = f_c = f_d$. Dopuszczając stratność procesu kompresji, redefiniowany jest warunek kontekstu pikseli podobnych (z dokładnością σ) w sposób następujący: $\forall_{i\in\{1,2,3\}}|d_i| \leqslant \sigma$. Przy spełnieniu powyższych warunków wartość piksela kodowana jest w trybie sekwencji próbek podobnych.

Oczekiwana jest wtedy seria kolejnych próbek o wartościach identycznych z f_a zakończona pojawieniem się piksela o innej wartości lub końcem aktualnego wiersza. Długość serii jest kodowana z wykorzystaniem adaptacyjnej wersji kodu Golomba EG_m (kod elementarny z ograniczeniem wartości rzędu m do potęgi dwójki). W przypadku przerwania serii pikseli podobnych kodowana jest różnica pomiędzy f i f_b (modulo rozmiar skończonego alfabetu wartości pikseli).

Nieliniowa predykcja trybu regularnego

W przypadku nie spełnienia warunku podobieństwa (identyczności) przez wartości pikseli sąsiedztwa kodowanego piksela stosowany jest bardziej złożony tryb regularny (dominujący w przypadku obrazów naturalnych), wykorzystujący nie-

liniową predykcję z trójelementowego kontekstu oraz statystyczne modelowanie kontekstu i adaptacyjny kod Golomba do kodowania błędu predykcji.

W standardzie kompresji bezstratnej JPEG-LS wykorzystano prosty kontekst predykcji trzeciego rzędu składający się z elementów o położeniu a, b i c jak na rys. 5.9. Model predykcji MED/MAP [223, 89] jest nieliniowy, określony dla wartości f następująco:

$$\hat{f} = \begin{cases} \min(f_a, f_b), & f_c \geqslant \max(f_a, f_b) \\ \max(f_a, f_b), & f_c \leqslant \min(f_a, f_b) \\ f_a + f_b - f_c, & \text{wpp} \end{cases} \tag{5.4}$$

gdzie wartość przewidywana $\hat{f}$ ustalana jest adaptacyjnie na okoliczność znajdującej się w najbliższym sąsiedztwie krawędzi obiektu (gdy f_c nie jest medianą pikseli sąsiedztwa), bądź też na okoliczność obszaru łagodnego (f_c jest medianą).

Dodatkowo, błąd predykcji $\epsilon = f - \hat{f}$ jest korygowany za pomocą składnika zależnego od kontekstu, aby skompensować systematyczny błąd odchylenia w predykcji. Obciążenie operatora predykcji jest możliwe wskutek wykorzystania predykcji medianowej oraz przybliżeń $\hat{f}$ do wartości całkowitej (w skończonym zbiorze punktów).

Po predykcji według (5.4) wartość $\hat{f}$ jest korygowana zależnie od odchylenia błędu predykcji przy danym konteście. Założono dwustronny geometryczny, symetryczny rozkład błędów predykcji ze środkiem (wartością średnią) w przedziale $< -1, 0 >$. Celem jest utrzymanie środka rozkładu błędów w tym przedziale. Procedura wykorzystuje sumaryczną wartość błędów próbek zanotowanych przy tym samym kontekście kodowania.

Zmienna odchylenia błędu `B[Q]` pozwala na aktualizację wartości korekcji predykcji `C[Q]` najwyżej o jeden w każdej iteracji. Wartość `C[Q]` obliczana jest zgodnie z procedurą jak niżej (przykład 5.1), podobnie jak aktualna postać `B[Q]`:

Przykład 5.1 *Oprogramowanie: korekcja błędu predykcji związana z obciążeniem operatora predykcji w JPEG-LS*

```
/* Aktualizacja zmiennych B[Q] i C[Q]; zmienna C[Q] zawiera wartość
   korekcji wartości przewidywanej Px; SIGN równe jeden oznacza
   kontekst 'dodatni' */
B[Q] = B[Q] + E; /* suma błędów predykcji E */
N[Q] = N[Q] + 1; /* liczba wystąpień Q od inicjalizacji */

if (B[Q] <= -N[Q]) {
   B[Q] = B[Q] + N[Q]; C[Q] = C[Q] - 1;
   if (B[Q] <= -N[Q])  B[Q] = -N[Q] + 1;
} else if (B[Q] > 0) {
   B[Q] = B[Q] - N[Q]; C[Q] = C[Q] + 1;
   if (B[Q] > 0) B[Q] = 0;
```

```
}
....
if (SIGN == +1)  /* korekcja przewidywań */
   Px = Px + C[Q];
else Px = Px - C[Q];
```

□

Skorygowany błąd predykcji jest następnie kodowany z wykorzystaniem adaptacyjnego kodu Golomba, którego podstawowy parametr (tj. rząd) dobierany jest na podstawie rozkładu wartości błędu predykcji przy danym kontekście.

Modelowanie kontekstu

Chcąc objąć modelem statystycznym entropijnego kodowania szerszy niż w predykcji obszar potencjalnych zależności danych zdefiniowano najbliższe sąsiedztwo 4 pikseli a, b, c, d, jak na rys. 5.9. Na jego podstawie określono wykorzystany w kodowaniu binarnym, skwantowany kontekst rzędu 4 składający się z następujących różnic:

$$\Delta\epsilon_1 = \epsilon_d - \epsilon_b, \quad \Delta\epsilon_2 = \epsilon_b - \epsilon_c, \quad \Delta\epsilon_3 = \epsilon_c - \epsilon_a \tag{5.5}$$

gdzie wartości błędów predykcji $\epsilon = f - \hat{f}$ obliczane są według modelu predykcji (5.4). Wykorzystując kontekst (5.5) do estymacji prawdopodobieństw warunkowych konstruowany jest model $P(\epsilon|\Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3)$.

Alfabet każdego z trzech elementów kontekstu $C^{(4)} = (\Delta\epsilon_1, \Delta\epsilon_2, \Delta\epsilon_3)$ zmniejszono za pomocą operatora kwantyzacji $Q(\cdot)$ o $2T+1$ regionach indeksowanych: $\{-T, -T+1, \ldots, -1, 0, 1, \ldots, T\}$, co daje $(2T+1)^3$ różnych kontekstów. Przykładowo, dla $T = 4$ mamy indeksy regionów: $\{-4, -3, \ldots, 3, 4\}$, co daje $9^3 = 729$ stanów modelu kontekstu. Kwantyzacja $Q(\cdot)$ nie musi być równomierna (o stałej szerokości przedziału kwantyzacji). W JPEG-LS dopuszczono kwantyzację nierównomierną, np. według następujących przedziałów domyślnych (zakładając $T = 4$): $\{0\}, \pm\{1, 2\}, \pm\{3, 4, 5, 6\}, \pm\{7, 8, \ldots, 20\}, \pm\{\Delta\epsilon > 20\}$. Procedurę kwantyzacji alfabetu kontekstu przedstawia poniższy fragment kodu (przykład 5.2).

Przykład 5.2 *Oprogramowanie: kwantyzacja alfabetu kontekstu przy kodowaniu błędów predykcji w JPEG-LS*

```
/* Różnica błędów predykcji E w punkcie i jest kwantowana do
   wartości QEi; T1,T2,T3 to dobierane granice przedziałów*/
if (Ei <= -T3) QEi = -4;
else if (Ei <= -T2) QEi = -3;
else if (Ei <= -T1) QEi = -2;
else if (Ei < 0) QEi = -1;
else if (Ei == 0) QEi = 0;
else if (Ei < T1) QEi = 1;
```

```
else if (Ei < T2) QEi = 2;
else if (Ei < T3) QEi = 3;
else Ei = 4;
```

□ Domyślnie T1=3, T2=7, T3=21. Okazuje się, że dla mniejszych zbiorów danych, np. obrazów o rozmiarach 64×64, lepiej jest użyć jeszcze silniejszej kwantyzacji kontekstów, np. do 63 stanów trójelementowego kontekstu z alfabetem składającym się z pięciu symboli zamiast dziewięciu [224]. Daje to poprawę efektywności kompresji, ponieważ model o mniejszej liczbie stanów może być bardziej wiarygodny.

Aby dodatkowo zmniejszyć liczbę stanów modelu z pamięcią ustala się znak kontekstu kodowanej wartości. Kontekst "ujemny" zamieniany jest na "dodatni", a model konstruowany jest wyłącznie na podstawie kontekstów "dodatnich". Adaptacyjnie modyfikowany model w chwili t (używany do kodowania ϵ_{t+1}) określa prawdopodobieństwo wystąpienia określonej wartości ϵ_{t+1} zależnie od kontekstu $C_{\mathbf{t}} = C_t^{(3)} = \mathbf{c} = (g_1, g_2, g_3)$ (określony w każdej chwili jak na rys. 5.9). Jeśli pierwszy niezerowy element kontekstu zaczynając od $g_1 = Q(\Delta\epsilon_1)$ jest ujemny, odwracane są znaki wszystkich elementów niezerowych. "Dodatni" kontekst $-\mathbf{c}$ warunkuje wtedy prawdopodobieństwo symbolu w kodzie binarnym. Założono w tym przypadku symetryczność rozkładu:

$$\Pr[\epsilon_{t+1} = \varepsilon | C_t^{(3)} = \mathbf{c}] = \Pr[\epsilon_{t+1} = -\varepsilon | C_t^{(3)} = -\mathbf{c})] \tag{5.6}$$

gdzie $\varepsilon \in A_{\mathcal{E}}$. Takie rozwiązanie zakłada kodowanie wartości $-\epsilon_{t+1}$ dla kontekstów "ujemnych", co należy uwzględnić podczas dekodowania.

Uzyskano zmniejszenie liczby stanów do $((2T+1)^3+1)/2 = 365$.

Binarne kodowanie błędu predykcji

Na etapie rozwoju standardu JPEG-LS jako kod binarny stosowano metodę Huffmana z modelem z pamięcią, a także optymalizowany algorytm adaptacyjny kodu Golomba. Pozwoliło to uzyskać bardzo szybką metodę kompresji obrazów o dużej efektywności. Dodatkowy wzrost efektywności kosztem większej złożoności obliczeniowej można uzyskać wykorzystując kodowanie arytmetyczne (proponowane rozszerzenie JPEG-LS) [225]. Używany jest tutaj analogiczny sposób kwantyzacji kontekstu i alfabetu na podstawie sąsiedztwa 5 pikseli.

W finalnej wersji standardu wykorzystano szybki kod Rice'a R_k rzędu k, gdzie k jest dobierane zależnie od kontekstu i zmieniane adaptacyjnie. Wartość k dla danego kontekstu jest uaktualniana za każdym razem, gdy kodowana jest próbka poprzedzona tym kontekstem. Metoda modyfikacji k wykorzystuje skumulowaną sumę modułów błędów predykcji wartości pikseli występujących przy danym kontekście jak niżej:

Przykład 5.3 *Oprogramowanie: dynamiczne ustalanie optymalnego rzędu kodu Rice'a JPEG-LS*

```
/* Zmienna A[0..364] zawiera sumę modułów błędów predykcji kontektu Q;
   w N[0..364] liczone są przypadki wystąpień Q od inicjalizacji; A[Q]
   i N[Q] są aktualizowane zgodnie z aktualnym błędem predykcji E*/
A[Q] = A[Q] + abs(E); /* aktualizacja zmiennych */
N[Q] = N[Q] + 1;
...
for(k=0; (N[Q]<<k)<A[Q]; k++); /* wyznaczenie rzędu k */
```

□

5.1.3 JPEG2000, czyli elastyczny paradygmat

W marcu 1997 rozpoczęto prace nad nowym standardem kompresji obrazów JPEG2000. Obecnie prowadzone są prace (niektóre już zakończono) nad następującymi częściami JPEG2000:

- część 1: podstawowy system kompresji obrazów pojedynczych;
- część 2: rozszerzenia (dodatkowe opcje systemu, m.in. możliwość definiowania własnych transformacji falkowych i kolorów, inny sposób kodowania ROI, nowe algorytmy kwantyzacji itp.);
- część 3: algorytm kompresji sekwencji obrazów (Motion JPEG 2000);
- część 4: testy zgodności (conformance);
- część 5: oprogramowanie referencyjne, tj. biblioteki w Javie (JJ2000) oraz C (JasPer);
- część 6: format pliku dla danych złożonych (obrazowanie dokumentów, artykułów prasowych itp.);
- część 7: (prace nad tą częścią zostały zaniechane);
- część 8: JPSEC, bezpieczeństwo danych;
- część 9: JPIP, interaktywne protokoły transmisji i API;
- część 10: JP3D, obrazowanie 3D;
- część 11: JPWL, aplikacje bezprzewodowe.

Standard JPEG2000 umożliwia kompresję obrazów o niemal dowolnej reprezentacji oryginalnej, tzn. możliwa jest duża liczba komponentów oraz bitów na próbkę w niemal nieograniczonej przestrzeni obrazu, a także różna orientacja dziedziny obrazu względem układu odniesienia. Liczba komponentów jest ograniczona do 214 (16384); natomiast ich maksymalna szerokość i wysokość to 232-1 próbki (przekracza 4 miliardy próbek). Liczba bitów na próbkę jest ograniczona do 38 bitów (ponad 4 bajty).

Nowy paradygmat

Zdolność uzyskiwania możliwie krótkiej reprezentacji danych obrazowych była dotychczas jednym z najważniejszych kryteriów decydujących o przydatności techniki kompresji. Przy obecnym rozwoju nowych technologii multimedialnych równie istotne okazały się nowe cechy funkcjonalne systemów kompresji-dekompresji, powodujące dużą użyteczność algorytmów w znacznie szerszym obszarze zastosowań. Wiele postulatów stawianych przez rosnący krąg potencjalnych użytkowników jest praktycznie nie do zrealizowania za pomocą funkcjonujących dotąd standardów kompresji obrazów, czyli przede wszystkim JPEG-a. Schemat kompresji wynikający z klasycznego paradygmatu definiuje zbiór dostępnych opcji kodera, które pozwalają z jednej strony dopasować algorytm kompresji do własności strumienia wejściowego (dynamiki danych, przestrzeni kolorów itp.), z drugiej zaś - dobrać parametry procesu kompresji (sposób kwantyzacji - np. tablice, szerokość przedziałów, czy kodowania - np. sekwencyjny, progresywny itd.). Po stronie dekodera standard nie dopuszcza żadnych możliwości wyboru sposobu odtwarzania strumienia - obowiązuje ustalony w koderze porządek i sposób rekonstrukcji, a strumień zakodowany jest rekonstruowany w całości. Realizację paradygmatu schematu kompresji ze standardu JPEG przedstawiono na rys. 5.10.a).

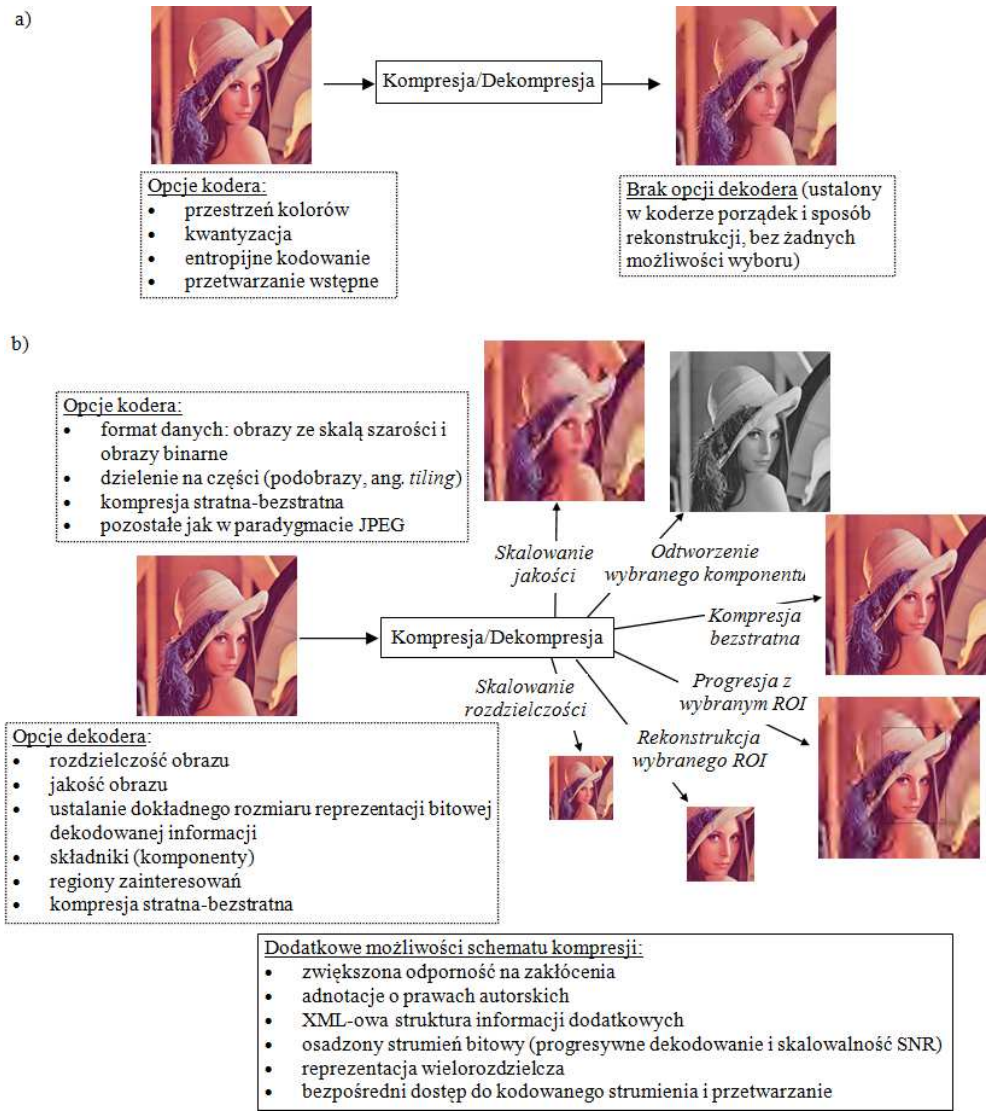
Koncepcja nowego paradygmatu kompresji (rys. 5.10.b)) została sformułowana w okresie intensywnych prac badawczych nad realizacją standardu JPEG2000. W porównaniu z paradygmatem z JPEG-a następuje znaczne rozszerzenie możliwości kształtowania skompresowanego strumienia danych (np. wprowadzanie regionów zainteresowań - ROI, mechanizmu podziału oryginału na części - ang. tile) oraz postaci dekompresowanej reprezentacji oryginału (np. sterowanie rozdzielczością, jakością obrazu, bitowym rozmiarem reprezentacji informacji rekonstruowanej).

Współczesny standard kompresji powinien być tworzony z myślą o szerokiej gamie zastosowań, takich jak Internet, różnorodne aplikacje multimedialne, obrazowanie medyczne, kolorowy fax, drukowanie, skanowanie, cyfrowa fotografia, zdalne sterowanie, przenośna telefonia nowej generacji, biblioteka cyfrowa oraz e-komercja (elektroniczna komercja, głównie z wykorzystaniem Internetu). Nowy system kodowania musi więc być skuteczny w przypadku różnych typów obrazów (binarne, ze skalą szarości, kolorowe, wieloskładnikowe) o odmiennej charakterystyce (obrazy naturalne, medyczne, sztuczne obrazy grafiki komputerowej, naukowe z różnych eksperymentów, z tekstem itd.). Ponadto, powinien zapewniać efektywną współpracę z różnymi technologiami obrazowania (akwizycji/generacji/ wykorzystywania obrazów): z transmisją w czasie rzeczywistym, z archiwizacją, gromadzeniem bazodanowym (np. w cyfrowej bibliotece), w strukturze klient/serwer, z ograniczonym rozmiarem bufora czy limitowaną szerokością pasma itp. Wymagania odnośnie nowego standardu kompresji najlepiej charakteryzują stwierdzenia z 'JPEG2000 call for proposals' z marca 1997 r.: "poszu-

kiwany jest standard dla tych obszarów, gdzie aktualne standardy nie potrafią zagwarantować wysokiej jakości lub wydajności, standard zapewniający nowe możliwości na rynkach, które dotąd nie wykorzystywały technologii kompresji i dający otwarte narzędzia systemowe dla aplikacji obrazowych."

Użyteczność falek w modelowaniu i realizacji Użyteczność analizy falkowej leży głównie w możliwościach skutecznego modelowania przy ich pomocy przestrzenno-częstotliwościowej charakterystyki obrazów naturalnych. Typowa postać obrazu jest mieszaniną obszarów o znacznych rozmiarach z treścią niskoczęstotliwościową (wolno zmieniające się tło sceny naturalnej, np. niebo, ściana w pokoju, czy też jednostajne, lekko zaszumione tło w obrazie medycznym), małych obszarów ze znaczącą treścią wysokoczęstotliwościową (silne krawędzie o dużych gradientach, wyraźne, drobne struktury) oraz obszarów (obiektów) pokrytych teksturami. W bazie falkowej znajdują się elementy o skończonym nośniku, które mają dobrą rozdzielczość częstotliwościową (przy słabszej przestrzennej) w zakresie częstotliwości niskich, czyli dobrze charakteryzują tło i wolnozmienne tekstury, a także elementy z dobrą rozdzielczością przestrzenną (przy słabszej częstotliwościowej) w zakresie częstotliwości wysokich, co daje dobrą lokalizację krawędzi w podpasmach dokładnej skali. Taka korelacja cech bazy z własnościami obrazów naturalnych prowadzi do silnej koncentracji energii sygnału w niewielkiej liczbie współczynników, przy czym większość informacji obrazu znajduje się w niewielkim obszarze podpasma najniższych częstotliwości dziedziny falkowej dekompozycji. Pozostała część informacji zawarta jest w wartościach współczynników rozrzuconych w niewielkich grupach wokół przestrzennej pozycji silnych krawędzi w obrazach uzupełniających różnych skal. Zaś zdecydowana większość współczynników ma wartości bliskie zeru i może być zupełnie pominięta lub silnie kwantowana w procesie stratnej kompresji, bez znaczącego wpływu na jakość rekonstrukcji. Wyższość wielorozdzielczej analizy falkowej w stosunku do STFT (ang. Short Time Fourier Transform) lub blokowej DCT (ang. discrete cosine transform) jest tutaj bezdyskusyjna, ponieważ lepiej, w sposób naturalny opisuje rzeczywisty sygnał.

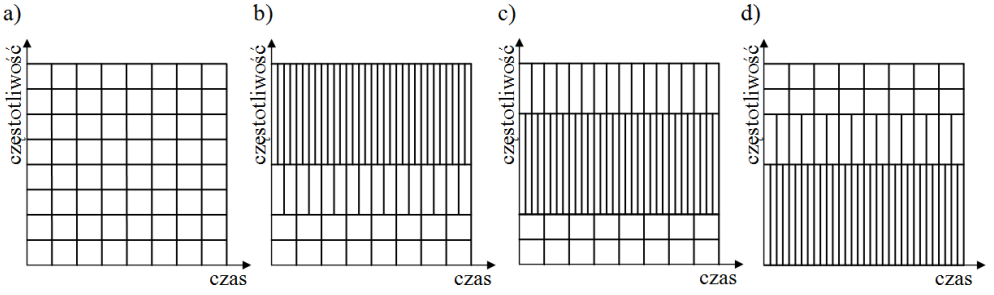
Dodatkową zaletą dekompozycji falkowej jest jej podatność na rozwiązania adaptacyjne, dotyczące zarówno stosowania bazy falkowej o nieskończonej ilości możliwych postaci, jak też samego schematu dekompozycji i podziału na podpasma. Adaptacyjną postać transformacji można lepiej dopasować do własności konkretnego obrazu czy grupy obrazów. Gładkość funkcji bazowych, ich rozmiar, symetryczność, decydują o możliwie najlepszym przybliżeniu lokalnych własności obrazu, a ich właściwy wybór wpływa znacząco na skuteczność kompresji. Zagadnienie doboru optymalnej bazy falkowej w kompresji konkretnego obrazu nie jest właściwie rozwiązywalne w sposób jednoznaczny, ale istnieje cały zbiór przesłanek (w postaci wiedzy a priori) pozwalających dobrać jądro przekształcenia w sposób prawie optymalny.



