

PODSTAWY TEORII INFORMACJI

- o źródłach, jednostkach i odkrywcach

Artur Przelaskowski

Warszawa, kwiecień 2022

(z wykorzystaniem fragmentów z A. Przelaskowski, Techniki Multimedialne,
OKNO 2016)

Wprowadzenie

Przedmiotem tego opracowania są fundamenty teorii informacji. Jako pojęcie pierwotne, na dodatek o stale odkrywanym, wręcz przełomowym znaczeniu dla zrozumienia Wszechświata i nieustannie odkrywającego największe tajemnice człowieka, informacja jest nieustannie definiowana na nowo, ale jej znaczenie rośnie.

Zwycięsko konkurująca z masą i energią w wyścigu o sens, wolność i zwycięstwo coś nad nic.

Jako informacja biologiczna została zaszyta na Początku w kodzie DNA, jak słowo ożywiające i napędzające rozwój istot żywych, kształtujące wielowiekowo ich doskonałość, nadające kształt istnieniu.

Jednocześnie nadaje sens najnowocześniejszym technologiom szóstej fali rewolucyjnej, naśladując zmysły w ich różnorodności wzmacniając zasięg i możliwości poznawcze.

A przy tym ciągle tajemnicza, ulotna, niepoznana, choć stale odkrywana, silniejsza i trwalsza od każdej formy ziemskiego istnienia, wieczna.

Struktura tej opowieści obejmuje przede wszystkim:

- próby definiowania informacji, czyli metaforycznie o największych tajemnicach i wyzwaniach;
- proste, formalne modele, opisy, struktury i schematy informacji, czyli konkretnie, fizycznie i prosto o narzędziach informacji;
- praktyczne ograniczenia informacji ustalane przez sygnały – mierzone i przetwarzane jako realne nośniki abstrakcji odwołującej się do zmysłów i zdolności umysłu;
- charakterystykę sygnałów (wybranych klas, koncepcję rozwinięć w bazach o modelowanych walorach użytkowych) – ”namacalnego” ciała informacji, za pomocą których można transformować poznanie i wiedzę;

- modele użytkowników, czyli opowieść o człowieku, na którą składa się ocena znaczenia i roli informacji względem realnych wyzwań i konsekwencji.

Warszawa, kwiecień 2022

Oznaczenia

Oznaczenia matematyczne

Zbiory

- \mathbb{N} – liczby naturalne
- \mathbb{Z} – liczby całkowite
- \mathbb{R} – liczby rzeczywiste
- \mathbb{R}^+ – liczby rzeczywiste dodatnie
- \mathbb{C} – liczby zespolone
- $|A|$ – liczba elementów zbioru

Operacje

- $\|f\|$ – norma
- $\langle f, g \rangle$ – iloczyn skalarny
- $f * g(t)$ – operacja splotu ciągłego w czasie
- $f * g_t$ – operacja splotu dyskretnego

Probabilistyka

- F – zmienna losowa o realizacji f
- $\mathbf{F} = \{F_1, F_2, \dots\}$ – pole losowe (ogólniej łańcuch losowy) o realizacji \mathbf{f}
- $E\{F\}$ – wartość oczekiwana zmiennej losowej F
- $R_F(k)$ – funkcja autokorelacji stacjonarnego procesu (łańcucha) losowego \mathbf{F}
- $\mathcal{K}_{m,n} = Cov(F_m, F_n)$ – macierz kowariancji wektora (łańcucha) losowego \mathbf{F}

Sygnały i funkcje

- $f(t)$ – sygnał (funkcja) ciągły; analogicznie ciągła funkcja wartości obrazu to $f(x, y)$

f_t – sygnał dyskretny (próbki), zaś dla obrazu dyskretnego mamy wartości pikseli $f_{x,y}$; analogicznie współczynniki filtru to g_n
 \hat{f} – wartość przewidywana (w predykcji za pomocą funkcji) sygnału f
 \tilde{f} – wartość aproksymowana (przybliżona) sygnału f

Przestrzenie

$L^2(\mathbb{R})$ – funkcje o skończonej energii, tj. $\int_{t \in \mathbb{R}} |f(t)|^2 dt < +\infty$
 $l^2(\mathbb{Z})$ – sygnały dyskretno o skończonej energii, tj. $\sum_{t \in \mathbb{Z}} |f_t|^2 < +\infty$
 C^m – przestrzeń funkcji mających ciągłą pochodną rzędu m
 \mathcal{H} – przestrzeń Hilberta
 $U \oplus V$ – suma prosta dwóch przestrzeni wektorowych