Rysunek 5.10: Porównanie paradygmatu schematu kompresji: a) z podstawowej wersji standardu JPEG; b) z nowego standardu JPEG2000 - część I, opartego na falkowej technice kompresji; SNR (ang. Signal to Noise Ratio).

Klasyczny schemat wielorozdzielczej dekompozycji Mallata ze strukturą logarytmiczną dobrze opisuje wspomniane własności typowych obrazów z wykładniczo opadającym widmem gęstości mocy. Dla obrazów o nieco innej charakterystyce, zawierających dużą ilość informacji w zakresie wysokoczęstotliwościowym (np. rozległe obszary z teksturą prążkową: czarne pasy na białym tle) bardziej efektywną dekompozycję otrzymuje się za pomocą bazy pozwalającej uzyskać dobrą lokalizację w dziedzinie częstotliwości podpasm wysokoczęstotliwościowych (ze względu dobre zróżnicowanie zgromadzonej tam dużej ilości informacji). Ko-

nieczne jest do tego narzędzie transformacji falkowej, pozwalające dobrać schemat dekompozycji w zależności od cech obrazu (źródła informacji). Choć teoretycznie można zbudować dowolną liczbę schematów dekompozycji, rozwiązując zagadnienie optymalizacji schematu dekompozycji w każdym konkretnym przypadku, to bardziej praktycznym jest rozwiązanie, gdzie można zbudować skończoną bibliotekę reprezentatywnych transformacji, dobierając z biblioteki, dzięki szybkiemu algorytmowi, optymalną postać transformacji dla konkretnego obrazu. Takim narzędziem jest dekompozycja za pomocą pakietu falek (*wavelet packets*), zwana też, bardziej algorytmicznie, schematem wyboru najlepszej bazy (*best basis*). Pakiety falek stanowią dużą bibliotekę transformacji silnie zróżnicowaną pod kątem ich własności dekompozycji przestrzenno-częstotliwościowej, ze zdolnością szybkiego przeszukiwania. Funkcje bazowe pakietu falek mają własności falek klasycznych, przy czym wzbogacone są często o większą liczbę oscylacji. Wykorzystując pakiet falek, można dostosować postać transformacji do praktycznie dowolnego spektrum sygnału - zobacz rys. 5.11. Implementacja dekompozycji z pakietem falek wymaga dodatkowych nakładów obliczeniowych, głównie na proces wyszukania optymalnego schematu transformacji dla danego obrazu, a także przesłania niewielkiej informacji dodatkowej do dekodera. Kosztem większego obciążenia obliczeniowego można niemal dowolnie 'blisko' dopasować postać transformacji do sygnału, np. wprowadzając zmienną w czasie segmentację, pozwalając na ewolucję pakietu falek wraz z sygnałem (silnie niestacjonarnym). Z drugiej strony, procedurę takiej transformacji można znacznie uprościć, dobierając ustaloną postać dekompozycji pakietu falek dla danego typu obrazów, jak to ma miejsce np. w standardzie FBI do kompresji obrazów z odciskami palców.



Rysunek 5.11: Podział dziedziny czas (przestrzeń)-częstotliwość w różnych schematach dekompozycji: a) dekompozycja równomierna, uzyskana za pomocą pakietu falek lub też STFT, b) klasyczna dekompozycja falkowa Mallata, c) dekompozycja pakietu falek, gdzie szerokość podpasm nie zmienia się ani równomiernie ani logarytmicznie, d) odwrócona dekompozycja falkowa, uzyskana z wykorzystaniem pakietu falek.

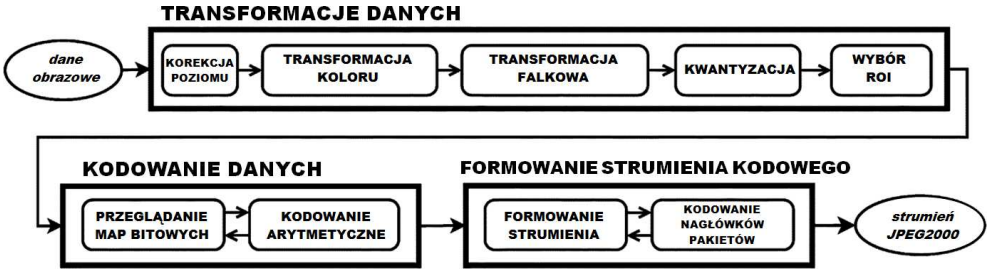
Ponadto, atrakcyjną cechą falkowej reprezentacji obrazu, od skal mało dokładnych (tj. skal dużych), z silną koncentracją informacji (energii) na bit danych, do

małych skal bardzo dokładnych (z małym przyrostem informacji na bit), jest jej naturalna progresywność. Reprezentacja ta umożliwia sterowanie kolejnością i rodzajem przekazywanej informacji, przy czym niewielkim kosztem można uzyskać strumień danych (prawie) optymalny w sensie R-D.

Kompresja konstruowana na podstawie dekompozycji falkowej jakby naturalnie odpowiada na większość wymagań stawianych w nowym paradygmacie kompresji, pozwalając je realizować w szybkich algorytmach obliczeniowych tak, że praca koderów w systemach czasu rzeczywistego jest możliwa.

Metoda kodowania

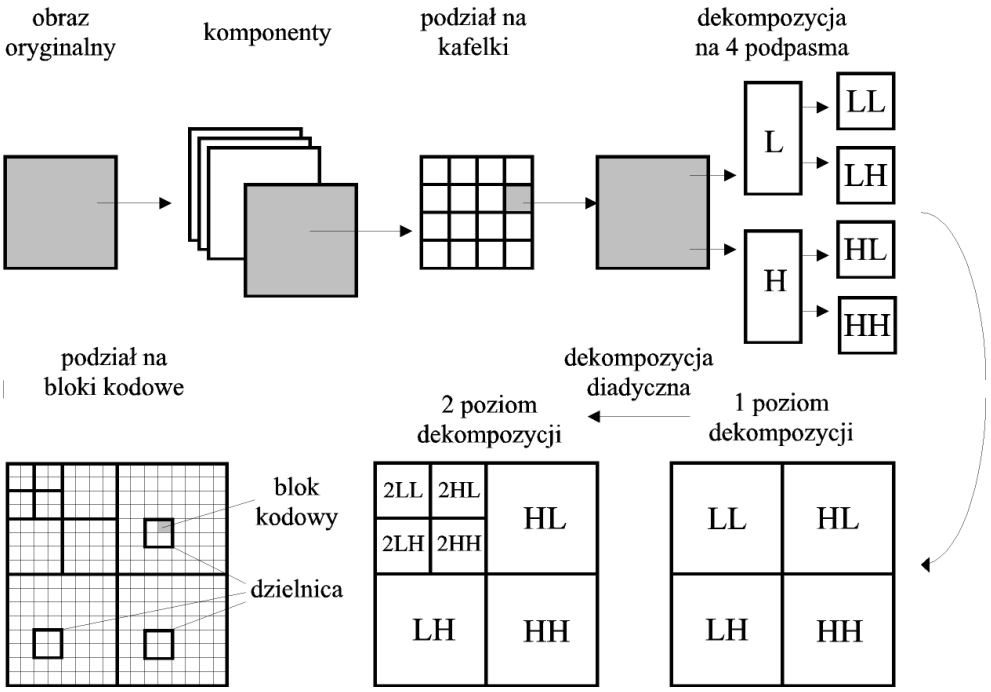
Na rys. 5.12 przedstawiono schemat blokowy algorytmu kompresji zaproponowany w standardzie. Dokładniej poszczególne operacje na danych pokazano na rys. 5.13.



Rysunek 5.12: Diagram ogólnego algorytmu kompresji według JPEG2000.

Pierwszym etapem kompresji jest formatowanie danych polegające na umiejscowieniu poszczególnych komponentów obrazu na siatce odniesienia (*reference grid*) oraz podziale ich na części (*tile*). Zastosowanie siatki umożliwia dokonywanie prostych operacji na obrazie skompresowanym (m.in. obrót o wielokrotność 90 stopni) oraz kompresję obrazów o komponentach o różnej wielkości. Podział obrazu na kafelki ma na celu zmniejszenie zapotrzebowania na pamięć, umożliwia przyspieszenie kodowania (dzięki możliwości równoczesnego kodowania wielu części obrazu) oraz daje dostęp do fragmentów obrazu bez konieczności dekompresji całości (należy zaznaczyć, że taka lokalna dekompresja jest możliwa także w inny sposób). Po podziale obrazu na części każda z nich jest dalej przetwarzane niezależnie.

Po podziale każda część jest poddawany transformacji kolorów. W cz. I standardu zastosowano transformację YCbCr (YCC) w wersji odwracalnej (YCC_R) oraz nieodwracalnej (YCC_I). Transformacja może być zastosowana tylko w przypadku, kiedy obraz zawiera co najmniej 3 komponenty, przy czym 3 pierwsze muszą mieć taką samą wielkość. Po transformacji kolorów wszystkie komponenty są dalej przetwarzane niezależnie. Cz. II standardu umożliwia dodanie innych (własnych) transformacji przestrzeni kolorów.



Rysunek 5.13: Poglądowy zarys struktur danych wykorzystywanych w kolejnych krokach podstawowego algorytmu kodowania JPEG2000.

5.2 Standardy rodziny MPEG

Grupa MPEG tworzy standardy reprezentacji naturalnych danych multimedialnych (tj. danych rejestrowanych jako odbicie natury, czyli otaczającego świata realnego, będącego źródłem informacji) oraz związanymi z nimi metadanych (opisów danych). Celem jest zapewnienie możliwości skutecznej i efektywnej wymiany danych między aplikacjami różnych wytwórców. Ogólniejszym zamierzeniem jest doskonalenie systemów multimedialnych, zwiększanie ich dostępności oraz użyteczności, dzięki szerokiej dystrybucji dokumentów norm i otwartości kodów referencyjnych [27].

Kolejne wersje dokumentów: MPEG-1, MPEG-2, MPEG-4, MPEG-7, czy też MPEG-21 miały nieco odmienny charakter ze względu na zmieniające się z czasem zapotrzebowania, rozszerzane obszary zastosowań, a niekiedy modyfikowany cel standaryzowania. Rozwój rodziny standardów MPEG ma przede wszystkim trudne do przecenienia walory aplikacyjne: dał przede wszystkim podstawy rozwoju cyfrowych archiwów filmowych (MPEG-1) i nowoczesnej telewizji cyfrowej (kompresja sekwencji wizyjnych i dźwięku w MPEG-2), odtwarzaczy MP3 (kompresja dźwięku z MPEG-1), współczesnych zapisów filmów formatu HD (MPEG-4 AVC), nowoczesnych systemów odtwarzania dźwięku (kodeki AC3 czy AAC z MPEG-4), wyszukiwarek obrazów, filmów i zapisów dźwiękowych (opis obiektów multimedialnych, deskryptory wizyjne czy dźwiękowe z MPEG-7) oraz ambitny program przezroczystości zintegrowanych multimediów różnych standardów, zastosowań, koncepcji wykorzystania (MPEG-21).

5.2.1 MPEG 1 i 2, czyli muzyka i telewizja cyfrowa

Norma **MPEG-1** składa się z trzech podstawowych części, dotyczących: a) kodowania wideo, b) kodowania audio, c) multipleksacji (łączenia) strumieni wideo i audio (część systemowa, uwzględniająca tak istotny dla multimediów aspekt integracji różnych strumieni informacji). Części te uzupełnia referencyjne oprogramowanie i dokument opisujący testy zgodności ze standardem. MPEG-1, przeznaczony początkowo do kompresji materiałów cyfrowych przechowywanych na lokalnie dostępnych nośnikach pamięci zewnętrznej (dysk, CD-ROM), z powodzeniem został wykorzystany w strumieniowaniu wideo w lokalnych sieciach komputerowych. Jednak największy wpływ na współczesny świat multimediów wywarł fragment tego standardu, dotyczący kodowania audio (*Audio Layer III*), powszechnie znany i używany do strumieniowania i archiwizacji dźwięku jako MP3 [3]. Określono kodowanie audio w trybie mono i stereo w trzech warstwach, będących kompromisem miedzy złożonością a wydajnością kompresji. Pierwsze dwie bazują na kodowaniu pasmowym, zaś trzecia (MP3) dodatkowo wykorzystuje mechanizmy

[3] Wśród młodzieży bardzo popularny stał się odtwarzacz nazywany "mp-trójką", pozwalający odsłuchiwać niezłej jakości muzykę w niemal każdych warunkach

dopasowania widma kodowanego sygnału do zdolności percepcyjnych odbiorcy w dziedzinie dyskretnej transformacji kosinusowej (DCT).

Podstawowy algorytm kompresji wideo bazuje na metodzie kompresji obrazów pojedynczych opisanej normą JPEG – z blokową dyskretną transformacją kosinusową, kwantyzacją z przedziałami i progami selekcji współczynników DCT opisanymi odpowiednią tablicą oraz binarnym kodowaniem skwantowanych wartości metodą entropijną. Metoda kompresji MPEG została jednak rozszerzona o dodatkowy mechanizm estymacji i kompensacji ruchu w celu usunięcia nadmiarowości czasowej, czyli istniejących zależności (korelacji) pomiędzy kolejnymi ramkami zapisu wideo.

Kompensacja ruchu wykorzystuje zasadę lokalnej predykcji kodowanego kadru na podstawie ustalonej ramki (lub ramek) odniesienia. Odbywa się w strukturze makrobloków, gdzie dla każdego kolejnego bloku obrazu o wymiarach 16×16 wyszukiwany jest (na etapie estymacji ruchu) najbardziej podobny blok ramki odniesienia, stanowiący predykcję informacji kodowanej. Różnicowy błąd predykcji makrobloków kodowany jest z wykorzystaniem blokowej DCT (transformacja niezależna w blokach 8×8) analogicznie jak w JPEG. Predykcja może się odbywać na podstawie ramek poprzedzających, a także następników przybierając formę interpolacji wzdłuż osi czasu (wymaga to zmiany naturalnej kolejności kodowania ramek znakowanych czasem, skorelowanej ze zmianą powrotną w dekoderze formującym strumień wyjściowy). Składnia strumienia MPEG przewiduje zapis tzw. wektorów ruchu, opisujących kolejno przesunięcie (przemieszczenie) najbardziej podobnego makrobloku obrazu odniesienia (referencyjnego) w położenie aktualnie kodowanego bloku ramki.

Kolejna norma, **MPEG-2** jest kontynuacją, rozszerzeniem MPEG-1 zarówno w wersji systemowej, jak i algorytmicznej-sygnałowej (wideo, audio). Wprowadza elementy skalowalności jakościowej (SNR), czasowej i przestrzennej strumienia wideo, rozszerza spektrum rozdzielczości, dołączając profil telewizji wysokiej rozdzielczości HDTV, dodaje warstwę transportową do transmisji w sieciach rozległych (zastosowanie telewizji cyfrowej). W kolejnych dokumentach MPEG-2 pojawia się protokół interakcji i sygnalizacji (DSM-CC – *Digital Storage Media Command and Control*), ochrona praw autorskich (IP *protection*) oraz określenie czasowej precyzji dostawy strumienia transportowego (RTI –*Real Time Interface*).

W MPEG-2 oprócz rozszerzenia liczby kanałów dźwiękowych i zestawu dopuszczalnych częstości próbkowania, dodano zupełnie nowy algorytm AAC (*Advanced Audio Coding*) wykorzystujący DCT, który przy tej samej jakości redukuje przepływność bitową strumienia audio o jedną trzecią w stosunku do MP3. Algorytm AAC zapewnia kilka trybów pracy z różną złożonością i wydajnością bitową.

W MPEG-2 wprowadzono możliwość dopisywania do strumienia dowolnych

danych użytkownika (jako metadane), opisujących dane sygnałowe – składnia i znaczenie metadanych wynikają już z konkretnej aplikacji.

5.2.2 MPEG 4, czyli pułapki złożoności natury

Zmiana koncepcji kodowania wideo jest chyba największym wyróżnikiem kolejnej normy – **MPEG-4**. Zrealizowano w niej koncepcję kodowania obiektowego, z definicją obszaru obiektu o dowolnym kształcie, zgodnie z paradygmatem metod kompresji tzw. drugiej generacji. Obok kompresji sekwencji ramek typowych (definiowanych w prostokątnej dziedzinie pikseli) dopuszczono kompresję sekwencji regionów dowolnych kształtów (obszarów o kształtach zadanych mapą binarną). W normie dodano także możliwość syntetycznego kodowania obrazu i dźwięku. W porównaniu z MPEG-2 udoskonalono możliwości skalowalności strumienia kodowanego oraz dodano mechanizm odporności na błędy (*error resilience*).

Obok mechanizmów śledzenia obiektów o naturalnych kształtach w sekwencji czasowej, w normie zaproponowano nowatorskie narzędzie do kompresji obrazów pojedynczych – kodowanie wizualnej tekstury VTC (*Visual Texture Coding*, przewidziane dla scen multimedialnych zawierających obrazy statyczne. W algorytmie VTC wykorzystano falkową dekompozycję obrazu, realizując kodowanie w nieco uproszczonej wersji niż w JPEG2000.

Możliwe jest także kodowanie obrazów syntetycznych. W zależności od rodzaju syntezy obrazu sztucznego zaproponowano: kodowanie siatek dwuwymiarowych, kodowanie siatek wielokątnych trójwymiarowych, animację sztucznej twarzy i sztucznego ciała. Zdefiniowano trzy typy modelowania wizualnego: modele geometryczne opisujące kształt i wygląd prezentowanych obiektów, modele kinematyczne określające przemieszczenia i zniekształcenia obiektów oraz modele dynamiczne (w tym biomechaniczne), definiujące siły działające na obiekty i naprężenia wewnątrz tych obiektów.

Norma wzbogaca część systemową o narzędzia kompozycji i interakcji z mediami cyfrowymi. Ochrona praw autorskich znajduje odbicie w dwóch protokołach IPMP (*Intelectual Property Management and Protection*): IPMP-H (odpowiednik MPEG-2) oraz IPMP-X z nowymi funkcjonalnościami. Odpowiednikiem protokołu interakcji i sygnalizacji z MPEG-2 (DSM-CC) jest w przypadku MPEG-4 DMIF (*Delivery Multimedia Integration Framework*). DMIF jest protokołem sesji kontrolującej strumieniowanie mediów cyfrowych za pośrednictwem technik transportowych w sieci.

MPEG-4 określa mechanizmy budowy scen multimedialnych złożonych z równoczesnych prezentacji wideo, audio, obrazów statycznych, animacji, efektów grafiki komputerowej. Kluczowym narzędziem temu służącym jest binarny format strumieniowania BIFS (*BInary Format for Streaming*), w którym opisano hierarchiczną strukturę sceny, z mechanizmem zmiany konfiguracji sceny i jej animacji [1]. Tekstowy opis sceny określono w rozszerzalnym tekstowym formacie XMT

(*eXtensible MPEG-4 Textual format*), który w postaci XMT-A jest tekstową wersją BIFS. Cennym rozwiązaniem zamieszczonym w MPEG-4 jest mechanizm interakcji wielu użytkowników współdzielących tę samą scenę multimedialną, zgodnie z mechanizmem "światów wielu użytkowników" (MUW – *multi-user worlds*).

W normie zdefiniowano składnię formatu plikowego MP4 – dla kodowanej reprezentacji treści multimedialnych oraz metadanych wspomagających strumieniowanie, edycję, lokalne odtwarzanie i wymianę treści. Wprowadzono także pojęcie informacji o multimedialnej treści obiektów (OCI – *Object Content Information*).

W obszarze audio zaproponowano w MPEG-4 trzy nowe techniki kompresji dźwięku z przepływnościami bitowymi w zakresie od 2 kbit/s do 64 kbit/s. Ponadto, określono dwa narzędzia reprezentacji danych wspierających tworzenie dźwięku syntetycznego: interfejs systemu "tekst w mowę" (*text to speech interface*) oraz interfejs systemów dźwięku strukturalnego. Pierwszy z nich jest wykorzystywany do zapisu parametrów dźwiękowych (takich jak czas trwania głoski), które równolegle z tekstem są wprowadzane na wejście syntetyzera dźwięku. Określa też parametry umożliwiające synchronizację dźwięku z animowaną twarzą, dopuszcza różne języki w tekście i międzynarodowe symbole dla głosek, przy czym informacja jest wprowadzana do oryginalnego tekstu techniką specjalnych znaczników. W systemie dźwięku strukturalnego zdefiniowano język orkiestry (*Structured Audio Orchestra Language*), który umożliwia konfigurowanie sztucznej orkiestry złożonej z "instrumentów" reprezentowanych przez dynamicznie ładowane strumienie. Ponadto, język definiowania nut (*Structured Audio Score Language*) reprezentuje nie tylko zapis nutowy, ale umożliwia określanie nowych dźwięków i modyfikację dźwięków istniejących.

Obok części 2 standardu, opisującej obiektowe kodowanie wideo, w okresie późniejszym opracowano zupełnie inną koncepcję kodowania wideo, wracając do schematu blokowego i transformacji DCT, jednak na zupełnie innych warunkach adaptacji przyjętych schematów do realiów treści obrazowej. Część 10 zawiera specyfikację metody, w ramach innej ścieżki standaryzacyjnej zwaną H.264 [91], która pozwala wyraźnie zwiększyć efektywność kodowania kosztem wyraźnego wzrostu złożoności obliczeniowej algorytmów. Co warto podkreślić, tak wysoką wydajność kodeka H.264 uzyskano metodą wielu drobnych modyfikacji, powodujących uelastycznienie modelu ruchu obiektów i wynikających z tego, konsekwentnych zmian w zakresie realizacji poszczególnych przekształceń, aż do formowania strumienia zakodowanych danych. Wprowadzono m.in. takie mechanizmy jak:

- predykcja wewnątrzramkowa INTRA, realizowana w blokach 4×4 (dla szczegółów) lub 16×16 (dla płaskich, łagodnych obszarów); nie jest to opcja obligatoryjna;

- wykorzystanie nadpróbkowania, sięgającego nawet 1/8 piksela do estymacji ruchu;

- wykorzystanie w estymacji i kompensacji ruchu bloków o różnej wielkości: $16 \times 16,\ 16 \times 8,\ 8 \times 16,\ 8 \times 8,\ 8 \times 4,\ 4 \times 8,\ 4 \times 4$;
- stosowanie wielu obrazów referencyjnych do kompensacji ruchu, z możliwością ustalenia dowolnych wag, przy czym kierunek ruchu nieistotny, a klatki typu B mogą być referencyjne;
- całkowitoliczbowa wersja transformacji DCT liczona w bloku 4×4, z możliwością rozszerzeń na dziedzinę 8×8 lub 16×16 poprzez wykorzystanie transformacji Hadamarda;
- adaptacyjny filtr redukcji efektów blokowych.

5.2.3 MPEG 7, czyli specyfika opisu

Standard MPEG-7 nie definiuje algorytmów kompresji audio i wideo (czy raczej dekompresji, jak poprzednie normy), ale normuje metadane multimedialne, a więc opisy danych cyfrowych lub też zawartej w nich treści. Opisy te są wykorzystywane do efektywnej charakterystyki, selekcji, doboru czy wyszukania określonych materiałów multimedialnych lub ich fragmentów.

Dokument systemowy określa narzędzia definiowania składni metadanych w języku XML oraz techniki kompresji postaci XML-owej do postaci binarnej BiM (*Binary Format for MPEG-7*). Informacja zawarta w schemacie XML, określającym składnię danego typu opisów, wykorzystywana jest do redukcji strukturalnej nadmiarowości informacji (nazwa elementu, nazwa atrybutu, itp.), natomiast dedykowane techniki kodowania służą do kompresji wartości elementów i atrybutów w dokumencie XML (IEEE 754 dla liczb zmiennoprzecinkowych, UTF_8 dla kodu UNICODE, kodowanie listy wartości, itp.)

Metadane

W MPEG-7 zaproponowano wiele mechanizmów definiowania metadanych, ustalono przede wszystkim bogaty język definicji opisów treści multimedialnej DDL (*Description Definition Language*) [216]. Język DDL jest nieznacznym rozszerzeniem XML Schema, który w MPEG-7 wykorzystywany jest w definiowaniu normowanych typów metadanych dla poszczególnych mediów cyfrowych. Aplikacja generująca metadane, przypisane określonemu strumieniowi danych musi generować opisy według składni zdefiniowanej w typach metadanych normy, a więc zgodnych z odpowiednimi schematami XML.

Główne atrybuty informacji o treści przypisanej obiektom (OCI) to: klasyfikatory treści, słowa kluczowe, opis w języku naturalnym, język strumienia audio, informacja o kontekście wytwórczym danego medium, kategoria przyzwolenia odbioru (np. ocena dopuszczalnego wieku widza danego programu). Metadane mogą

być włączane do opisu obiektu audio-wizualnego (AV), do strumienia elementarnego lub też same mogą tworzyć strumień elementarny. W strumieniu OCI mogą się znajdować opisy wielu obiektów AV.

Typy metadanych zostały podzielone na następujące kategorie:

- typy sygnałowe, odnoszące się do informacji zawartej w danych zarejestrowanego sygnału cyfrowego, np. tekstura czy kształt konturu obiektów obrazu, struktura kolorów, linia melodyczna audio; metadane tego typu są generowane automatycznie na podstawie zarejestrowanego sygnału określonego rodzaju;
- typy semantyczne, dotyczące znaczenia obiektów, relacji między nimi i zrozumiałych zdarzeń świata rzeczywistego (naturalnego) zapisanych w sygnale;
- typy strukturalne, normujące organizację metadanych w struktury danych i ich kolekcje;
- typy kontrolne, nawiązujące do informacji niezbędnych w zarządzaniu i udostępnianiu medium cyfrowego, takich jak: kontekst techniczny (typ nośnika, typ kompresji, format i rozmiar pliku), kontekst wytwórczy (autor, producent, typ narzędzia wytwórczego, klasyfikacja produktu, np. muzyka filmowa), kontekst dystrybucyjny (dystrybutor, prawa własności, koszty).

Szereg technik automatycznej ekstrakcji cech z obrazów, ich sekwencji oraz z materiału dźwiękowego określono w dokumentach wideo i audio tego standardu.

Deskryptory wizualne

Wśród deskryptorów wizualnych można wyróżnić kilka zasadniczych grup:

- opis struktur podstawowych
  - rozmieszczenie siatki (*Grid Layout*) – kompozycja deskryptorów wizualnych obliczanych niezależnie w siatce bloków pokrywających dany obraz,
  - widok złożony w 2W-3W (*2D-3D Multiple View*) – kompozycja deskryptorów wizualnych obliczanych niezależnie dla wielu obrazów 2W lub 3W tej samej sceny,
- deskryptory koloru
  - dominujące kolory (*Dominant Colors*) – informacje o dominujących kolorach, ekstrahowane z obiektu wizualnego (ramka, region ramki, sekwencja ramek, obszar czasowy),

  - skalowalny kolor (*Scalable Color*) – informacja o histogramie kolorów skalowalna ze względu na liczbę poziomów kwantyzacji koloru w przestrzeni HSV (*Hue, Saturation, Value*),
  - rozmieszczenie koloru (*Color Layout*) – informacja o współczynnikach DCT w blokach ramki dla składowych Y (luminancja) oraz Cb i Cr (chrominancja),
  - struktura koloru (*Color Structure*) – informacja o strukturalnym histogramie kolorów w przestrzeni kolorów HMMD (*Hue, Min, Max, Difference*),

- deskryptory tekstury
  - jednolitość tekstury (*Homogenous Texture*) – informacja o teksturze jednorodnej bazująca na analizie spektrum transformacji Radona z wykorzystaniem zestawu filtrów Gabora,
  - przegląd tekstury (*Texture Browsing*) – zwarta informacja o teksturze na podstawie analizy kierunków dominujących w amplitudowym spektrum transformaty Gabora obrazu oryginalnego,
  - histogram krawędzi (*Edge Histogram*) – histogram dla elementarnych typów krawędzi liczonych w różnych konfiguracjach bloków,

- deskryptory kształtu
  - kształt regionu (*Region Shape*) – informacja o kształcie obszaru na podstawie momentów Zernicke (transformacja ART),
  - kształt konturu (*Contour Shape*) – informacja o kształcie konturu na podstawie analizy CSS (transformacja do skalowalnej przestrzeni krzywizny),
  - trójwymiarowe widmo kształtu *3D Shape Spectrum*)

- deskryptory ruchu:
  - ruch kamery (*Camera Motion*) – informacja o parametrach ruchomej kamery,
  - tor ruchu (*Motion Trajectory*) – informacja o trajektorii ruchu obiektu interpolowanej funkcją kawałkami wielomianową, co najwyżej stopnia drugiego,
  - aktywność ruchu (*Motion Activity* – aktywność ruchu mierzona na podstawie estymacji ruchu w makroblokach 16×16.

Warto zwrócić uwagę przede wszystkim na trzy zdefiniowane w MPEG-7 deskryptory tekstury obiektów wizyjnych [218, 219]:

- Homogeneous Texture Descriptor – HTD, który służy do opisu jednorodnych tekstur obiektów wizyjnych
- Texture Browsing Descriptor – TBD, który opisuje percepcyjne właściwości postrzegania tekstur
- Edge Histogram Descriptor – EHD, który nie opisuje regionów o danych teksturach, lecz rozkład i kierunkowość krawędzi pomiędzy regionami o różnych teksturach w obrazie

Rysunek 5.14: Pięc typów krawędzi ekstrahowanych w EHD

Edge Histogram Description (Histogram krawędzi) Histogram krawędzi opisuje rozkład przestrzenny obrazu za pomocą pięciu typów krawędzi pokazanych na rysunku 5.14: czterech kierunkowych i jednym bezkierunkowym. Ostatni typ krawędzi nie jest określony a priori w żaden sposób i jest częścią obrazu. Zostaje on z niego wydzielony za pomocą procedury pokazanej na rysunku 5.15. Obraz wejściowy jest dzielony na bloki, a następnie pod–bloki, w obrębie których wykonywana jest operacja detekcji krawędzi.

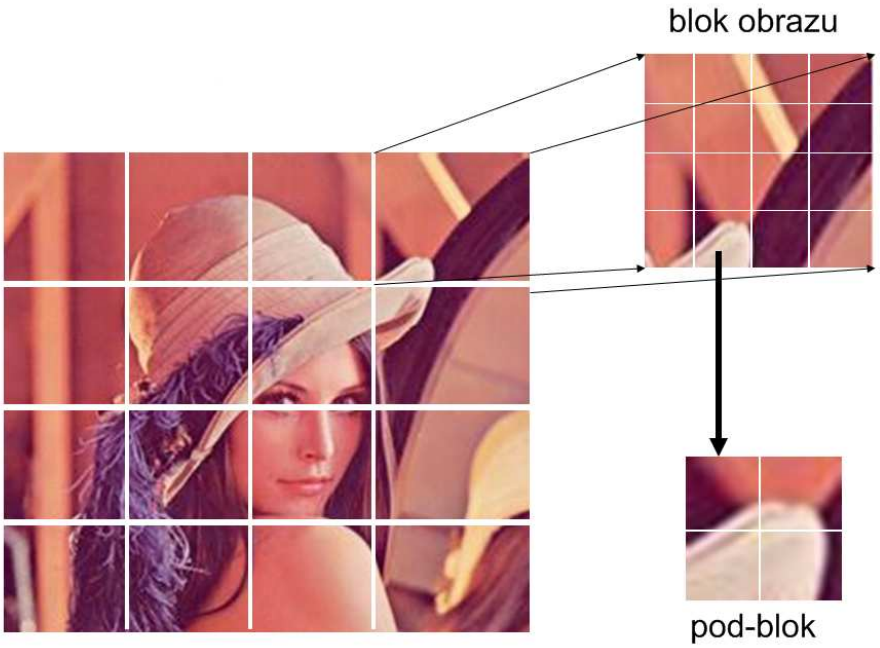
Krawędzie odgrywają ważną rolę w postrzeganiu obrazów, szczególnie obrazów naturalnych, dlatego też powyższy deskryptor uważany jest za najważniejszy przy porównywaniu grup obrazów za pomocą wzorca [220].

Homogeneous Texture Descriptor (Deskryptor homogennych tekstur) Deskryptor homogennych tekstur służy do opisu jednorodnych tekstur obiektów wizyjnych. Metoda wykorzystywana do ekstrakcji tego deskryptora została zaproponowana w pracy [221] i polega na wyznaczeniu parametrów statystycznych każdego z kanałów dwuwymiarowego widma obrazu statycznego [219]. Podział widma na rozłączne podpasma odbywa się przy użyciu zastawu filtrów zamodelowanych funkcją Gabora:

$$G_{s,r}(\omega,\theta) = \exp[\frac{-(\omega-\omega_s)^2}{2\sigma_s^2}]\exp[\frac{-(\theta-\theta_r)^2}{2\tau_s^2}] \tag{5.7}$$

gdzie s, r to indeksy podpasm, ω, θ to argumenty pulsacji i kąta, natomiast σ, τ to odchylenie standardowe pulsacji i kąta.

Deskryptor HTD dzieli widmo obrazu na 30 kanałów o sześciu kierunkach orientacji i pięciu zakresach częstotliwości, tak jak pokazano na rysunku 5.16.



Rysunek 5.15: Zasada składania bloków i definiowanie pod–bloków w EHD

Dla każdego z kanałów wyznaczana jest średnia energia składników widma oraz ich wariancja. Składnia deskryptora HTD jest następująca [219, 218]:

$$HTD = [f_{DC}, f_{SD}, e_1, \ldots e_{30}, d_1, \ldots d_{30}] \tag{5.8}$$

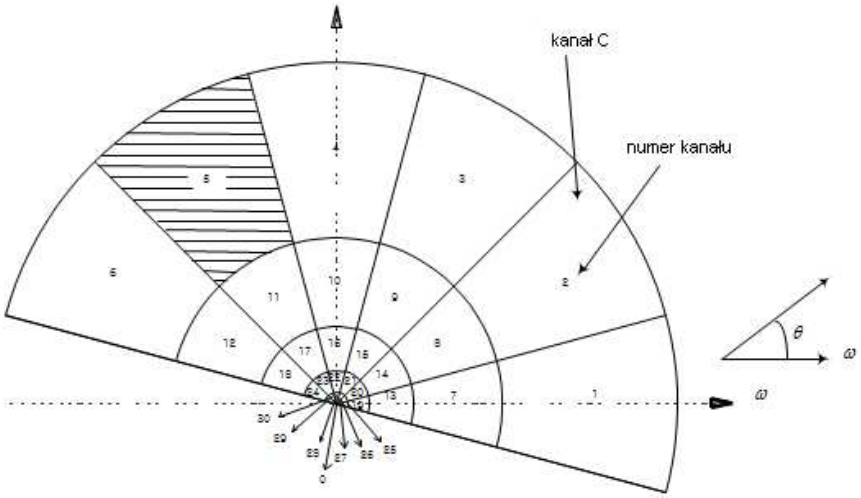
gdzie $[f_{DC}$ to jednorodnie skwantowana wartość średniej luminancji obrazu, f_{SD} to jednorodnie skwantowana wartość odchylenia standardowego luminancji punktów obrazu od średniej, e_i to średnia wartość energii i–tego kanału, natomiast d_i to wartość dewiacji energii i–tego kanału.

Texture Browsing Descriptor (Percepcyjny deskryptor tekstury) Percepcyjny deskryptor tekstury opisuje tekstury obiektów wizyjnych poprzez wyznaczenie parametrów odpowiadających ludzkiemu sposobowi postrzegania [2]. Składnia opisu deskryptora TBD jest następująca:

$$TBD = [v_1, v_2, v_3, v_4, v_5] \tag{5.9}$$

gdzie v_1 to parametr regularności bądź struktury tekstury (*regularity*), v_2, v_3 to parametry opisujące kierunkowość tekstury *(directionality)*, natomiast v_4, v_5 to parametry opisujące ziarnistość tekstury (*ang. coarseness*).

Algorytm ekstrakcji cech TBD został przedstawiony w [218, 222] i także opiera się na filtracji widma z użyciem funkcji Gabora (wzór 5.7).



Rysunek 5.16: Podział widma na kanały dla obliczeń deskryptora HTD. Rysunek zaczerpnięto z [218]

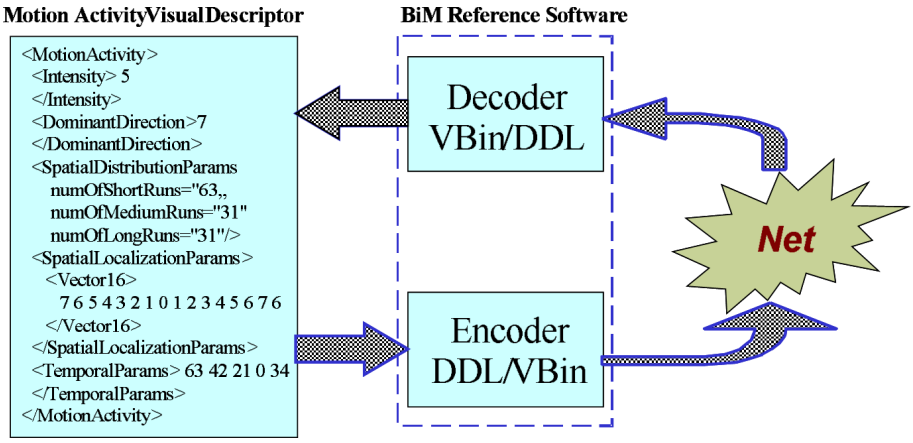
Deskryptor TBD ma na celu wyznaczenie opisu, który bardziej odpowiada percepcyjnym właściwościom postrzegania człowieka. Nie gwarantuje to jednak uzyskiwania zadowalających rezultatów wyszukiwania [219]. Liczba parametrów deskryptora jest ograniczona do pięciu (wzór 5.9). Dodatkowo wartości poszczególnych parametrów są silnie kwantowane, co zapewnia bardzo ograniczoną objętość bitową deskryptora, co niekorzystnie wpływa na precyzję wyszukiwania z użyciem TBD i ogranicza zakres jego zastosowań [219]. Dodatkowo, deskryptor TBD charakteryzuje się bardzo dużą złożonością obliczeniowa ekstrakcji opisu, co dyskwalifikuje zastosowanie tego deskryptora w aplikacjach działających w czasie rzeczywistym.

Deskryptory rozpoznawania twarzy Zdefiniowano także specyficzne, bardziej zaawansowane deskryptory, m.in. rozpoznawania twarzy:

- deskryptor rozpoznawania twarzy (*Face Recognition*) – informacja o obrazie twarzy uzyskana na podstawie skwantowanych do 5 bitów pierwszych 48 współczynników transformaty KLT,

- zaawansowany deskryptor rozpoznawania twarzy (*Advanced Face Recognition*) – informacja o obrazie twarzy uzyskana w wyniku hierarchicznej liniowej analizy dyskryminacyjnej LDA, dokonanej na informacji spektralnej globalnej i lokalnej w blokach, z opcjonalną wstępną normalizacją twarzy do pozy frontalnej.

Deskryptory wizualne oprócz formatu XML mają także zdefiniowaną formę binarną, bardziej upakowaną niż binarna postać BiM – w oprogramowaniu referencyjnym zamieszczono kodeki postaci XML na postać binarną (rys. 5.17.



Rysunek 5.17: Schemat blokowy kodeków deskryptorów wizualnych MPEG-7, zaczerpnięty z [27].

Deskryptory audio

Wśród deskryptorów opisujących dźwięk i mowę warto wymienić przede wszystkim:

- sygnatura audio (*Audio Signature*) – skalowalna informacja o lokalnych statystykach widma dźwięku,
- deskryptory brzmienia instrumentów (*Musical Instrument Timbre*) – szereg deskryptorów charakteryzujących brzmienie instrumentu w terminach bazowych charakterystyk widmowych, takich jak centroid widma harmonicznego, jego rozproszenie, itp.
- deskryptor melodii (*Melody*) – złożony deskryptor zawierający między innymi sygnaturę czasową i linię melodyczną w postaci ciągu zmian melodycznych.
- deskryptory rozpoznawania i indeksowania dźwięku (*General Sound Recognition and Indexing*) – szereg deskryptorów pozwalających dokonać rozróżnienia dźwięków na poziomie ogólnym, np. pomiędzy muzyką, mową, a szumem, lub bardziej szczegółowym, np. między głosem mężczyzny, kobiety i dziecka (podstawowym modelem jest ukryty łańcuch Markowa).

5.2.4 MPEG-21, czyli integracja

Obecnie ostatni z serii, intensywnie rozwijany MPEG-21 definiuje narzędzia do opakowania zarówno danych, jak i metadanych multimedialnych w celu sprawnej ich zintegrowanej dystrybucji na nośnikach pamięci i w sieci rozległej. Zasadniczą wizją jest zapewnienie przezroczystego i możliwie poszerzonego wykorzystania zasobów multimedialnych dostępnych w sposób niemal nieograniczony poprzez zintegrowanie różnego typu sieci i urządzeń oraz współpracę różnych społeczności w skali globalnej. Chodzi o zbudowanie na tych warunkach infrastruktury dla dostarczania i konsumpcji multimediów poprzez integracje rozwiązań juz istniejących oraz wskazanie potrzeb opracowania nowych standardów porządkujących współdziałanie.

Norma MPEG-21 zawiera uniwersalne mechanizmy definiowania praw odnoszących się do materiału multimedialnego REL (*Rights Expression Language*) oraz koncepcje adaptacji i przetwarzania obiektów cyfrowych: DIA (*Digital Item Adaptation*) i DIP (*Digital Item Processing*). W standardzie występują obiekty cyfrowe (*digital items*), złożone z materiałów cyfrowych i ich opisów, przez które użytkownicy realizują różne formy interakcji. Struktura obiektu cyfrowego opisana jest dokumentem XML, który nosi nazwę deklaracji obiektu cyfrowego.

Przykładem multimedialnego obiektu cyfrowego jest film wideo wraz z zestawem ścieżek dźwiękowych, opisem tekstowym, mechanizmami interakcji, zdjęciami z procesu produkcji, opiniami o filmie, wywiadami z aktorami, itp.

5.3 Podsumowanie

Do najistotniejszych zagadnień poruszonych w tym rozdziale należy charakterystyka standardów multimedialnych rodzin JPEG i MPEG, ze szczególnym zwróceniem uwagi na proces ich doskonalenia, zakres zastosowań oraz stosowne rozwiązania algorytmiczne.

Warto zwrócić uwagę przede wszystkim na algorytmy kompresji JPEG, JPEG2000 oraz MPEG-2 i H.264, specyficzne ich dostosowanie do charakterystyki odbiorcy informacji, wręcz nieograniczonego zakresu definiowania danych źródłowych oraz wymagań współczesnych systemów przekazu informacji wielostrumieniowej. Ważne jest ponadto zrozumienia zasad indeksowania oraz ocena skuteczności procesu wyszukiwania danych po zawartości.

Zadania do tego rozdziału podano na stronie 371.

Ćwiczenie pozwalające na eksperymentalną weryfikację wybranych realizacji standardów multimedialnych zamieszczono na stronie 393.

Bibliografia

[1] A. Puri, T. Chen (2000) Multimedia systems, standards, and networks. Marcell Dekker, Inc., Nowy Jork.

[2] K. Keeler, J. Westbrook (1995) 'Short encodings of planar graphs and maps. Disc. Appl. Math. pp. 239-252.

[3] M. Deering (1995) Geometry compression. In SIGGRAPH'95 Conference Proceedings, pp. 13-20.

[4] K. Hosaka (1986) A new picture quality evaluation method. Proc Int Picture Coding Symposium, Tokyo, Japan.

[5] A.M. Eskicioglu, P.S. Fisher (1995) Image quality measures and their performance. IEEE Tran Comm 43(12):2959–65.

[6] J.A. Swets (1979) ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiology 14:109–121.

[7] M.P. Sampat, M.K. Markey, A.C. Bovik (2005) Computer-aided detection and diagnosis in mammography. Handbook of Image and Video Processing (2nd edition):1195–1217.

[8] H. Li, K.J.R. Liu, S.C.B. Lo (1997) Fractal modeling and segmentation for the enhacement of microcalcifications in digital mammograms. IEEE Trans Med Imag 16(6):785–98.

[9] W. Kwiatkowski (2010) Wprowadzenie do kodowania. BEL Studio, Warszawa.

[10] E.R. Fossum (1997) CMOS image sensors: electronic camera-on-a-chip. IEEE Tran Elec Dev 44(10):1689–98.

[11] S. Królewicz (2005) Charakterystyka wybranych cech współczesnych średnio- i wysokorozdzielczych danych teledetekcyjnych, [w:] Biskupin ... i co dalej? Zdjęcia lotnicze w polskiej archeologii, (red.) J. Nowakowski, A. Prinke, W. Rączkowski. Poznań, Instytut Prahistorii UAM.

[12] A. Smailbegovic, J.V. Taranik, F. Kruse (2000) Importance of spatial and radiometric resolution of AVIRIS data for recognition of mineral endmembers in the Geiger Grade Area, Nevada, U.S.A. Proceedings of the Nirth JPh airborne earth science workshop.

[13] T. Krzymień, Z. Kulka, M. Moraszczyk, D. Sawicki, T. Smakuszewski, K. Wnukowicz oraz koordynatorzy: W. Skarbek, R. Rak (2004) Wstęp do inżynierii multimediów. Seria: Akademickie podręczniki akademickie. Politechnika Warszawska, stron 217.

[14] R.V. Mayorga (2003) Towards computational sSapience (wisdom): a paradigm for sapient (wise) systems. Proc IEEE Integration of Knowledge Intensive Multi-Agent Systems, Boston, USA, pp. 158–166.

[15] L. Floridi (2007) In defence of the veridical nature of semantic information. Eur J Anal Philosophy 3:31–41.

[16] Y. Bar-Hillel, R. Carnap (1953) An outline of a theory of semantic information. Przedrukowane w Y. Bar-Hillel Bar-Hillel (1964) Language and information. Reading, Mass., London, Addison-Wesley, pp. 221–274.

[17] A.N. Kolmogorov (1956) Some fundamental problems in the approximate and exact representation of functions of one or several variables. Proc III. Math Congress USSR 2:28–29. MCU Press, Moskwa.

[18] A.N. Kolmogorov, V.M. Tikhomirov (1959) ϵ-entropy and ϵ-capacity. Uspekhi Mat. Nauk 14:3–86 (Engl Transl. Amer. Math. Soc. Transl. 2(17):277–364).

[19] R.M. Gray (1990) Entropy and information theory. Springer-Verlag, New York.

[20] A. Przelaskowski (2010) Uproszczone wspomaganie obrazowej diagnostyki medycznej. Przegląd telekomunikacyjny LXXXIII(12):1748–1755.

[21] C.E. Shannon (1948) A mathematical theory of communications', Bell System Technical Journal, 27:379-423, 623-656.

[22] S. Sahni, B.C. Vemuri et al. (1998) State of the art lossless image compression algorithms. IEEE Proc International Conference on Image Processing, Chicago, Illinois, pp. 948–952.

[23] W. Skarbek (1993) Metody reprezentacji obrazów cyfrowych. Akademicka Oficyna Wydawnicza PLJ, Warszawa.

[24] M. Rabbani, P.W. Jones (1991) Digital image compression techniques. SPIE Optical Engineering Press, TT 7, Bellingham, Washington, USA.

[25] N.S. Chang, K.S. Fu (1979) A relational database system for images. Purdue University. TR-EE 7928.

[26] V.N. Vapnik. A.Y. Chervonenkis (1971) On the uniform convergence of relative frequencies of events to their probabilities. Theory of probability and its applications 16:264–80.

[27] W. Skarbek (2004) Podstawy multimediów. Politechnika Warszawska, Warszawa.

[28] T. Deselaers (2002) Features for image retrieval. PhD thesis, Aachen Univeristy.

[29] H.A. Elsalamony (2010) Automatic video stream indexing and retrieving based on face detection using wavelet transformation. Proc IEEE 2nd Int Conf Sig Proces Sys (ICSPS) pp. V1-153 – V1-157.

[30] J. P. Eakins, M. E. Graham (2000) Content-based image retrieval. Technical report, JISC Technology Application Program.

[31] B. Bin Zheng (2009) Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives. Algorithms 2(2):828–849.

[32] P.Welter, C. Hockena, T.M. Desernoa, C. Grouls (2009) Workflow management of content-based image retrieval for CAD support in PACS environments based on IHE. JCAR 5(4):393–400.

[33] P. Boniński (2008) Metody indeksowania obrazów medycznych na potrzeby radiologii cyfrowej, rozprawa doktorska, Politechnika Warszawska.

[34] M.C. Oliveira, W. Cirne, P.M. de Azevedo Marques (2007) Towards applying content-based image retrieval in the clinical routine. Future Generation Computer Systems 23(3):466–474 .

[35] T. Pfund, S. Marchand-Maillet (2002) Dynamic multimedia annotation tool. In G. Beretta and R. Schettini, editors, Proc SPIE Photonics West Conference on Internet Imaging III, pp. 216–224.

[36] C. Carson, J. Hellerstein, J. Malik (1999) Blobworld: A system for region-based image indexing and retrieval. Proc Third International Conference On Visual Information Systems (VISUAL'99), pp. 509–516.

[37] R.C. Veltkamp, M. Tanase (2002) Content-based image retrieval systems: a survey. Technical Report UU-CS-2000-34 (revised October 28, 2002).

[38] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain (2000) Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Machine Intel 22(12):1349–1380.

[39] H. Eidenberger (2004) Visual Information Retrieval. PhD thesis, Technischen Universitat Wien.

[40] D. Heesch, S. Ruger (2003) Performance boosting with three mouse clicks - relevance feedback for cbir. Proc 25th European Conf on Information Retrieval Research, pp. 363–376, Springer.

[41] K. Sklinda, P. Bargiel, A. Przelaskowski, T. Bulski, J. Walecki, P. Grieb (2007) Multiscale extraction of hypodensity in hyperacute stroke. Medical Science Monitor 13(Suppl 1):5–10, Proc XXXVIII Congress of the Polish Medical Society of Radiology.

[42] D. Comaniciu, P. Meer, D. Foran, A. Medl (1998) Bimodal system for interactive indexing and retrieval of pathology images. Proc Fourth IEEE Workshop on Applications of Computer Vision (WACV'98), pp. 76–81.

[43] S.T. Perry, P.H. Lewis (1998) A novel image viewer providing fast object delineation for content based retrieval and navigation. Proc SPIE Conference on Storage and Retrieval for Image and Video Databases VI, pp. 436–445.

[44] A. Winter and C. Nastar (1999) Differential feature distribution maps for image segmentation and region queries in image databases. Proc IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'99), pp. 9–17.

[45] G. Ciocca, R. Schettini (1999) Similarity retrieval of trademark images. Proc Int Con Image Anal Proces.

[46] R.G. Kuehni (2002) The early development of the Munsell system. Color Research and Application 27(1): 20-–27.

[47] D.M. Squire, W. Müller, H. Müller, T. Pun (2000) Content-based query of image databases: inspirations from text retrieval. Pattern Recognition Letters 21:1193–1198.

[48] C. Carson, S. Belongie, H. Greenspan, J. Malik (1997) Region-based image querying. Proc IEEE Con Comp Vis Patt Recog (CVPR'97), pp. 42–51.

[49] W. Niblack, R. Barber et al. (1993) Qbic project: querying images by content, using color, texture and shape. In W. Niblack (editor) Proc SPIE Conf Storage and Retrieval for Image and Video Databases, pp. 173-—187.

[50] S. Sclaroff, L. Taycher, M. La Cascia (1997) Imagerover: a content-based browser for the world wide web. Proc IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 2–9.

[51] T. Gevers, A.W.M. Smeulders (1996) A comparative study of several color models for color image invariants retrieval. Proc First International Workshop ID-MMS'96, pp. 17-—26.

[52] J. M. Geusebroek, R. van den Boogaard, A.W.M. Smeulders, H. Geerts (2001) Color invariance. IEEE Tran Pattern Anal Machine Intel 23(12):1338—-1350.

[53] C. Vertan, N. Boujemaa (2000) Using fuzzy histograms and distances for color image retrieval. Proc Challenge of Image Retrieval (CIR),pp. 85–89.

[54] S. Siggelkow (2002) Feature histograms for content-based image retrieval. PhD thesis, Albert-Ludwigs-Universit at Freiburg.

[55] L.A. Zadeh (1965) Fuzzy sets. Information and Control 8(3):338-–353.

[56] J. Han, K.-K. Ma (2002) Fuzzy color histogram and its use in color image retrieval. IEEE Tran Im Proc 11(8):944–952.

[57] C. Vertan, N. Boujemaa (2000) Embedding fuzzy logic in content based image retrieval. Proc. 19th Int'l Meet North Am Fuzzy Inf Proc Soc (NAFIPS), pp. 85–89.

[58] H. Muller, N. Michoux, D. Brandon, A. Geissbuhler (2004) A review of content-based image retrieval systems in medical applications — clinical benefits and future directions. Int J Medical Informatics 73(1):1–23.

[59] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, T.S. Huang (1998) Supporting ranked boolean similarity queries in mars. IEEE Trans Knowledge Data Eng 10(6):905—-925.

[60] J. Ze Wang, G. Wiederhold, O. Firschein, S. Xin Wei (1997) Wavelet-based image indexing techniques with partial sketch retrieval capability. Proc Fourth Forum on Research and Technology Advances in Digital Libraries, pp. 13-—24.

[61] W. Ma, B. Manjunath (1996) Texture features and learning similarity. Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR'96), pp. 425–430.

[62] S. Santini, R. Jain (1996) Gabor space and the development of preattentive similarity. Proc 13th International Conference on Pattern Recognition (ICPR'96), pp. 40—44.

[63] R. Milanese, M. Cherbuliez (1999) A rotation, translation and scale-invariant approach to content-based image retrieval. J Visual Commun Image Represent 10:186–196.

[64] C.-R. Shyu, C.E. Brodley et al. (1999) Assert: A physician-in-the-loop content-based retrieval system for hrct image databases. Comput Vis Image Understand 75(1-2):111–132.

[65] J.S. Weszka, C.R. Dyer, A. Rosenfeld (1976) A comparative study of texture measures for terrain classification. IEEE Trans Sys Man Cybernetics 6(4):269–285.

[66] W.-J. Kuo, R.-F. Chang, C.C. Lee, D.-R. Chen W.K. Moon (2002) Retrieval technique for the diagnosis of solid breast tumors on sonogram. Ultrasound Med Biol 28(7):903–909.

[67] R. Baeza-Yates, B. Ribeiro-Neto (1999) Modern Information Retrieval. Addison Wesley.

[68] Manuel Castells on the Network Society: http://www.tidec.org/geovisions/Castells.html

[69] J.Brockman (1995) The Third Culture: Beyond the Scientific Revolution. Simon & Schuster.

[70] T.H. Eriksen (2003) Tyrania chwili. PIW.

[71] A. Przelaskowski (2005) Kompresja danych: podstawy, metody bezstratne, kodery obrazów. Wydawnictwo BTC, Warszawa.

[72] N. Sloane, A.D. Wyner (eds.)(1993) Claude Elwood Shannon : collected papers (New York).

[73] J. Ziv, A. Lempel (1977) A universal algorithm for sequencial data compression., IEEE Trans Inf Theory 23(3):337–343.

[74] J. Ziv, A. Lempel (1978) Compression of individual sequences via variable-rate coding. IEEE Trans Inf Theory 24(5):530–536.

[75] J.A. Storer, T.G. Szymanski (1982) Data compression via textual substitution. J ACM 29:928–951.

[76] T. Welch (1984) A technique for high-performance data compression. IEEE Computer 17(6):8–19.

[77] G. Langdon (1984) An introduction to arithmetic coding. IBM J Res Dev 28(2):135–149.

[78] G. Langdon, C. Haidinyak (1994) Context-dependent distribution shaping and parametrization for lossless image compression. Proc SPIE 2298 (Applications of Digital Image Processing XVII):62–70.

[79] I. Witten, R. Neal, J. Cleary (1987) Arithmetic coding for data compression. Comm ACM 30(6):520–540.

[80] P.G. Howard, J.S. Vitter (1994) Arithmetic coding for data compression. Proc IEEE, 82(6):857–865.

[81] P.G. Howard, J.S. Vitter (1993) Fast and efficient lossless image compression. Proc IEEE Data Compression Conference, pp. 351–360.

[82] A. Moffat, R.M. Neal, I.H. Witten (1998) Arithmetic coding revisited: a guided tour from theory to practice. ACM Trans Inf Sys 16(3):256–294.

[83] J.M. Shapiro (1993) Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans Sig Proc 41(12):3445–3462.

[84] C.C. Cutler (1952) Differential quantization for television signals. U.S. Patent 2,605,361.

[85] T. Endoh, T. Yamazaki (1986) Progressive coding scheme for multilevel images. Proc Picture Coding Symposium, pp. 21–22.

[86] P. Roos, M.A. Viergever, M.C.A. van Dijke, J.H. Peters (1988) Reversible intraframe compression of medical images, IEEE Trans Med Imag 7(4):328–336.

[87] P. Roos, M.A. Viergever (1991) Reversible interframe compression of medical images: a comparison of decorrelation methods". IEEE Trans Med Imag 10(4):538–547.

[88] X. Wu (1996) Lossless compression of continuous -tone images via context selection and quantization. IEEE Trans Im Proc 5(6):656–664.

[89] M. Weinberger, G. Seroussi, G. Sapiro (2000) The LOCO-I lossless image compression algorithm: principles and standarization into JPEG-LS. IEEE Trans Im Proc 9(8):1309–1324.

[90] ISO/IEC 15444 International Standard (JPEG2000),Information technology - JPEG 2000 image coding system, 2000.

[91] ITU-T Rec. H.264ISO/IEC 14496-10 International Standard (MPEG-4:10), Klagenfurt, AT, Information Technology - Coding of audio-visual objects - Part 10: Advanced video coding (FDIS), 2003.

[92] ISO/IEC 11544 International Standard (JBIG), Information Technology - Coded representation of picture and audio information - Progressive bi-level image compression, 1993.

[93] ISO/IEC 14492 International Standard (JBIG2), Information Technology - Lossy/lossless coding of bi-level Images, 2001.

[94] ISO/IEC JTC 1/SC 29/WG 1, JPEG LS image coding system, ISO Working Document ISO/IEC JTC 1/SC 29/WG 1 N399 - WD14495, June 1996.

[95] L. Bottou, P.G. Howard, Y. Bengio (1998) The Z-coder adaptive binary coder. Proc IEEE Data Compression Con, pp. 13–22.

[96] Graphics Interchange Format (1990) version 89a, CompuServe Incorporated, http://www.w3.org/Graphics/GIF/spec-gif89a.txt

[97] T. Boutell, A. Dilger, et al: PNG (Portable Network Graphics) specification, www.w3.org/TR/REC-png-multi.html

[98] A. Luthra, P. Topiwala (2003) Overview of the H.264/AVC video coding standard. Proc SPIE vol. 5203, Applications of Digital Image Processing XXVI, pp. 417–431.

[99] S.P. Lloyd (1982) Least squares quantization in PCM. IEEE Tran Inform Theory IT-28:129–137. Jest to reprodukcja manuskryptu raportu technicznego Bell Laboratories z 1957 roku.

[100] J. Max (1960) Quantizing for minimum distortion. IRE Tran Inform Theory IT-6:7–12.

[101] D.L. Donoho, M. Vetterli, R.A. DeVore, I. Daubechies (1998) Data compression and harmonic analysis. Invited paper, IEEE Trans. Inform. Theory, Special Issue, Inform. Theory: 1948-1998 Commemorative Issue 44(6):2435–2476.

[102] A. Moffat (1990) Implementing the PPM data compression scheme. IEEE Trans Comm COM-38 11:1917-1921.

[103] F.M. Willems, Y.M. Shtarkov, T.J. Tjalkens (1995) The context-tree weighting method: basic properties. IEEE Trans. Information Theory, IT-41, pp. 653-664.

[104] Lossless data compression software benchmarks/comparisons: http://www.maximumcompression.com/

[105] SPMG/JPEG-LS V.2.2 codec, ftp://spmg.ece.ubc.ca/pub/jpeg-ls/ver-2.2/, Signal Processing & Multimedia Group at the University of British Columbia, USA.

[106] http://www.cl.cam.ac.uk/m̃gk25/jbigkit/

[107] X. Wu, CALIC codec, ftp://ftp.csd.uwo.ca/pub/from_wu/v.arith/

[108] A. Przelaskowski (2004) Binarny koder obrazów ze skalą szarości. Prace Naukowe Instytutu Telekomunikacji i Akustyki Politechniki Wrocławskiej, nr 85, seria: Konferencje, nr 29, X Sympozjum Nowości w Technice Audio i Wideo, str. 243-252.

[109] http://www.msoftware.co.nz/downloads_page.php

[110] S.G. Chang,B. Yu, M. Vetterli (2000) Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Proces 9:1532–1546.

[111] G.E. Sarty, M.S. Atkins, R.A. Pierson (1998) Sharper MRI of ovarian and uterine masses via wavelet-based compression. 20th Annual Canadian Western Society for Reproductive Biology Workshop.

[112] P. Brownrigg, P.R.G. Bak (2005) What is the legal risk in using irreversible compression on diagnostic images in medical management? - a legal assessment and review of the regulatory framework in the USA and Canada. SCAR 2005 Annual Meeting, Orlando, USA.

[113] W. Kopaliński (2004) Słownik wyrazów obcych i zwrotów obcojęzycznych. Państwowe Wydawnictwo Naukowe.

[114] P. Bargieł (2007) Komputerowe metody poprawy jakości medycznych danych obrazowych. PRozprawa doktorska, Politechnika Warszawska.

[115] G. Galinski, W. Skarbek (2006) Wyszukiwanie obrazów z wykorzystaniem kolorów dominujących. Proc V Sympozjum Naukowe Techniki Przetwarzania Obrazu, Serock, pp. 406–412.

[116] T. Deselaers, Daniel D. Keysers, H. Ney (2004) Features for image retrieval: a quantitative comparison. Lecture Notes in Computer Science, 26th DAGM Symposium on Pattern Recognition.

[117] L. Lucchese, S.K. Mitra (1999) Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane. Proc IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'99), pp. 74–78.

[118] H. Tamura, S. Mori, T. Yamawaki (1978) Texture features corresponding to visual perception. IEEE Trans Systems, Man and Cybernetics 8(6):460–473.

[119] H. Tamura, N. Yokoya (1984) Image database systems: a survey. Pattern Recognition 17(1):29–43.

[120] T. Deselaers, D. Keysers, H. Ney (2004) Fire - flexible image retrieval engine: Imageclef 2004 evaluation. In CLEF 2004.

[121] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, W. Equitz (1994) Efficient and effective querying by image content. J Intellig Inf Sys 3(3/4):231–262.

[122] B. Terhorst (2003) Texturanalyse zur globalen bildinhaltsbeschreibung radiologischer. Technical report, Institut fur Medizinische Informatik.

[123] B.B. Chaudhuri, N. Sarkar (1995) Texture segmentation using fractal dimension. PAMI 17(1):72–77.

[124] W. Niblac, Barber et al (1994) The qbic project querying images by content using color, texture and shape. Proc SPIE Storage and Retrieval for Image and Video Databases.

[125] W. Equitz and W. Niblack (1994) Retrieving images from a database using texture algorithms from the qbic system. IBM Research Report.

[126] M. Flickner, H. Sawhney et al. (1995) Query by image and video content: The qbic system. IEEE Comput Mag 28:23–32.

[127] J. R. Bach, C. Fuller et al. (1996) The virage image search engine: An open framework for image management. Proc SPIE Storage and Retrieval for Image and Video Databases.

[128] P. Boniński, A. Przelaskowski (2006) Local wavelet features for content-based image retrieval in medical domain. Proc NTIAV 2006 Conference.

[129] Y.A. Aslandogan, C.T.C. Yu, C. Liu, K.R.Nair (1995) Block Design, implementation and evaluation of score (a system for content based retrieval of pictures). Proc 11th Int Con Data Engineering, pp. 280–287. IEEE Computer Society.

[130] A. Pentland, R. W. Picard, S. Sclaroff (1996) Photobook: Content-based manipulation of image databases. Int J Comp Vis 18(3):233–254.

[131] S. Marcus, V.S. Subrahmanian (1996) Foundations of multimedia database systems. J ACM 43(3).

[132] W.-S Li, K.S. Candan (1998) A hybrid objectbased image database system and its modeling, language, and query processing. Proc 14th International Conference on Data Engineering.

[133] J.Z. Li, M.T. Özsu, D. Szafron, V. Oria (1997) Moql: a multimedia object query language. Proc 3rd International Workshop on Multimedia Information Systems.

[134] Nowa Encyklopedia Powszechna (1996) t.4 str. 647, Wydawnictwo Naukowe PWN. Warszawa.

[135] Materiały B. Smitha umieszczone na stronie http://ontology.buffalo.edu/

[136] T. Gruber (1993) A transactional approach to portable ontologies. Knowledge Acquisition 5(2):199–221.

[137] T.R. Gruber, G. Olsen (1994) An ontology for engineering mathematics, Fourth International Conference on Principles of Knowledge Representation and Reasoning, Bonn, Morgan, Kaufmann Publishers, San Francisco, California, pp. 258–269.

[138] T. Gruber, What is ontology – http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

[139] W.N. Borst (1997) Construction of engineering ontologies for knowledge sharing and reuse, PhD thesis, University of Tweente Enschede, Nederlands.

[140] M.R. Gensereth, N.J. Nilson (1987) Logical foundations of artificial intelligence. Morgan Kaufman.

[141] N. Guariano, P. Giaretta, P. (1995) Ontologies and knowledge bases: towards a terminological clarification. In N.Mars (ed.) Towards very large knowledge bases: knowledge building and knowledge sharing, IOS Press, Amsterdam, pp 25–32.

[142] N. Guarino (1998) Formal ontology in information systems. Proc FOIS'98, IOS Press , pp. 3–15.

[143] A. Bassara (2004) I weź tu dogadaj się. Gazeta IT 1(20).

[144] B. Chandrasekaran, J.R. Josephson, V.R. Benjamins (1999) What are ontologies, and why we need them?. IEEE Inteligent Systems, pp 20–26.

[145] B. Swarout, P. Ramesh, K. Knight, T. Russ (1997) Toward distributed use of largesScale ontologies, Proc AAAI'97 Spring Symposium on Ontological Engineering, Stanford University, pp 138–148.

[146] A. Maedche (2002) Ontology learning for the semantic web. Boston: Kluwer Academic Publ.

[147] D.B. Lenat, R.V. Guha (1990) Building large knowledge based systems: representation and inference. In the CYC Project Reading Mass.: Addison-Wesley.

[148] G.A. Miller, R. Beckwith et al. (1990) WORDNET: an on line lexical database. Int J Lexicography 3-4:235–312.

[149] GUM – Generalized Upper Model, http://www.fb10.uni-bremen.de/anglistk/langpro/webspace/jb/gum/inc

[150] J.F. Sowa (1995) Top-level ontological categories. Int J Human-Computer Studies 43(5-6):669–685.

[151] A. Maedche, S. Staab (2002) Measuring similarity between ontologies. Proc CIKM, LNAI vol. 2473.

[152] N.N. Friedman, D.D. Hafner (1997) The state of the art in ontology design a survey and comparative review. AI Magazine 18(3):53–74.

[153] J. Baumeister, J. Reutelshoefer, F. Puppe (2011) Engineering intelligent systems on the knowledge formalization continuum. Int J Appl Math Comput Sci 21(1):27-–39.

[154] W.K. Pratt, Digital Image Processing, 3nd ed., Wiley & Sons, New York, 2001.

[155] A. Wróblewska (2008) Metody wspomagania detekcji zmian patologicznych w mammografii, rozprawa doktorska, Politechnika Warszawska.

[156] S.M. Pizer, E.P. Amburn el at. (1987) Adaptive histogram equalization and its variations. Comp Vis, Graph Im Proc 39:355–68.

[157] K. Zuiderveld (1994) Contrast limited adaptive histogram equalization. Graphics Gems IV, Academic Press.

[158] S. Osher, J.A. Sethian (1988) Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. Journal of Computational Physics 79:12–49.

[159] J.M. Prewitt (1970) Object enhancement and extraction. In B.S. Lipkin, A. Rosenfeld, eds. Picture Processing and Psychopictorics, Academic Press, New York.

[160] I. Sobel (1978) Neighborhood coding of binary images for fast contour following and general binary array processing. Computer Graphics Image Process 8(1):127–135.

[161] L.G. Roberts (1965) Machine perception of three-dimensional solids. In J. T. Tippet et al. Optical and Electro-Optical Information Processing, MIT Press, Cambridge, MA, pp. 159–97.

[162] R.A. Kirsch (1971) Computer determination of the constituent structure of biological images. Comput Biomed Res 4(3):315–28.

[163] F. Samopa, A. Asano (2009) Hybrid image thresholding method using edge detection. IJCSNS Int J Comp Sc Net Sec 9(4):292-99.

[164] W.F. Schreiber (1970) Wirephoto quality improvement by unsharp masking. J Pattern Recognition 2(2):111-–121.

[165] J. W. Tukey (1977) Exploratory data analysis, Addison-Wesley, Reading, MA.

[166] A.C. Bovik, T.S. Huang, D.C. Munson (1987) The effect of median filtering on edge estimation and detection. IEEE Trans Pattern Anal Machine Intell PAMI-9:181–94.

[167] W.K. Pratt, T.J. Cooper, I. Kabir (1985) Pseudomedian filter. Proc SPIE 534:34–43.

[168] H.P. Kramer, J.B. Bruckner (1975) Iterations of nonlinear transformation for enhancement of digital images. Pattern Recognition 7:53–58.

[169] J. Serra (1982,1988) Image analysis and mathematical morphology. Vol. I, vol. II. Academic Press.

[170] J.M. Lester, J.F. Brenner, W.D. Selles (1980) Local transforms for biomedical image analysis. Comp Graph Im Proc 13(1):17–30.

[171] J. Rogowska (2000) Overview and fundamentals of Medical Image Segmentation. In Handbook of Medical Imaging pp.69–81, Academic Press.

[172] S.M. Lai, X.Li, W.F. Bishof (1989) On techniques for detecting circumscribed masses in mammograms. IEEE Tran Med Im 8(4):377–386.

[173] J.W. Cooley, J.W. Tukey (1965) An algorithm for the machine calculation of complex Fourier series. Mathematics of Computation 19(90):297-–301.

[174] Y. Katznelson (1976) An introduction to harmonic analysis, Dover Books on Mathematics.

[175] M. Nieniewski (1998) Morfologia matematyczna w przetwarzaniu obrazów. Akademicka Oficyna Wydawnicza PLJ.

[176] J. Liu, W. Gao, S. Huang, W.L. Nowiński (2008) A model-bBased, semi-global segmentation approach for automatic 3-D point landmark localization in neuroimages. IEEE Tran Med Imag 27(8):1034–44

[177] L. Demaret, N. Dynb, A. Iske (2006) Image compression by linear splines over adaptive triangulations. Signal Processing 86:1604-–16.

[178] M. Kass, A. Witkin, D. Terzopoulos (1988) Snakes: active contour models. Int J Computer Vision 1(4):321–331.

[179] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham (1995) Active shape models - their training and application. Computer Vision and Image Understanding 61:38-–59.

[180] S. Osher, R. Fedkiw (2002) Level set methods and dynamic implicit surfaces. Applied Mathematical Sciences 153, Springer-Verlag, New York.

[181] S. Osher, N. Paragios (2003) Geometric level set methods in imaging, vision and graphics. Springer-Verlag, New York.

[182] I. Daubechies (1992) Ten lectures on wavelets. Society for industrial and applied mathematics.

[183] I. Daubechies (1993) Orthonormal bases of compactly supported wavelets II. Variations on a theme, SIAM Journal on Mathematical Analysis 24(2):499–519.

[184] G. Beylkin, B. Torresani (1996) Implementation of operators via filter banks, autocorrelation shell and Hardy wavelets. Applied and Computational Harmonic Analysis 3:164–185.

[185] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies (1992) Image coding using wavelet transform. IEEE Trans Image Proces IP-1:205–20.

[186] Y. Meyer (1986) Ondelettes, fonctions splines et analyses graduees. Wykłady na Uniwersytecie Turyńskim, Włochy.

[187] S. Mallat (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Tran Pat Anal Mach Intell 11:674–693.

[188] A. Przelaskowski (2002) Falkowe metody kompresji danych obrazowych. Prace naukowe, Elektronika z. 138, Oficyna Wydawnicza Politechniki Warszawskiej.

[189] Y. Zhang, M. Brady, S. Smith (2001) Segmentation of brain MR images through a hidden Markov random field model and the Expectation-Maximization algorithm. IEEE Trans Med Imag 20(1):45–57.

[190] M. Alfo, L. Nieddu, D. Vicari (2008) A finite mixture model for image segmentation. Stat Comput 18:137-–50.

[191] T. Adali, Y. Wang (2001) Image analysis and graphics for multimedia presentation. In Multimedia Image and Video Processing, L. Guan, S.-Y Kung, and J. Larsen (editors), CRC Press: Boca Raton, FL, pp. 201–41.

[192] J. Hawkins, S. Blakeslee (2004) On intelligence: how a new understanding of the brain will lead to the creation of truly tntelligent machines. Times Books.

[193] R. Penrose, S. Hameroff, H.P. Stapp, D. Chopra (2011) Consciousness and the universe: quantum physics, evolution, brain & mind. Cosmology Science Publishers.

[194] V.N. Vapnik. A.Y. Chervonenkis (1974) Theory of pattern recognition. Nauka, Moskwa (po rosyjsku).

[195] V.N. Vapnik (1995) The nature of statistical learning theory. Springer-Verlag New York, Inc.

[196] S. Theodoridis, K. Koutroumbas (2009) Pattern recognition. Fourth edition, Academic Press.

[197] M. Krzyśko, W. Wołyński, T. Górecki, M. Skorzybut (2008) Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości. WNT, Warszawa.

[198] K.P. Bennett, C. Campbell (2000) Support Vector Machines: hype or hallelujah? SIGKDD Explorations 2(2):1–13.

[199] S. Ayache, G. Quenot, J. Gensel (2007) Classifier fusion for SVM-based multimedia semantic indexing. Advances in Information Retrieval, 4425:494–504.

[200] M. Zampoglou, T. Papadimitriou, K.I. Diamantaras (2007) Support vector machines content-based video retrieval based solely on motion information. Proc IEEE Workshop on Machine Learning for Signal Processing, pp. 176–180.

[201] T. Podsiadły-Marczykowska (2011) Metody wspomagania procesu interpretacji badań mammograficznych z wykorzystaniem modelu ontologicznego, rozprawa doktorska, Instytut Biocybernetyki i Inżynierii Biomedycznej PAN.

[202] R. Neches, R.E. Fikes, T. Finin, T.R. Gruber, T. Senator, W.R. Swartout (1991) Enabling technology for knowledge sharing, AI Magazine 12(3):36–56.

[203] D.A. Huffman (1952) A method for the construction of minimum redundancy codes, Proc IRE, 40:1098–1101.

[204] N. Faller (1973) An adaptive system for data compression, Proc IEEE Asilomar Conf Circuits, Systems, and Computers, pp. 593–7.

[205] R.G. Gallagher (1978) Variations on a theme by Huffman, IEEE Tran Inf Theory 24(6):668–74.

[206] D.E. Knuth (1985) Dynamic Huffman coding, J Algorithms 6:163–80.

[207] J.S. Vitter (1987) Design and analysis of dynamic Huffman codes, J ACM 34(4):825–45.

[208] X. Lin (1991) Dynamic Huffman coding for image compression, praca magisterska, University of Nebraska.

[209] M. Nelson (1991) The data compression book, M&T Books.

[210] CCITT (ITU Recommendation T.4), Standardization of Group 3 facsimile apparatus for document transmission, 1980, amended in 1984 and 1988.

[211] CCITT (ITU Recommendation T.11), Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus, 1984, amended in 1988.

[212] ISO/IEC 11172 International Standard (MPEG-1), Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbits/s, 1993.

[213] ISO/IEC 13818 International Standard (MPEG-2), Information technology – Generic coding of moving pictures and associated audio information, 1996.

[214] ISO/IEC 10918 International Standard (JPEG), Information technology – Digital compression and coding of continuous-tone still images, 1994.

[215] W.B. Pennebaker, J.L. Mitchell (1992) "JPEG still image data compression. New York: Van Norstrand Reinhold.

[216] Norma 15938-2 organizacji ISO/IEC, komitetu JTC1/SC29, grupy roboczej WG11: Information Technology – Multimedia Content Description Interface – Part 2: Descriptions Definition Language, December 2001.

[217] A. Filip, J. Urbaniec (2008) Wizualizacja w e-nauczaniu: przerost formy nad treścią? E-edukacja dla rozwoju społeczeństwa – materiały z IV ogólnopolskiej konferencji 'Rozwój e-edukacji w ekonomicznym szkolnictwie wyższym', strony 146-150.

[218] ISO/IEC 15938-3/FCD. Information technology — multimedia content description interface – part 3 visual. Technical report, ISO/IEC, 1997.

[219] P. Owczarek and T. Rosiński (2003) Model eksperymentalny opisu treści wizyjnych. Proc VIII Warsztaty Telekomunikacyjne, pp. 171–176.

[220] M. Ostrowski (2002) Reprezentacja obiektów multimedialnych w MPEG-7. Praca magisterska, Politechnika Warszawska.

[221] B. S. Manjunath, W. Y. Ma (1996) Texture features for browsing and retrieval of image data. IEEE Tran Pat Anal Mach Intel 18(8):837–842.

[222] B.S. Manjunath, P. Salembier, T. Sikora (2002) Introduction to MPEG–7. John Willey & Sons, Ltd.

[223] H. Murakami, S. Matsumoto, Y. Hatori, H. Yamamoto (1987) 15/30 Mbit/s universal digital TV codec using a median adaptive predictive coding method. IEEE Tran Comm 35(6):637–645.

[224] Digital Imaging Group (1997) "FlashPix format specification". Version 1.0.1.

[225] ISO/IEC JTC1/SC29/WG1 (1996) "LOCO-A: An arithmetic coding extension of LOCO-I". Doc. N342.

[226] Baza obrazów mammograficznych zarchiwizowana w bezstratnym JPEG: http://marathon.csee.usf.edu/Mammography/Database.html

[227] Recomendation ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television pictures".

Słowniczek pojęć

Deskryptor – opis słowny (tekstowy, zazwyczaj w strukturze XML) lub liczbowy (binarny) charakteryzujący czy wręcz identyfikujący obiekt lub obiekty informacyjne (multimedialne) ze względu na określony atrybut (tj. wybraną właściwość obiektów); często pojęcie to obejmuje także sposób konstruowania opisu tekstowego lub algorytm wyznaczania liczbowej reprezentacji cech obiektu.

Deskryptorem jest metoda wyznaczania, algorytm, zasady składni opisu, w końcu zestaw liczb czy opis tekstowy, które charakteryzują, czy wręcz identyfikują obiekty ze względu na określony atrybut (tj. wybraną właściwość obiektów).

H.264 – standard multimedialny dotyczący zasadniczo metod kompresji audio i wideo, czyli sekwencji wizyjnych skojarzonych z dźwiękiem, niekiedy opisem tekstowym i innego typu znacznikami informacji, a także możliwymi formami interakcji.

Informacja – to wszystko, co służy bardziej skutecznej realizacji zamierzonego celu, czyli co jest użyteczne dla odbiorcy, odpowiada jego zapotrzebowaniom czy oczekiwaniom.

Indeks – zbiór informacji dotyczących określonej kolekcji danych multimedialnych. Dokładniej, jest to zbiór cech, czy wartości danego atrybutu ((tj. wybranej właściwości obiektu) wraz z listą identyfikatorów obiektów opisanych daną cechą.

JPEG – rodzina standardów dotyczących zasadniczo metod kompresji obrazów statycznych lub też sekwencji obrazów statycznych na sposób odwracalny lub nieodwracalny. Obejmuje także różnorodne uwarunkowania związane z przekazem, archiwizacją i obróbką obrazów, związane z szerokim wachlarzem aplikacji multimedialnych.

Kod – reguła (zasada, funkcja, przekształcenie) przyporządkowująca ciągowi symboli wejściowych (opisanych modelem źródła informacji) wyjściową reprezen-

tację kodową, która jest sekwencją bitową o skończonej długości, utworzoną z bitowych słów kodowych charakterystycznych dla danej metody.

Kompresja – proces przekształcania pierwotnej reprezentacji ciągu danych w reprezentację o mniejszej liczbie bitów; odwrotny proces rekonstrukcji oryginalnej reprezentacji danych na podstawie reprezentacji skompresowanej nazywany jest dekompresją. W szerszym znaczeniu jest to proces wyznaczania efektywnej reprezentacji danych, według kryteriów obowiązujących w danym zastosowaniu.

Media cyfrowe – określona forma użytkowania treści multimedialnych, taka jak Internet, telewizja cyfrowa, telefonia komórkowa, poczta elektroniczna, dystrybucje DVD, itp. Podstawową cechą mediów cyfrowych jest ich zdalna dostępność przez sieci telekomunikacyjne, z której wynika konieczność unormowania sposobów reprezentacji danych i ich opisów (metadanych) jako warunek skutecznej wymiany informacji.

MPEG – rodzina standardów multimedialnych, normujących sposoby kompresji multimediów, wymiany informacji, opisu obiektów multimedialnych, a także integracji mediów cyfrowych pod kątem różnorodnych aplikacji.

Multimedia – różne środki przekazu informacji, przy czym ta różnorodność dotyczy w pierwszej kolejności informacji (rodzaj, semantyka, treść), w drugiej - form przekazu (reprezentacja, jakość), a dopiero na końcu chodzi o zróżnicowanie środków (technologia, skala, zasady).

Obraz – to mieszanina trzech podstawowych składników: konturów wydzielających obiekty w obrazie, tekstur charakteryzujących wewnętrzne właściwości tych obiektów oraz tła. Treść przekazywana w obrazie zasadniczo zależy od rodzaju obiektów występujących w obrazie, ich właściwości, takich jak rozmiar, kształt, tekstura, spójność, jednorodność, itd., a także od wzajemnych relacji pomiędzy obiektami. Obrazy są kluczowym nośnikiem wykorzystywany we współczesnych systemach informacyjnych – według różnych szacunków poprzez formę obrazu dociera do nas blisko 80-90% wszystkich informacji.

Paradygmat – powszechnie przyjęte, sprawdzone wzorce teorii, pojęć, metod czy algorytmów, ogólniej rozwiązań znajdujących istotne odniesienia praktyczne. Jest zwykle przyjmowany na zasadzie konsensusu większości badaczy, może ulec modyfikacji lub być zastąpionym przez nowy paradygmat - lepszy wzorzec w danej dziedzinie.

Strumieniowanie – przesyłanie danych w formie strumienia, któremu towarzyszy wykorzystywanie kolejno napływających danych bezpośrednio po ich otrzymaniu.

Wyszukiwarka – narzędzie pozwalające na odnajdywanie określonej treści multimedialnej. Wyszukiwarka wykorzystuje przede wszystkim interfejs użytkownika pozwalający formułować zapytania w określonej formie (tj. zadania dotyczące przedmiotu przeszukiwań), bazy danych obiektów multimedialnych, opisanych indeksami w postaci odpowiednio przygotowanych struktur danych, a także silnik wyszukujący, który na podstawie indeksów pozwala odnaleźć obiekty najbardziej podobne do zapytania.

Załączniki

Dodatek A

Zestaw zadań

Multimedia jako zintegrowany przekaz międzyludzki

Pytania

1. Czym charakteryzuje się multimedialny przekaz informacji, jakie są jego zalety, a jakie ograniczenia? Odpowiedź należy uzasadnić na przykładach.
2. Proszę wyjaśnić ideę usług multimedialnych oraz potencjału stosowanych technologii na przykładzie internetowego serwisu ogólnopolskiej gazety codziennej.
3. Proszę zdefiniować pojęcie "multimedia" ze względu na:
   a) rolę i istotę przekazu danych,
   b) źródła przekazu danych,
   c) zastosowania (przeznaczanie/wykorzystanie/aplikacje),
   d) stosowane narzędzia/sprzęt,
   e) wykorzystywane standardy.
4. Proszę scharakteryzować pojęcie dźwięku – czym jest fizycznie, jak można opisać dźwięk w zakresie:
   a) widma częstotliwościowego,
   b) sposobu propagacji,
   c) subiektywnych doświadczeń słuchacza,
   d) metod analizy, możliwych zniekształceń, oceny jakości?
5. Proszę narysować i krótko omówić schemat typowego procesu produkcyjnego materiału dźwiękowego.

6. Proszę opisać system odtwarzania dźwięku 5.1. Jakie są trendy rozwoju metod przestrzennego odtwarzania dźwięku?

7. Proszę scharakteryzować znane rodzaje mikrofonów. Jakie cechy zestawów mikrofonowych są istotne przy doborze optymalnych warunków pracy w studiu nagrań?

8. Jakie są podstawowe, kolejne procedury przetwarzania wykorzystywane w metodach kompresji dźwięku?

9. Proszę podać ograniczenia ludzkiej percepcji wykorzystywane w metodach przetwarzania dźwięku i obrazu.

10. Proszę opisać przykładowe aplikacje multimedialne wykorzystujące metody grafiki komputerowej. Jak uzyskiwany jest realizm scen graficznych?

11. Proszę opisać rolę metod cieniowania w budowaniu scen metodami grafiki komputerowej.

12. Proszę wymienić i krótko scharakteryzować procesory dźwięku, które są często uzupełnieniem stołów mikserskich w studiach nagrań.

13. Proszę wyjaśnić pojęcia:

    a) funkcja jasności, pole obrazu,

    b) kontrast, histogram, zdolność rozdzielcza,

    c) tekstura, widmo spektralne,

    d) scena multimedialna, obiekt multimedialny,

14. Na czym polegają efekty maskowania równoczesnego oraz nierównoczesnego i jaki jest ich wpływ na percepcję dźwięku?

15. Proszę omówić najnowsze technologie wykorzystywane w przetwornikach obrazów.

16. Proszę opisać zjawiska fizyczne wykorzystywane do wyświetlania obrazów i podstawowe problemy konstrukcyjne współczesnych wyświetlaczy.

17. Należy podać przykłady nowoczesnych technologii multimedialnych, które realizują koncepcję wielostrumieniowego przekazu multimedialnego.

18. Proszę o krótką odpowiedź na następujące pytania:

    a) czym różni się barwa dźwięku od wysokości dźwięku?

    b) czym różni się oświetlenie od cieniowania obiektów w grafice komputerowej?

Zagadnienia problemowe

1. Proszę dokonać klasyfikacji znanych urządzeń multimedialnych oraz mediów cyfrowych ze względu na stopień integracji możliwie wielu strumieni przekazu informacji. Jakie są kierunki rozwoju rynku multimediów?

2. Jak wykorzystywany jest tekst oraz możliwości interakcji w przekazie multimedialnym?

3. Proszę scharakteryzować metody oceny jakości obrazów poprzez zwięzłe wyjaśnienie następujących zagadnień:

   a) jak należy rozumieć pojęcie jakości obrazów w kontekście konkretnej aplikacji, czym różni się ocena jakości od oceny użyteczności?

   b) jakie są podstawowe metody oceny jakości obrazów?

   c) co jest bardziej istotne w odbiorze informacji obrazowej: dobry kontrast obrazu czy też wysoka zdolność rozdzielcza?

Zadania obliczeniowe

1. Porównano efektywność dwóch metod A i B poprawy jakości obrazów. Uzyskano następujące wyniki reprezentatywne dla 50 obrazów przetworzonych tymi metodami:

   a) błąd średniokwadratowy (względem obrazów źródłowych) na poziomie około $MSE_A = 150$ oraz $MSE_B = 32$ (średnie wartości błędu MSE na zbiorze obrazów testowych),

   b) średnia ocen subiektywnych w skali -3 ... 3 z opisem semantyki obrazów (od znacznego pogorszenia do znaczącej poprawy jakości obrazów): $s_A = 1,52$ oraz $s_B = 1,35$.

   Należy określić bardziej efektywną metodę przetwarzania obrazów. Czy na podstawie tych wyników można stwierdzić jednoznacznie jej użyteczność w danej aplikacji wykorzystującej informację obrazową?

2. Proszę zaprojektować subiektywny test oceny jakości zdjęć pejzaży, których wynikiem będzie:

   a) bezwzględna ocena obrazów źródłowych, pozwalająca ustalić kolejność najlepszych zdjęć według zespołu ekspertów,

   b) wskazanie najlepszej metody przetwarzania zdjęć, na podstawie oceny jakości najlepszych zdjęć źródłowych względem przetworzonych trzema technikami poprawy percepcji.

Reprezentowanie informacji

Pytania

1. Jaka jest definicja pojęcia informacji w kontekście przekazu multimedialnego? Jakie formy obiektywizacji rozumienia informacji są stosowane w mediach cyfrowych?

2. Jaka jest definicja kompresji danych? Proszę wyjaśnić następujące pojęcia:

   a) reprezentacja danych,

   b) odwracalność i nieodwracalność procesu kompresji,

   c) efektywność kompresji,

   d) modelowania źródeł informacji,

   e) binarne kodowanie.

   miary efektywności kompresji,

3. Proszę wymienić znane standardy i metody kompresji dźwięku oraz mowy. Dlaczego możliwa jest kompresja dźwięku? Jakiego typu nadmiarowość może być usunięta oraz jakie zjawiska charakteryzujące ludzką zdolność percepcji dźwięku są tutaj wykorzystywane?

4. Proszę podać znane standardy i metody kompresji obrazów. Dlaczego możliwa jest kompresja obrazów? Jakiego typu nadmiarowość może być usunięta oraz jakie zjawiska charakteryzujące ludzką zdolność percepcji obrazów są tutaj wykorzystywane?

5. Na czym polega indeksowanie treści multimedialnej?

6. Co to są deskryptory obrazów, do czego są wykorzystywane? Proszę omówić wybrany przykład deskryptora.

7. Czym różnią się metody przetwarzania, analizy i rozpoznawania obrazów? Proszę podać przykłady algorytmów oraz zastosowań tych metod.

8. Czemu służy proces indeksowania treści multimedialnej, jakie rozwiązania/mechanizmy wykorzystywane są w systemach indeksowania danych?

9. Proszę wymienić i krótko omówić podstawowe metody obróbki obrazów: czemu służą, jakie efekty można uzyskać, jaka jest ich rola w przekazie multimedialnym?

10. Proszę omówić podstawowe metody segmentacji obrazów. Do czego są one wykorzystane w aplikacjach multimedialnych?

11. Proszę scharakteryzować metody przetwarzania obrazów w celu poprawy ich jakości w ocenie obserwatorów, jak też w celu zwiększenia efektywności metod analizy obrazów.

12. Proszę odpowiedzieć na pytania:

    a) dlaczego niektóre metody kompresji nazywane są stratnymi?

    b) dlaczego w konstrukcji kodów binarnych wykorzystywana jest struktura drzewa?

    c) czym różnią się deskryptory wizualne i audio od deskryptorów tekstowych?

    d) czym się różni model źródła informacji z pamięcią i bez?

13. Proszę zdefiniować pojęcie entropii przy różnych modelach źródeł informacji i odpowiedzieć na pytania:

    a) jaka jest relacja pomiędzy modelem źródła, oszacowaną ilością informacji a możliwą efektywnością kodu?

    b) w jaki sposób można uzyskać kod optymalny?

14. Proszę wyjaśnić pojęcie nieliniowej aproksymacji sygnałów. Dlaczego metody analizy i aproksymacji sygnałów są wykorzystywane do odszumienia sygnałów użytecznych w przekazie informacji?

Zagadnienia problemowe

1. Jakie są ograniczenia dotyczące wykorzystania modeli Markowa wyższych rzędów w kodowaniu danych?

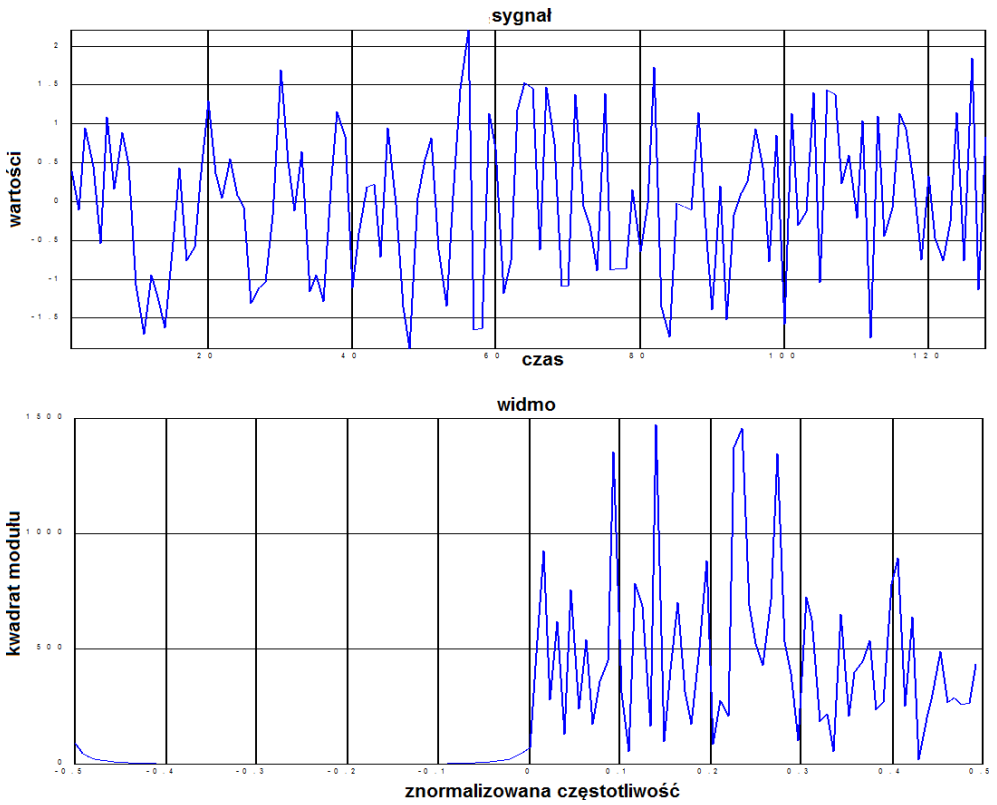
2. Jaką rolę pełni model w kodowaniu, indeksowaniu oraz obróbce danych multimedialnych? Proszę podać oczekiwane cechy modelu oraz przykładowe rozwiązania.

3. Proszę podać schemat blokowy reprezentatywnej metody przetwarzania obrazów w celu poprawy percepcji informacji wizyjnej.

4. Proszę omówić przykładową metodę rozpoznawania wzorców, bazującą na wstępnej segmentacji obszarów zainteresowania, doborze cech oraz klasyfikacji obiektów zgodnie z wybraną aplikacją.

5. Zaproponuj deskryptor charakteryzujący kolor obrazu, jednocześnie określając sposób jego wykorzystania w praktyce.

6. Poniżej przedstawiono rysunki (rys. A.1) zaszumionego sygnału o liniowej modulacji częstotliwościowej oraz jego widmo częstotliwościowe (kwadrat modułu współczynników Fouriera). Należy rozważyć możliwości odszumienia sygnału za pomocą dostępnych metod przetwarzania, biorąc pod uwagę zarówno działania w dziedzinie sygnału, jak też w dziedzinie częstotliwościowej. Jakie problemy można napotykać przy realizacji tego zadania?



Rysunek A.1: Zaszumiony sygnał (góra) wraz z widmem częstotliwościowym (dół).

Zadania obliczeniowe

1. Proszę zbadać, czy podane kody symboli są jednoznacznie dekodowalne: $A_{\mathcal{K}_1} = \{10, 000, 11, 010\}$, $A_{\mathcal{K}_2} = \{1, 01, 010\}$, $A_{\mathcal{K}_3} = \{1, 10, 00\}$. Dodatkowo należy zweryfikować poprawność kodów za pomocą nierówności Krafta (MacMillana).

2. Które z podanych niżej alfabetów słów kodowych określają kod jednoznacznie dekodowalny (proszę opisać weryfikację każdego z kodów)?
$A_{\mathcal{K}_1} = \{0, 01, 001, 0001\}$, $A_{\mathcal{K}_2} = \{1, 101, 011\}$, $A_{\mathcal{K}_3} = \{0, 101, 11, 100\}$, $A_{\mathcal{K}_4} = \{0, 01, 11\}$.

3. Proszę podać przykłady kodów przedrostkowych o silnie zróżnicowanych długościach słów, a także kody nie będące przedrostkowymi, które są jednoznacznie dekodowalne.

4. Nadawca zna bardzo istotne słowo "l,a,t,a,w,i,e,c", które pomoże odbiorcy odszyfrować tajne dokumenty. Przekazywanie informacji jest jednak bardzo utrudnione i wymaga minimalnej reprezentacji bitowej. Proszę obliczyć, ile informacji musi przekazać nadawca, a następnie zaproponować metodę kodowania.

5. Należy zdefiniować miarę ilości informacji, a następnie obliczyć, które źródło informacji – o rozkładzie $P_{S1} = p_0 = 1/8, p_1 = 7/8$ czy też o rozkładzie $P_{S2} = p_0 = 1/2, p_1 = 1/2$ – dostarcza więcej informacji.

6. W systemie wyszukiwania obrazów po zawartości uzyskano następujące wyniki:

Zapytanie	Liczba zwróconych obrazów	Pozycje poprawnych obrazów w zbiorze zwróconych	Liczba obrazów podobnych do zapytania w bazie
1	20	1,3,4,5,13,15,17,20	10
2	20	1,2,3,4,5,6,7,8,9,10,12,14,15,20	40
3	20	2,3,4,5,9,10,13,16,17,18	20
4	20	1,2,5,6,7,8,12,16,17,19	36
5	20	1,5,6,13,14,17	13

Proszę ocenić selektywność wyszukiwarki obliczając precyzję, przywołanie i stopę sukcesu.

Komputerowe przetwarzanie informacji – metody

Pytania

1. Proszę opisać sposób realizacji (najlepiej algorytm) podstawowych operacji na obrazie, takich jak:

   a) korekcja kontrastu i jasności, korekcja gamma,

   b) wyznaczenie histogramu oraz możliwe korekcje kontrastu poprzez przekształcenia histogramowe,

   c) dokonanie filtracji splotowej liniowej z wykorzystaniem filtru zdefiniowanego maską, z uwzględnieniem sposobu przetwarzania pikseli granicznych,

   d) dokonanie kontekstowej operacji nieliniowej.

2. Proszę opisać dziedzinę falkową przekształconego obrazu oraz użyteczne sposoby kodowania współczynników falkowych.

3. Proszę o odpowiedź na następujące pytania związane z indeksowaniem obrazów:

   a) co zawiera typowy indeks, co go definiuje?

   b) co to jest funkcja podobieństwa, podaj jakiś jej przykład?

   c) jak mierzy się efektywność wyszukiwania po zawartości przy zapytaniach przez przykład?

4. Proszę o odpowiedź na następujące pytania związane z obróbką obrazów:

   a) co to jest obraz, na czym polega istota informacji obrazowej?

   b) jakie znasz sposoby modelowania obrazów?

   c) jakie metody kompresji bezstratnej uwzględniają specyfikę obrazów?

5. Proszę wyjaśnić, na czym polega rozumienie obrazów i jaką rolę w tym procesie odgrywają metody komputerowe.

Zagadnienia problemowe

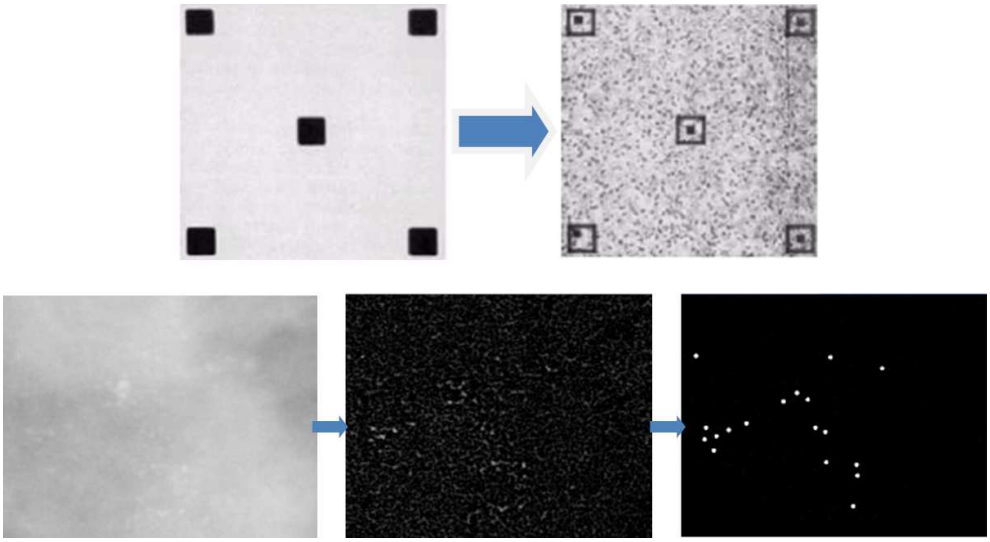
1. Proszę zaproponować i krótko omówić metody obróbki obrazów, które służą poprawie jakości obrazu jak na rys. A.2. Jak wyglądałby zapis operacji obrotu twarzy z obrazu o kąt 90°?

   Sugerowane metody przetwarzania należy zweryfikować za pomocą dogodnego narzędzia, np. programu IMLAB.

Rysunek A.2: Obraz twarzy poddawany obróbce w celu poprawy jakości obrazu.

2. Proszę opisać możliwe metody kwantyzacji bloku danych 8×8, charakteryzując ich przypuszczalną efektywność w zastosowaniach kompresji danych.

3. Proszę scharakteryzować metody przetwarzania obrazów, które pozwolą uzyskać efekty jak na rys. A.3.



Rysunek A.3: Efekty przetwarzania obrazów źródłowych.

4. Proszę wyjaśnić rolę filozofii: "jak rozumieć i pokazać to, co da się policzyć?" w doskonaleniu metod komputerowego rozumienia obrazów. Należy uwzględnić wybrany przykład zastosowań.

Zadania obliczeniowe

1. Proszę podać przykłady filtrów: wygładzającego, wykrywającego krawędzie, podkreślającego krawędzie, a także przykład operacji punktowej (bezkontekstowej) na obrazie.

2. Proszę dokonać filtracji metodą splotu bloku obrazu jak na rysunku poniżej, wykorzystując filtr g zdefiniowany maską jak na rys. A.4.

33	25	23	18	19
32	30	34	26	25
024	27	21	19	17
17	24	16	19	19

blok obrazu
(proszę przetwarzać jedynie obszar zaznaczony linią przerywaną)

$g=$

0	**1**	**0**
0	**0**	**0**
0	**-1**	**0**

Rysunek A.4: Filtracja bloku obrazu (po lewej) za pomocą filtru g (po prawej).

3. Proszę wykryć krawędź we fragmencie bloku obrazu jak na rys. A.5. Jaki filtr należy zastosować, by uzyskać efekt wyraźnej i możliwie cienkiej krawędzi? Jak będzie wyglądać fragment obrazu po filtracji? Jaką regułę ustalenia punktów krawędziowych użyć, jaki ewentualnie mechanizm rozszerzenia bloku obrazu należy wykorzystać?

133	125	123	122	118
130	123	134	125	81
122	132	61	69	70
141	58	63	59	61

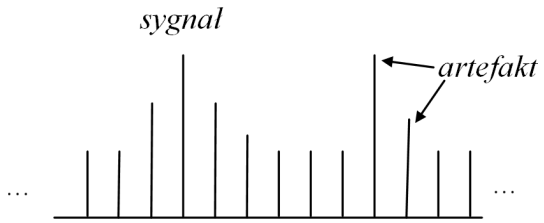
Rysunek A.5: Przykładowy blok obrazu poddawany filtracji w celu wykrycia krawędzi.

4. Dla podanego fragmentu obrazu (rys. A.6), zawierającego wartości funkcji jasności pikseli w bloku, należy zaproponować metody poprawy kontrastu, wzrostu jasności oraz filtracji dolnoprzepustowej w celu usunięcia szumów. Następnie trzeba przetworzyć obraz zgodnie z zaproponowaną procedurą.

31	34	32	35	32
35	33	34	36	33
29	27	29	32	30
32	37	33	28	31
28	26	31	33	29

Rysunek A.6: Przykładowy blok wartości funkcji jasności pikseli, poddawany prostym operacjom poprawy jakości obrazu.

5. Proszę zaproponować filtr, który możliwie skutecznie wzmocni sygnał użyteczny i stłumi/usunie artefakt w ciągu próbek jak na rys. A.7.



Rysunek A.7: Ciąg próbek sygnału ze wskazanymi artefaktami, poddany przetwarzaniu, które wzmocni sygnał względem artefaktów.

6. Należy dokonać filtracji dolnoprzepustowej wektora danych całkowitoliczbowych: $[0, -2, -3, 1, -2, 0, 1, 5]$ metodą splotu. W tym celu trzeba zaproponować postać filtru oraz ustalić sensowne warunki brzegowe. Czym się różni filtr modalny od medianowego?

Użytkowanie informacji, czy narzędzia

Pytania

1. Proszę odpowiedzieć na pytania:
   a) dlaczego efektywność kodu Golomba wzrasta po uporządkowaniu alfabetu źródła?
   b) jak wpływa rząd kodu Golomba na efektywność kodowania?
   c) jakie korzyści związane są z wykorzystaniem wykładniczej wersji kodu Golomba (przedziałowego)?

Zagadnienia problemowe

1. Wyjaśnij, kiedy kod Golomba dorównuje, a kiedy nawet przewyższa efektywnością kod Huffmana. Zaproponuj metodę adaptacji rzędu kodu Golomba do lokalnych zmian rozkładu prawdopodobieństw symboli alfabetu.

2. Zaproponuj mechanizm przełączania modeli wektorów rekonstruowanych w algorytmie kompresji obrazów z wektorową kwantyzacją bloków. Wykorzystaj w tym celu charakterystykę wektorów danych źródłowych, która różnicuje treść typowych rodzinnych fotografii.

Zadania obliczeniowe

1. Mając dany ciąg źródłowy symboli $\mathbf{s}_{we}$ = (a,c,c,h,b,i,a,h) należy określić ilość informacji dostarczonej przez to źródło oraz nadmiarowość reprezentacji tej sekwencji zakodowanej metodą Huffmana. Podaj słowa kodu Huffmana o możliwie zrównoważonej długości.

2. Należy zakodować możliwie efektywnie następujący ciąg danych źródłowych: $\mathbf{s}_{we}$ = (5,1,1,5,5,3,2,1,7,3,3,3).

3. Dany jest następujący ciąg danych źródła informacji o alfabecie binarnym: $\mathbf{s}_{we} = (1, 0, 0, 1, 1, 1, 0, 1, 0)$. Proszę zakodować ten ciąg metodą RLE oraz Golomba.

4. Proszę wyznaczyć kod drzewa binarnego o 6 liściach, z symbolami o wagach odpowiednio 5,8,3,7,1,2, przy czym należy podać możliwie wiele rozwiązań, wykorzystując m.in kod Huffmana i kod Golomba.

Pragmatyzm multimediów

Pytania

1. Proszę opisać algorytm kompresji według standardu JPEG. Co to jest transformacja DCT, w jakim celu wykonuje się kwantyzację współczynników tej transformaty, jak są one kodowane?

2. Proszę wymienić właściwości danych, które zostały opisane deskryptorami w normie MPEG-7. Jak konstruowane są deskryptory złożone?

3. Proszę podać przykładowe deskryptory dźwięku, obrazu pojedynczego i filmu.

4. Czemu służą wektory ruchy wyznaczane w kompresji wideo definiowanej w standardach rodziny MPEG? Proszę opisać wybraną metodę estymacji oraz kompensacji ruchu, stosowaną w MPEG.

5. Dlaczego w standardzie JPEG:

   a) stosuje się podział obrazu na bloki?

   b) wykorzystuje się metodę RLE?

   c) stosuje się kwantyzację, jakie znasz metody kwantyzacji?

   d) można kodować obrazy progresywnie?

6. Czy składowe luminancji i chrominancji obrazów przetwarzane są tak samo w kompresji według MPEG? Na czym polegają ewentualne różnice? Proszę przy okazji wyjaśnić różne struktury próbkowania (podpróbkowania).

Zagadnienia problemowe

1. Wyjaśnij ideę technik multimedialnych na przykładzie internetowego serwisu ogólnopolskiej gazety codziennej. Jakie standardy multimedialne wspierają realizację portalu gazety codziennej? Jaką rolę w usprawnieniu takiego medium odgrywają kodeki danych multimedialnych?

2. Proszę scharakteryzować paradygmat, ewentualnie paradygmaty kompresji wideo przyjęte w ramach standardów MPEG.

3. Proszę porównać schematy blokowe standardów kompresji JPEG i JPEG2000. Co daje zastąpienie DCT transformacją falkową?

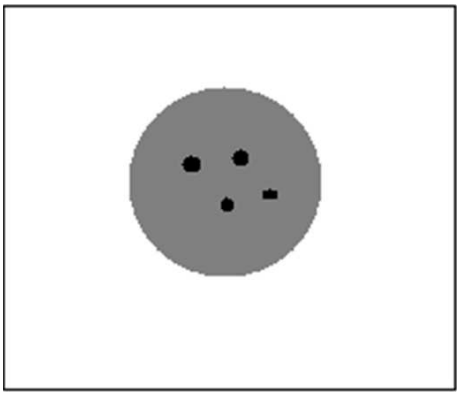
Zadania obliczeniowe

1. Proszę dokonać kwantyzacji bloku o rozmiarach 8×8 jak na rys. A.8, zawierającego współczynniki DCT, zgodnie z regułami standardu JPEG. Proszę ustalić postać tablicy kwantyzacji, a następnie wyznaczyć postać macierzy wartości skwantowanych współczynników. Dodatkowo proszę podać rozkład wartości progów selekcji dla poszczególnych współczynników oraz narysować uzyskaną strefę przenoszenia energii sygnału.

784.2	6.630	-6.45	-5.46	2.25	0.000	-3.82	-2.80
-34.9	10.89	-0.46	5.318	4.209	3.641	3.270	-2.62
0.726	10.11	12.68	-0.72	2.956	-5.22	-0.91	3.229
-19.8	2.843	21.02	7.299	-5.43	-7.31	7.198	2.446
9.000	-2.50	20.10	1.802	-8.00	5.061	-0.48	1.331
-21.7	-3.61	8.453	-1.76	-5.13	-3.66	0.128	0.899
2.405	37.31	-0.16	3.307	2.564	0.964	2.072	-3.61
40.72	-3.57	-18.2	-1.70	3.892	0.977	2.199	-0.53

Rysunek A.8: Przykładowy blok współczynników transformaty DCT, poddawany kwantyzacji zgodnie z koncepcją normy JPEG.

2. Proszę naszkicować, jak będzie wyglądać progresywna oraz sekwencyjna rekonstrukcja obrazu z rys. **??**. Jakie rodzaje progresywnego porządkowania kodowanych danych są możliwe?



Rysunek A.9: Prosty obraz kodowany według porządku sekwencyjnego oraz z progresją.

Dodatek B

Ćwiczenia

Zaproponowano sześć ćwiczeń pogłębiających wybrane zagadnienia "Technik multimedialnych", przede wszystkim ze względu na praktyczne aspekty dostępnych metod, urządzeń i oprogramowania. Kolejno przedstawiono:

1. Rejestracja obrazu i dźwięku
2. Postrzeganie przekazu informacji
3. Kompresja multimediów
4. Indeksowanie danych multimedialnych
5. Komputerowe przetwarzanie danych multimedialnych
6. Użytkowanie multimediów

B.1 Rejestracja obrazu i dźwięku

Celem ćwiczenia jest badanie wybranych metod rejestracji danych multimedialnych, w tym: weryfikację zasad próbkowania i kwantyzacji poprzez przyswojenie podstaw teoretycznych (np. zasad próbkowania), eksperymentalną weryfikację efektów rejestracji z wykorzystaniem gotowych narzędzi oraz analizę wyników i sformułowanie wniosków.

B.1.1 Program ćwiczenia

Przewidywana jest realizacja następujących zadań:

- ocena efektów próbkowania i kwantyzacji rejestrowanych sygnałów,
- testowanie metod rejestracji dźwięku,

- testowanie metod rejestracji obrazów.

Ocena efektów próbkowania i kwantyzacji rejestrowanych sygnałów obejmuje

- wstępną charakterystykę procesu akwizycji sygnałów, określeniu roli i zasad procesów próbkowania i kwantyzacji oraz konstruowania formatu zapisu danych cyfrowych;
- badanie efektów próbkowania przy regulacji częstości próbkowania w schemacie równomiernym oraz przykładowym nierównomiernym – obserwacja zjawiska aliasingu przy malejącej częstości próbkowania, ocena wpływu filtru ograniczającego pasmo sygnału na redukcję aliasingu, ocena zależności jakości sygnału (przede wszystkim jego widma) od częstości próbkowania (rola nadpróbkowania);
- badania wpływu schematu kwantyzacji na jakość sygnałów cyfrowych, przede wszystkim wielkości przedziału kwantyzacji w schemacie równomiernym oraz różnicy pomiędzy schematem równomiernym i nierównomiernym (Lloyda-Maxa) przy tej samej liczbie przedziałów kwantyzacji; przypisanie rejestrowanym próbkom binarnej reprezentacji kodów stałej długości (dwójkowego, Greya, U2 itp.).

Rodzaj wykorzystywanych sygnałów jest dowolny – w przypadku rejestracji dźwięku możliwa jest subiektywna ocena efektów akwizycji.

Testowanie metod rejestracji dźwięku polega przede wszystkim na:

- rejestracji zapisów dźwiękowych za pomocą różnych zestawów mikrofonowych, uwarunkowań zapisu (orientacja przestrzenna, studyjnie, na "ulicy", itp.), z użyciem kart dźwiękowych z użytecznym oprogramowaniem;
- dokładnej ocenie właściwości (widmo, zależności czasowe, dynamika, brzmienie) rejestrowanych form muzycznych, mowy, zapisu głosu różnych osób;
- ocenie jakości zapisów przy różnych zmiennych częstościach próbkowania, konfiguracjach ograniczeniach dynamiki, zapisie do różnych formatów (MP3, aac, itp.), testowa ocena efektów modyfikowania głosów poszczególnych osób, synteza (miksowanie) wspólnych wypowiedzi, wykorzystanie możliwych procedur masteringu, itp.

Testowanie metod rejestracji obrazów sprowadza się do badania uwarunkowań zapisu obrazów za pomocą kamer i aparatów fotograficznych podłączanych do komputera z wykorzystaniem różnego typu interfejsów. Do rejestracji sygnału z kamer analogowych można wykorzystać karty *frame grabber* z regulowaną charakterystyką. W przypadku kamer cyfrowych, dedykowane kart rejestracji obrazów

(nawet np. karty telewizyjne, graficzne) warto wykorzystać oryginalne sterowniki pozwalające zmieniać warunki akwizycji. W ramach ćwiczenia należy:

- dokonać podglądu zapisywanego strumienia wideo oraz jego archiwizacji, z doborem parametrów rejestracji i wyświetlania w możliwie szerokim zakresie;
- zbadać wpływ takich parametrów rejestracji jak zdolność rozdzielcza, dynamika, rodzaj kompresji oraz konwersje przestrzeni barw na jakość rejestrowanego wideo;
- dokonać próby poprawy jakości rejestrowanych danych za pomocą zmiany uwarunkowań procesu akwizycji (dobór oświetlenia, rozstaw kamer, dynamika sceny, wykorzystanie wbudowanych metod przetwarzania wstępnego), formatu zapisu (np. parametrów MPEG); należy także obserwować wpływ zmiennych warunków akwizycji na rozmiar rejestrowanego strumienia.

B.1.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:

- Matlab – Signal Processing Toolbox,
- Matlab – Image Acquisition Toolbox,
- Matlab – Data Acquisition Toolbox,
- Matlab – funkcje ze strony http://www.falstad.com/mathphysics.html
- narzędzia Wavosaur lub inne (np. GoldWave), aplety ze strony http://www.falstad.com/mathphysics.html
- edytor dźwięku Audacity http://audacity.sourceforge.net/
- inne.

B.1.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów uzyskanych efektów rejestracji multimediów oraz sformułowanych wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- wpływ próbkowania i kwantyzacji na efekty akwizycji sygnałów, przede wszystkim w kontekście zjawiska aliasingu oraz optymalnego wykorzystania dynamiki rejestrowanych sygnałów,
- uwarunkowania procesu rejestracji obrazów i dźwięku,

- zasady służące poprawie jakości rejestrowanego przekazu multimedialnego.

B.2 Postrzeganie przekazu informacji

Celem ćwiczenia jest badanie procesu postrzegania przekazu multimedialnego, zarówno od strony zapewnienia odpowiednich warunków wizualizacji czy odsłuchu informacji, jak też możliwości percepcji wszystkich istotnych elementów przekazu, z zachowaniem satysfakcjonującego odbiorcę poziomu jakości.

B.2.1 Program ćwiczenia

Przewidywana jest realizacja następujących zadań:

- analiza specyfiki treści dostarczanej z różnych źródeł informacji oraz formy przekazu;
- badanie procesu percepcji przekazywanej informacji o charakterze multimedialnym;
- ocena użyteczności danych multimedialnych.

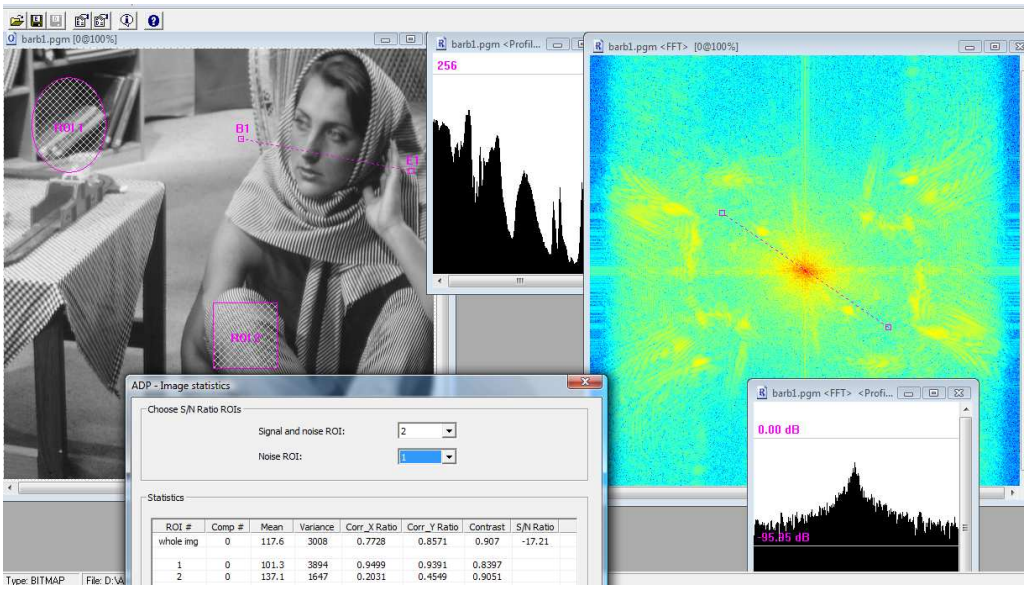
Analiza specyfiki treści dostarczanej z różnych źródeł informacji oraz formy przekazu polega na przeprowadzeniu wstępnej charakterystyki różnego typu danych multimedialnych: obrazów statycznych, sekwencji wizyjnych, filmów oraz zapisów mowy i muzyki. Chodzi tutaj przede wszystkim o odwołanie do ludzkich modeli postrzegania i odczytywania treści (ograniczenia postrzegania obrazów i dźwięku, zróżnicowanie postrzegania, komponenty treści, rozpoznawanie obiektów, wzajemnych relacji, itd.). Należy na wybranych, własnych przykładach omówić specyfikę postrzegania obrazów i dźwięku, podkreślając przede wszystkim jej znaczenie w kontekście różnych zastosowań.

Badanie procesu percepcji przekazywanej informacji o charakterze multimedialnym dotyczy dwóch zasadniczych aspektów:

- obliczeniowej oceny ogólnej jakości danych, przy czym należy wziąć pod uwagę przede wszystkim te cechy sygnału, które mają bezpośredni związek z percepcją przenoszonej informacji;
- oceny percepcji danych w teście subiektywnej oceny jakości według różnych scenariuszy.

Szacując jakość obrazów i dźwięku, stanowiących podstawę przekazu multimedialnego, należy wykorzystać różne miary obliczeniowe. Wstępna ocena jakości wybranej grupy zróżnicowanych obrazów testowych powinna dotyczyć obliczeń kontrastu lokalnego (w wybranych regionach zainteresowań), histogramu poszczególnych komponentów i poziomu szumów na podstawie widma częstotliwościowego oraz analizy statystycznej. Zdolność rozdzielczą systemów obrazowania można

oszacować pośrednio, analizując widma oraz profile obrazowe istotnych krawędzi czy dokonując charakterystyki dominujących tekstur. Przykładową analizę jakości obrazu przedstawiono na rys. B.1.

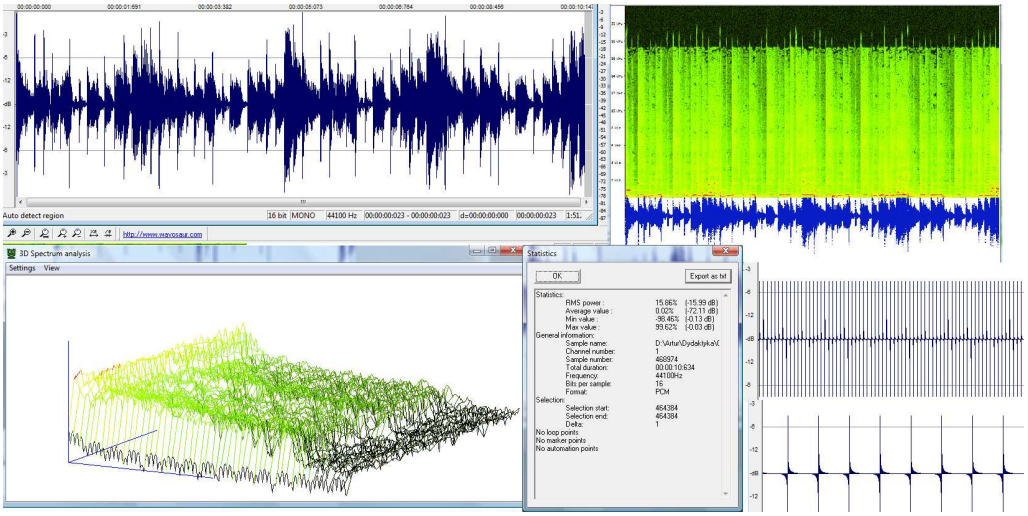


Rysunek B.1: Przykładowa ocena jakości obrazu z wyświetlonym profilem, dwoma regionami zainteresowania, 2W widmem częstotliwościowym oraz 1W profilem widma, podstawową statystyką z uwzględnieniem zaznaczonych regionów; wykorzystano narzędzie MDI.

W przypadku dźwięku należy przeanalizować przede wszystkim postać widma, sonogram, zróżnicowanie zapisu wielokanałowego, statystykę, itp. dla różnych sygnałów dźwiękowych oraz mowy. Przykładową analizę pokazano na rys. B.2. Warto zwrócić uwagę na różnice w obliczeniowej ocenie dźwięku, występujące pomiędzy różnymi typami muzyki (można skorzystać z nagrań własnych oraz referencyjnych), zapisami mowy o zróżnicowanej barwie i cechach charakterystycznych.

Odnosząc względem siebie sygnały rejestrowane przy różnych uwarunkowaniach, czy też wstępnie przetworzone za pomocą odmiennych procedur, należy wykorzystać miary porównawcze, takie jak MSE czy PSNR, jak też złożone miary obliczeniowe, uwzględniające ludzkie zdolności percepcji i/lub optymalizowane testem subiektywnym: SSIM, VQM, OMW.

Obok miar obliczeniowych, warto skorzystać z testów obserwacyjnych (subiektywnych), odpowiednio dobierając skale i rodzaj formułowanych ocen (względne, absolutne, porównawcze z porządkowaniem, itp.). Testy te mają służyć przede wszystkim badaniu zdolności postrzegania obiektów, wybranych struktur czy cech przez obserwatorów. Powinny to odzwierciedlić projektowane procedury realizacji



Rysunek B.2: Przykładowa ocena jakości zapisu dźwiękowego z wyświetlonym przebiegiem czasowym sygnału oraz czasowo-częstotliwościowym widmem sonogramu (góra), 3W widmem częstotliwościowym, prostą statystyką oraz przetworzonymi postaciami sygnału, gdzie uwypuklono wybrane jego cechy. Wykorzystano narzędzie Wavosaur.

testów – w ich planowaniu warto wykorzystać odpowiednie normy ITU-R [227].

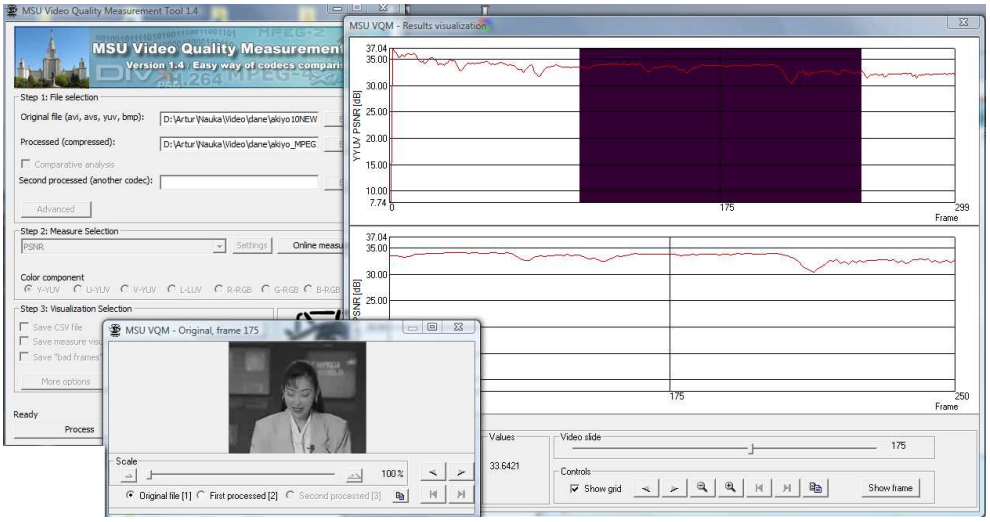
Pomocne w realizacji testów subiektywnych, ale także przy obliczaniu złożonych miar obliczeniowych, korelujących z oceną subiektywną, mogą być narzędzia MSU (http://www.compression.ru/video/) – pokazano to na rys. B.3. Wykorzystując aplikację MSU perceptual video quality, przedstawioną na rys. B.4, można zaprojektować procedurę testu, a następnie test zrealizować. Za pomocą prostych procedur statystycznych można też dokonać zestawienia i analizy wyników.

Ocena użyteczności danych multimedialnych dotyczy weryfikacji przekazu multimedialnego ze względu na jego użyteczność w obliczu konkretnego zastosowania. Użyteczność rozumiana jest tutaj jako przydatność odczytywanej informacji w realizacji określonego zadania, mierzona najczęściej sprawnością realizacji tego zadania. Przykładowo, w medycznej diagnostyce obrazowej ocena użyteczności dotyczy rozpoznania w obrazie symptomów określonej patologii.

Procedura oceny może polegać na detekcji określonej treści (obiektu, właściwości) w zbiorze danych testowych, a analizę wyników można przeprowadzić z wykorzystaniem krzywej ROC, korzystając z pomocy specjalistów danej dziedziny lub też ogólnie znawców przedmiotu.

B.2.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:



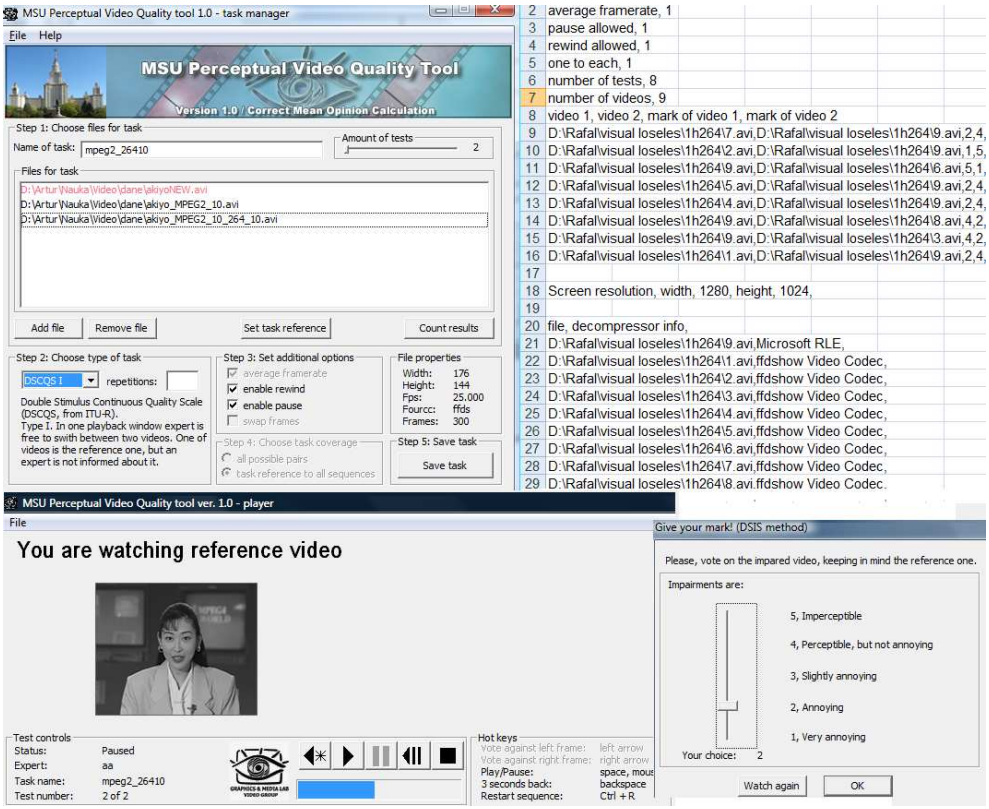
Rysunek B.3: Przykład oceny jakości sekwencji wizyjnej – po prawej wykres błędu PSNR dla ramek całej sekwencji kodowanej według MPEG-2, względem wideo źródłowego. Poniżej tego wykresu – szczegółowa analiza błędu w wybranym fragmencie, z możliwością podglądu jakości wybranej ramki.

- narzędzia MDI (www.ire.pw.edu.pl/ arturp/Dydaktyka/Media_tools/), Wavosaur, aplety ze strony http://www.falstad.com/mathphysics.html
- narzędzia IMLAB, CXIMAGE, VirtualDub, MSU Video Quality Measurement Tool, MSU perceptual video quality

B.2.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów zróżnicowanych efektów oceny jakości i użyteczności multimediów oraz sformułowanych wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- zróżnicowanie obiektów multimedialnych ze względu na specyfikę przekazywanej informacji,
- wiarygodne miary oceny jakości danych;
- charakterystykę procedur wykorzystywania przekazu, szczegółów percepcji treści i form użytkowania.



Rysunek B.4: Ocena jakości sekwencji wizyjnej metodą obserwacyjną z wykorzystaniem narzędzi MSU – obok narzędzia do projektowania testu obserwacyjnego pokazano przykładowe wyniki (góra) oraz fragment interfejsu do oceny wideo wraz ze skalą ocen (dół).

B.3 Kompresja multimediów

Celem ćwiczenia jest ocena efektywności wybranych metod kompresji danych multimedialnych, z uwzględnieniem optymalizacji algorytmów za pomocą doboru zestawu parametrów i eksperymentalnej weryfikacji skuteczności kodowania zbiorów danych testowych z wykorzystaniem gotowych implementacji kodeków. Spodziewanym efektem jest dyskusja wyników oraz formułowanie wniosków.

B.3.1 Program ćwiczenia

Przewidywana jest realizacja następujących zadań:

- charakterystyka kompresowalności danych źródłowych;
- porównanie efektywności odwracalnych kodeków danych;
- porównanie efektywności kodeków danych z selekcją informacji.

Charakterystyka kompresowalności danych źródłowych polega na ocenia podatności na kompresję wybranych rodzajów danych źródłowych. Należy skorzystać z referencyjnych zestawów danych o charakterze uniwersalnym, tzw. korpusów, np. Calgary, http://links.uwaterloo.ca/calgary.corpus.html, Cantenbury http://corpus.canterbury.ac.nz/ czy Silesia http://www.data-compression.info/Corpora/SilesiaCorpus/.

Podstawową miarą ilości informacji wyznaczającą granice kompresowalności jest entropia. Należy więc ocenić i porównać wartości entropii dla różnego typu danych testowych. Zadanie polega na wyznaczeniu wartości entropii przy różnych wielkościach alfabetu modelu źródła oraz kontekstu modelu Markowa, wykorzystując narzędzia WinEntropia oraz Entropia. Stosując różne metody odwracalnych przekształceń danych, jak np. tranformacja Burrowsa-Wheelera czy zmiana alfabetu danych, trzeba prześledzić ich wpływ na wartości liczonych entropii, a następnie omówić uzyskane rezultaty.

Porównanie efektywności odwracalnych kodeków danych polega w pierwszej kolejności na ocenie skuteczności kodowania archiwizerów, czyli uniwersalnych narzędzi do możliwie efektywnej kompresji różnego typu danych. Wśród testowanych kodeków powinny się znaleźć takie aplikacje jak: WinRK, UHBC, ABC, 7-Zip, WinRAR, WinZip, PAC, SBC, PAQ, KGB, Compressia i wiele innych, aktualnych. Miarą efektywności jest średnia bitowa uśredniona po zbiorze wszystkich plików korpusu testowego. Warto też zanalizować średnie uzyskiwane w poszczególnych kategoriach danych (obrazy, audio, dokumenty wielokomponentowe, programy wykonawcze, teksty, źródłowe pliki w językach programowania itp.). Podobny test można przeprowadzić na wybranych grupach kodeków dedykowanych określonemu rodzajowi danych, np. dla odwracalnych kodeków audio.

Porównanie efektywności kodeków danych z selekcją informacji służy selekcji oraz ocenie zróżnicowań skuteczności najpopularniejszych kodeków, dopuszczających zniekształcenia danych źródłowych. Na przykładzie obrazów pojedynczych (analogicznie można zrobić analizę wybranych kodeków audio czy wideo), należy przeprowadzić porównanie efektywności kompresji (poprzez zebranie krzywych stopnia zniekształceń R-D) wybranych kodeków (przykładowy zestaw: DPCM, VQ, DCT, SBC, EZW, Spiht, JPEG,JPEG2000) poprzez

- optymalizację (dobór parametrów) każdego z testowanych kodeków;
- porównanie efektywności zestawu zoptymalizowanych kodeków w zakresie ustalonych średnich bitowych;
- wskazanie optymalnej metody kompresji dla każdego z obrazów testowych.

W eksperymentach należy wykorzystać kilka typowych obrazów testowych (np. Lena, Goldhill, Barbara, Target, Baboon) generalizując wnioski z analizy wyników. Użytecznym narzędziem w realizacji tego zadania jest VcDemo http://siplab.tudelft.nl/content/image-and-video-compression-learning-tool-vcdemo.

B.3.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:

- narzędzia WinEntropia, Entropia – www.ire.pw.edu.pl/ arturp/Dydaktyka/Media_tools/,
- archiwizery i inne dostepne kodeki,
- VcDemo, IMLAB, CXIMAGE, VirtualDub i inne.

B.3.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów uzyskanych efektów kompresji multimediów oraz sformułowanych wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- kompresowalność różnego typu danych, w tym zbiorów łączących w sobie dane o różnym charakterze,
- możliwą do uzyskania efektywność kompresji odwracalnej w przypadku różnego typu danych – wartości uzyskanych średnich bitowych skompresowanych plików warto odnieść do granicznych wartości entropii, wyznaczonych dla tych samych plików źródłowych,

- uzyskiwane wartości stopni kompresji z selekcją informacji, przy zachowaniu wizualnej i odsłuchowej bezstratności tego procesu (tj. braku dostrzegalnych zniekształceń w stosunku do oryginału).

B.4 Indeksowanie danych multimedialnych

Celem ćwiczenia jest badanie schematów indeksowania treści multimedialnej oraz ocena efektów wyszukiwania z wykorzystaniem dobieranych zestawów deskryptorów. Istotne jest wykorzystanie wiarygodnych, reprezentatywnych i statystycznie istotnych zasobów multimedialnych z możliwie dokładną specyfiką ich zawartości. Opis eksperymentów, uzasadnienie zaplanowanych procedur testowych, dyskusja wyników i formułowanie wniosków są spodziewanym efektem realizacji tego ćwiczenia.

B.4.1 Program ćwiczenia

Generalnie ćwiczenie to dotyczący podstaw indeksowania i wyszukiwania treści multimedialnej. Przewidywana jest realizacja następujących zadań:

- ocena sposobów indeksowania;
- konstruowanie zapytań;
- wyszukiwanie multimediów po zawartości.

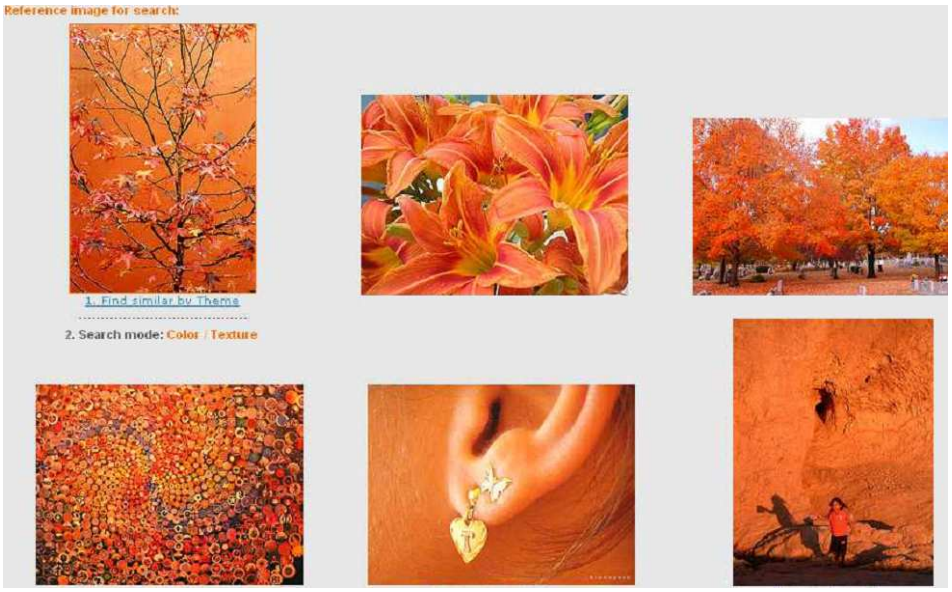
Ocena sposobów indeksowania sprowadza się do analizy wpływu doboru zestawu deskryptorów na efektywność wyszukiwania. Za pomocą wybranych wyszukiwarek dostępnych w sieci, np. Google, Muvis, TinEye, Tilmoto, Anaktisi, GazoPa, a także aplikacji dostępnych na stronie narzędzi www.ire.pw.edu.pl/ arturp/Dydaktyka/Media_tools/, wykorzystując bazy zróżnicowanych danych multimedialnych należy ocenić użyteczność różnych zestawów deskryptorów treści. Miarą ich efektywności jest selektywność wyszukiwania.

Konstruowanie zapytań dotyczy formułowania ciągu zapytań tekstowych (TBIR) oraz po zawartości obiektów (CBIR) lub też łączonych (hybrydowych). W przypadku obrazów zapytania mogą dotyczyć słów kluczowych, graficznego zarysu sceny (np. błękit u góry i żółty obszar na dole), wyboru kategorii możliwych odpowiedzi (logiczne operacje łączące np. samochody z kobietami i deszczem). Przy indeksowaniu zawartością typową formą jest zapytanie przez przykład, przy czym można tutaj wykorzystać własne zdjęcie czy inną formę obrazu, jak też wskazać jeden z obrazów dostępnych w bazie, wyszukanych po wstępnym zapytaniu słowem kluczowym.

W przypadku dźwięku mogą to być słowa zapytania dotyczące rodzaju muzyki, tytułu czy nazwiska wykonawcy. Zapytaniem może być też prosty motyw muzyczny skomponowany za pomocą wygodnego edytora czy też zbiór z nagraniem mowy identyfikowanej osoby lub zapisem muzyki, itp.

Wyszukiwanie multimediów po zawartości koncentruje się na weryfikacji efektów wyszukiwania w kontekście oczekiwań odbiorcy. Ważny jest tutaj przede wszystkim sposób selekcji i porządkowania odpowiedzi, przy określonych kryteriach podobieństwa treści oraz sposobach wyrażania satysfakcji z udzielonej odpowiedzi. Istotne w tej ocenie jest zbieranie subiektywnych ocen użytkowników za pomocą przygotowanych skal użyteczności odpowiedzi.

Po zinterpretowaniu zbioru odpowiedzi należy oszacować selektywność wyszukiwania za pomocą takich miar, jak precyzja, stopa sukcesu i, w miarę możliwości, przywołanie. Porównując wyniki dla różnych zestawów deskryptorów można ustalić korzystne warunki wyszukiwania określonej treści obrazowej. Przykładowe efekty wyszukiwania zamieszczono na rys. B.5.



Rysunek B.5: Przykładowe rezultaty wyszukiwania: zapytanie znajduje się w lewym górnym rogu, natomiast wszystkie z ukazanych odpowiedzi są poprawne (spełniają kryterium podobieństwa zbliżonego koloru i tekstury).

B.4.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:

- narzędzie – wyszukiwarki sieciowe:
  VideoGoogle (http://www.robots.ox.ac.uk/ vgg/research/vgoogle),
  CIRES (http://amazon.ece.utexas.edu/ qasim/research.htm),
  MUVIS (http://muvis.cs.tut.fi),
  Tiltomo (http:/www.tiltomo.com),
  Retrievr (http://labs.systemone.at/retrievr/),

http://xcavator.net
TinEye (http:/www.tineye.com) i inne.

B.4.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów uzyskanych efektów wyszukiwania treści multimedialnych i sformułowanych wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- podobieństwo wyszukiwanych obiektów wskazywane (oceniane) przez użytkowników,
- charakter błędów popełnianych przy różnych formach zapytań,
- analizę przyczyn braku skuteczności różnych schematów wyszukiwania.

B.5 Komputerowe przetwarzanie danych multimedialnych

Celem ćwiczenia jest badania wybranych metod komputerowego przetwarzania multimediów, przede wszystkim metod przetwarzania oraz analizy obrazów i dźwięku. Zadaniem jest porównanie efektywności różnych metod przetwarzania, podjęcie próby optymalizacji uzyskiwanych efektów oraz sformułowania wniosków dotyczących realnych możliwości zwiększenia walorów przekazu informacji wskutek poprawy cech użytecznych dla odbiorcy.

B.5.1 Program ćwiczenia

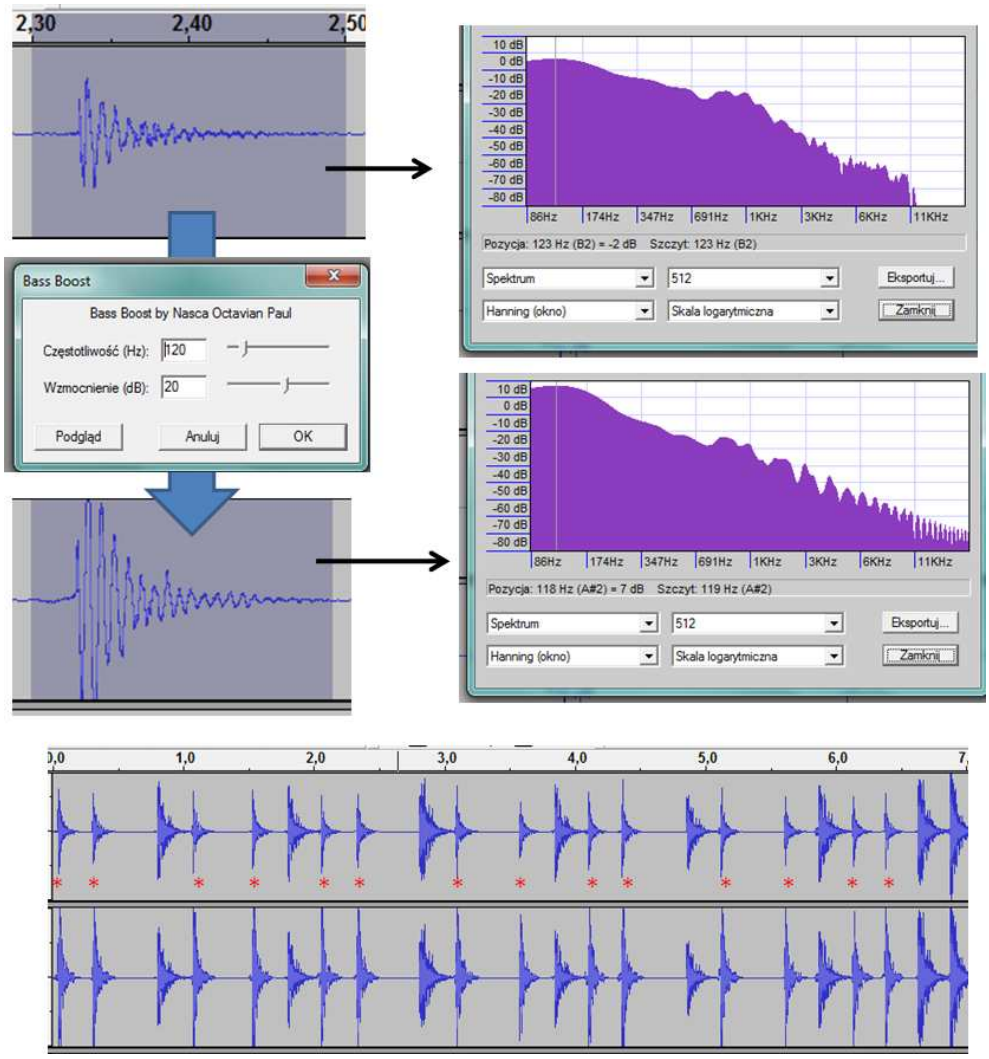
Przewidywana jest realizacja następujących zadań:

- charakterystyka treści danych źródłowych;
- przetwarzanie dźwięku (analiza widma, filtracje, efekty dodatkowe)
- przetwarzanie obrazów (korekcje histogramu, filtracje, segmentacja)

Charakterystyka treści danych źródłowych polega na wydzieleniu obiektów, wzorców, specyficznych elementów przekazu, których postrzeganie ma znaczenie i wpływa na zrozumienie treści przez odbiorcę. Na przykładzie wybranych obrazów i zapisów dźwięku należy opisać różnorodne postacie obiektów czy wzorców, możliwą hierarchiczność i skalowalność takiego opisu, związek pomiędzy rozpoznawaną treścią a informacją użyteczną w przekazie.

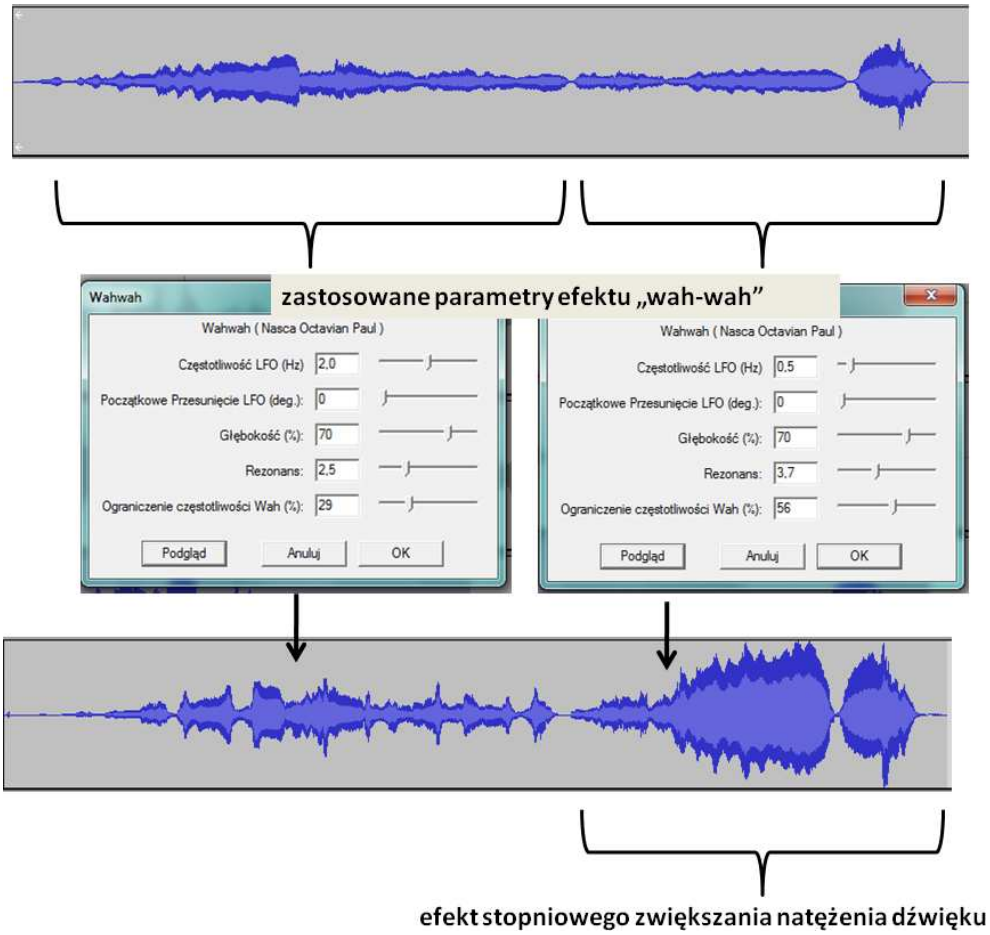
Przetwarzanie dźwięku sprowadza się do prostego zadania obserwacji wpływu różnych postaci filtrów na jakość dźwięku, jego odszumienie (korektę szumów), korektę błędów rejestracji uwydatnienie niektórych tonów, zmianę dynamiki, wysokości poszczególnych dźwięków, modulację brzmienia poszczególnych zapisów (np. wybranych instrumentów), przy czym istotne jest porównanie wrażenia subiektywnego z obliczeniową oceną jakości w postaci przebiegu czasowego, widma częstotliwościowego, statystyki, sonogramu. Można przeprowadzić krótki test subiektywny oceny jakości dźwięku z pomocą kilku słuchaczy, badając np. wpływ redukcji pasma lub podbijania niektórych składowych widma na jakość odbieranego dźwięku.

Dodatkowo, należy obserwować użyteczność efektów dodatkowych: pogłosu, studni, chóru, opóźnienia, itp. jako formy udoskonalenia zapisu dźwiękowego. Przykładowy efekt wzmocnienia brzmienia niższych dźwięków za pomocą funkcji *Bass Boost* pokazano na rys. B.6. Na rys. B.7 przedstawiono z kolei efekt modulacji brzmienia wybranej melodii za pomocą efektu "kaczki" (*wah-wah*), czyli przesuwania filtru przetwarzającego dźwięk po skali częstotliwości w czasie trwania dźwięku oraz poprzez stopniowe wzmocnienie dźwięku w zadanym fragmencie.



Rysunek B.6: Fragment rejestracji zawierający pojedyncze uderzenie przed i po operacji *Bass Boost*, przy częstotliwości wzmacnianej 120 Hz; wzmocnienie tej składowej widma częstotliwościowego (amplitudowego) widać na wykresach z prawej strony. Dolny panel zawiera dłuższy fragment rejestracji przed i po operacji – uderzenia poddane operacji *Bass Boost* zaznaczono czerwoną gwiazdką; źródło: Dominika Malińska, sprawozdanie z projektu Technik Multimedialnych.

Zadanie obróbki dźwięku może być rozszerzone na sygnały jednowymiarowe. W tym celu należy wykorzystać filtry cyfrowe o różnorodnej charakterystyce, operacje na widmie częstotliwościowym, zmiany w zakresie fazy i amplitudy harmonicznych, nakładanie dwóch sygnałów o różnym widmie - efekt dyspersji, interferencje dźwięków pochodzących z kilku źródeł – zestaw apletów realizujących podobne funkcje jest dostępny pod adresem:
http://www.falstad.com/mathphysics.html.



Rysunek B.7: Efekty ”kaczki” oraz stopniowego wzmacniania dźwięku, zastosowane do modyfikacji brzmienia ścieżki z prowadzącą melodią skrzypiec; na poszczególnych fragmentach zastosowano różne filtry *wah-wah*: filtr w części drugiej charakteryzuje się wolniejszą oscylacją przebiegającą wokół wyższej częstotliwości niż filtr z części pierwszej; źródło: Dominika Malińska, sprawozdanie z projektu Technik Multimedialnych.

Przetwarzanie obrazów polega przede wszystkim na ocenie efektów przetwarzania kilku obrazów o zróżnicowanej jakości oraz przekazywanej informacji. Obok prostych metod korekcji histogramu, filtracji liniowej i nieliniowej, służącej odszumieniu oraz poprawie widoczności cech obiektów (krawędzi, kształtów, tekstur), należy poddać analizie także metody segmentacji obrazów. Wśród testowanych metod powinny znaleźć miejsce: proste progowanie z doborem wartości progu, grupowanie metodą k-średnich, rozrost regionów, metoda wododziałowa, metoda aktywnych konturów czy kształtów. Przykładowe efekty obróbki obrazów pokazano na rys. B.8. Interesującym zagadnienie jest też dobór odpowiedniej metody przetwarzania wstępnego (redukcja szumów, podkreślenie gradientów) w celu uzy-

skania lepszych efektów segmentacji obiektów obrazu bądź też ich klasyfikacji (po doborze właściwej przestrzeni cech).

Rysunek B.8: Przykłady przetwarzania i analizy obrazu: w górnym szeregu znajduje się obraz źródłowy, efekt wykrywania krawędzi oraz przetwarzania za pomocą metod morfologii matematycznej (otwarcie), zaś w dolnym szeregu efekty segmentacji przy różnej liczbie klas obiektów.

B.5.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:

- narzędzia MDI, Wavosaur lub inne
- edytor dźwięku Audacity http://audacity.sourceforge.net/
- narzędzia IMLAB, CXIMAGE, MaZDa, VirtualDub

B.5.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów uzyskanych efektów przetwarzania multimediów oraz sformułowanych charakterystyk, spostrzeżeń i wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- zróżnicowanie przekazywanych treści multimedialnych ze względu na warunki akwizycji, jakość, charakter występujących obiektów czy wzorców, ich

właściwości, wymowę całości, przeznaczenie – sposób wykorzystania (użytkowania) itp.,

- potencjał poszczególnych metod przetwarzania multimediów, możliwości kaskadowego łączenia metod, poziom uzyskiwanej poprawy postrzegania treści,
- efektywność rozważanych metod wydzielania obiektów oraz możliwości różnicowania ich kształtu, krawędzi, tekstury,
- rolę przetwarzania jako uzupełnienie procesu akwizycji informacji w kontekście różnych aplikacji.

B.6 Użytkowanie multimediów

Celem ćwiczenia jest poznanie wybranych właściwości standardów rodziny JPEG i MPEG poprzez przyswojenie podstaw teoretycznych, eksperymentalną weryfikację efektów działania narzędzi implementujących normy oraz analizę wyników i sformułowanie wniosków.

B.6.1 Program ćwiczenia

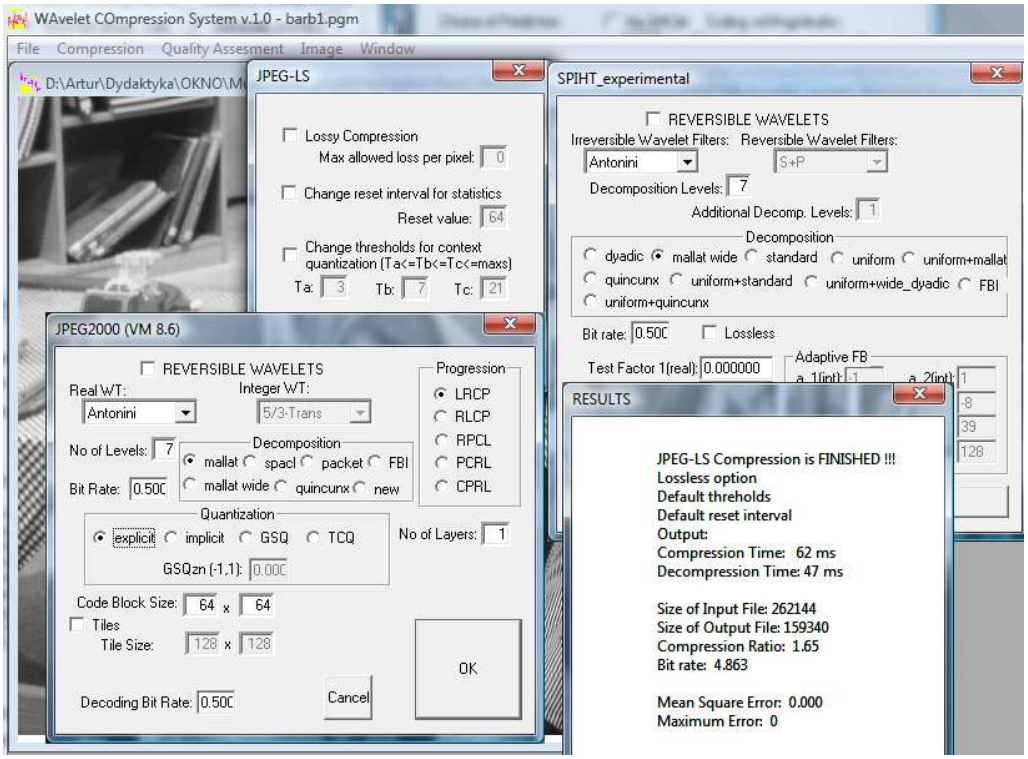
Przewidywana jest realizacja następujących zadań:

- optymalizacja kodeka JPEG-LS (kodowanie odwracalne, dobór parametrów, porównanie z innymi kodekami)
- optymalizacja kodeków JPEG i JPEG2000 (kodowanie nieodwracalne, dobór parametrów, porównanie z innymi kodekami)
- optymalizacja kodeków sekwencji wizyjnych (porównanie implementacji MPEG-2 i H.264, dobór parametrów)
- optymalizacja efektów wyszukiwania na bazie deskryptorów MPEG-7 (analiza bazy danych, ustalanie kryteriów podobieństwa, dobór deskryptorów)

Optymalizacja kodeka JPEG-LS dotyczy w pierwszej kolejności doboru parametrów kwantyzacji kontekstu: *Reset value* oraz wartości progów T1, T2 i T3. Eksperymenty powinny być przeprowadzone na zbiorze testowym kilku obrazów o zróżnicowanych charakterze. Ustaloną, efektywną wersję kodeka należy następnie porównać z innymi kodekami odwracalnymi, wyznaczając średnią bitową BR uśrednioną po wszystkich obrazach testowych (ewentualnie w podgrupach obrazów naturalnych, medycznych, sztucznych, specjalistycznych, itp.) oraz średnie czasy kompresji/dekompresji. Można skorzystać z implementacji kodeka JPEG-LS pod nazwą JLS Encoder (*University of British Columbia*) lub też z narzędzia WACOS (rys. B.9). Wśród kodeków porównywanych z JPEG-LS mogą się znaleźć JPEG2000 oraz formaty dostępne w narzędziach IMLAB oraz CXIMAGE.

Dodatkowo, należy przeprowadzić optymalizacje kodeka COMPUT, gdzie dostępne są trzy rodzaje kodowania (zobacz rys. B.10):

- z 2W skanowaniem (emphScaning_Encoding), gdzie można dobrać porządek przeglądania pikseli oraz metodę 1W kodowania,
- z 2W modelem predykcji (emphPrediction-ContextEncoding), gdzie można ustalić postać modelu oraz 1W metodę kodowania,
- z 2W modelem predykcji oraz kwantyzacją 2W kontekstu (emphPrediction-ContextEncoding), gdzie można ustalić postać modelu predykcji oraz 2W

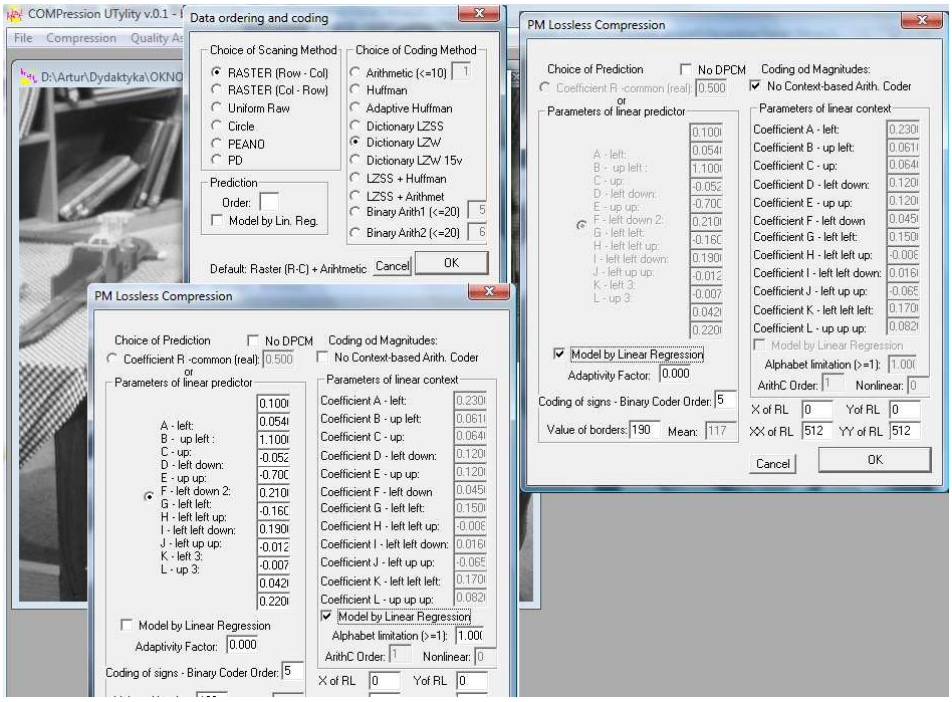


Rysunek B.9: Interfejs programu WACOS umożliwiający dobór parametrów kodeków bezstratnych JPEG-LS, JPEG2000 oraz SPIHT oraz generujący raport końcowy procesu kompresji/dekompresji (czasy kodowania/dekodowania, odpowiednie rozmiary plików, wartość średniej bitowej, weryfikacja odwracalności procesu kompresji).

kontekst stosowany w modelowaniu obrazu na użytek kodowania arytmetycznego.

Najlepsze wśród uzyskanych wyników należy porównać z tymi, które otrzymano dla JPEG-LS i innych kodeków, przeprowadzić dyskusje, wyciągnąć wnioski.

Optymalizacja kodeków JPEG i JPEG2000 obejmuje w pierwszej kolejności dobór parametrów każdego z kodeków, a następnie porównanie efektywności kodowania według algorytmów norm JPEG i JPEG2000 na zbiorze obrazów testowych (dobranych tak, by były zróżnicowane pod kątem ilości szczegółów, wielkości obiektów zainteresowania, rodzaju tła). Efektywność kodowania należy ocenić za pomocą miar obliczeniowych (PSNR, MSE, maksymalna różnica punkcie), a także w teście oceny subiektywnej (można to tego celu wykorzystać narzędzie MSU perceptual video quality). Warto zwrócić uwagę na odmiennych rodzaj zniekształceń wprowadzanych w obu kodekach standardowych.



Rysunek B.10: Interfejs programu COMPUT umożliwiający dobór parametrów bezstratnych kodeków, konstruowanych według 3 schematów uwzględniających właściwości obrazów: 2W skanowanie (okno u góry), 2W predykcję (po prawej) oraz 2W predykcję z kwantyzacją 2W kontekstu (u dołu).

Wśród dobieranych parametrów kodeka JPEG warto zwrócić uwagę na postać tablicy kwantyzacji dla luminancji i chrominancji, rodzaj kodowania skwantowanych współczynników (kodem Huffmana lub arytmetycznym), wpływ konwersji źródłowej przestrzeni barw RGB na YCrCb oraz podpróbkowania składowych chrominancji.

Znacznie większa liczba stopni swobody w optymalizacji kodeka JPEG2000 obejmuje przede wszystkim: postać filtrów falkowych i liczba poziomów dekompozycji, rozmiar bloków kodowych, rodzaj transformacji przestrzeni barw, schemat kwantyzacji (z różnym rozkładem wartości przedziału kwantyzacji w poszczególnych skalach), rodzaj progresji (brak, zorientowana na warstwę, rozdzielczość, ROI, komponent, liczba warstw), parametry kodowania arytmetycznego (rodzaje modeli kontekstowych, częstość czyszczenia modelu, możliwość stosowania trybu bez kodowania).

Dodatkowo, interesującym eksperymentem jest porównanie odporności na zakłócenia strumieni JPEG i JPEG2000 oraz ocena wpływu parametrów kodowania na tę odporność.

Optymalizacja kodeków sekwencji wizyjnych dotyczy przede wszystkim oceny możliwości kodowania według MPEG-2 oraz H.264, badania występujących ograniczeń oraz potencjału umożliwiającego dalszy rozwój tych standardów. Głównym przedmiotem optymalizacji mają być schematy estymacji i kompensacji ruchu o charakterze międzyramkowym, ale też wewnątrzramkowym. Ustalenie sposobu kodowania poszczególnych ramek, by redukcja nadmiarowości czasowej oraz przestrzennej była możliwie duża, może być uzupełnione regulacją pozostałych parametrów kodowania (rodzajem filtru wygładzającego nieprzyjemne efekty blokowe, schematem kodowania, próbkowaniem przestrzeni barw, rodzajem przekształceń dziedziny obrazu, itp.).

Optymalizacja efektów wyszukiwania na bazie deskryptorów MPEG-7 Za pomocą narzędzi zawierających implementacje kilku podstawowych deskryptorów wizyjnych, wykorzystując bazę zróżnicowanych obrazów o określonych właściwościach należy sformułować ciąg zapytań. Następnie, po zinterpretowaniu zbioru odpowiedzi oszacować selektywność wyszukiwania za pomocą takich miar jak precyzja, przywołanie, stopa sukcesu. Porównując wyniki dla różnych zestawów deskryptorów można ustalić korzystne procedury wyszukiwania określonej treści obrazowej.

Dostępne są deskryptory koloru, tekstury, kształtu, tła sceny, itd., które można przetestować dla różnych danych testowych pod kątem wymagań wysokiej skuteczności wyszukiwania. Ponadto, złożenie różnego rodzaju deskryptorów w jeden schemat indeksujący, z uwzględnieniem potrzeb konkretnej aplikacji, służy rozwiązaniu bardziej uniwersalnemu i niezawodnemu.

B.6.2 Wykorzystywane narzędzia

Aby zrealizować poszczególne zadania, należy wykorzystać odpowiednio:

- narzędzia JLSEncoder, JPER2000 oraz COMPUT, WACOS i inne
- narzędzia JPER2000, WACOS, MJPEG oraz VcDemo i inne
- narzędzia VcDemo, VirtualDub
- narzędzie PI

B.6.3 Sprawozdanie

Obok rezultatów przeprowadzonych badań, wybranych – interesujących przykładów uzyskanych efektów kompresji i indeksowania multimediów oraz sformułowanych wniosków, podczas omawiania najistotniejszych wyników należy zwrócić uwagę przede wszystkim na:

- ograniczony potencjał metod kompresji bezstratnej,

- problem poziomu dopuszczalności zniekształceń oraz ich zróżnicowanie dla różnych algorytmów kompresji stratnej,
- ograniczenia metod wyszukiwania po zawartości obrazów
- dostosowane do możliwości wykorzystanie w aplikacjach.