Informacja

A_S – alfabet źródła informacji S
 S^N – rozszerzenie źródła S stopnia N
 $H(S)$ – entropia łączna źródła S
 $H(S | C)$ – entropia warunkowa źródła względem kontekstu C
 $C^{(m)}$ – kontekst źródła Markowa rzędu m
 $I(S_1; S_2)$ – średnia informacja wzajemna źródeł S_1 i S_2

Transformacje

$\mathcal{F}\{f(t)\} = F(\omega)$ – transformacja Fouriera sygnału $f(t)$
 $\mathcal{F}\{f_t\} = F_k$ – dyskretna transformacja Fouriera sygnału f_t
 $\mathcal{L}\{f_t\} = f(z)$ – transformacja z sygnału f_t ,
 $\mathcal{S}\mathcal{F}\{f(t)\} = Sf(d, t)$ – okienkowa (krótkoczasowa) transformacja Fouriera sygnału $f(t)$
 $\mathcal{W}\{f(t)\} = Wf(s, t)$ – transformacja falkowa sygnału $f(t)$
 ϕ – funkcja skalująca
 ψ – falka

Kompendium

Słowo kompendium pochodzi od łacińskiego wyrazu *compendere* oznaczającego "ważyć (coś) razem". Jest to w miarę kompletny, zwięzły i konkretny zasób wiedzy na wybrany temat, stanowiący swego rodzaju jej "esencję", wykładnik czy praktyczne, konkretne przełożenie. Kompendium może także stanowić zwarte podsumowanie jakiegoś większego dzieła lub pracy naukowej. W większości przypadków koncentruje się na wybranych, kluczowych zagadnieniach będących przedmiotem badań, studiów lub innych form konkretnych zainteresowań (np. hydrologii, medycyny, metrologii etc.). W tym kontekście encyklopedia może być rozumiana w swej uniwersalnej istocie jako kompendium wszystkich kompendiów, obejmującą wiedzę z szerokiego wachlarza dziedzin poznania.

Altruizm

Altruizm to pomaganie innym, dzielenie się z innymi, informowanie innych.

Bóg

Godne podziwu jest, że nigdy żaden autor kanoniczny nie posłużył się naturą, by dowieść Boga. Wszyscy starają się wzbudzić wiarę w Niego. Dawid, Salomon itd., żaden z nich nie powiedział: "Nie ma próżni, zatem jest Bóg". W czym okazali się bystrzejsi od najbystrzejszych ludzi późniejszych czasów, którzy wszyscy się tym posługiwali. To bardzo doniosłe. Natura pomieściła każdą swoją prawdę jedynie w niej samej (Blaise Pascal).

Czas

Czas jest substancją, z której jestem stworzony (Jorge Luis Borges).

Estymacja

Estymacja to ogólnie określanie przybliżonej wartości (szacowanie, przybliżanie).

W statystyce matematycznej oznacza określanie (szacowanie) parametrów całej populacji na podstawie badania jej próby losowej (przeciętna, wariancja itp.). Celem jest uogólnienie wyników badania próby losowej na **nieznaną** postać i parametry rozkładu zmiennej losowej całej populacji oraz szacowanie błędów wynikających z tego uogólnienia.

Estymacja parametryczna obejmuje metody znajdowania nieznanymi wartościami parametrów rozkładu, podczas gdy nieparametryczna dotyczy metod wyznaczania nieznanymi postaci rozkładu populacji. Estymacja nieparametryczna jest często zastępowana prostszymi metodami bazującymi na weryfikacji hipotez statystycznych, w których najpierw stawiamy przypuszczenie dotyczące rozkładu, a następnie sprawdzamy ich słuszność (co zwykle oznacza brak przesłanek do odrzucenia przyjętej hipotezy).

Informacja

Informacja to coś, czego nie wiedzieliśmy przedtem, co ogranicza poziom niepewności, zaspokaja ciekawość, pomaga rozwiązać problem, urealnia sytuację, jest obiektem oczekiwań, otrzeźwia, wyrzywa ze snu, pomaga ...

Informacja jest różnicą, która robi różnicę (G.Bateson, antropolog kulturowy).

Informacja nie jest ani materią ani energią (Norbert Wiener, matematyk)

Nie da się w żaden sposób wyrazić pojęcia informacji - pojęcia przekazywania wiadomości, w kategoriach pojęć fizyki i chemii, nawet jeśli ta ostatnia wyjaśnia działania mechanizmów molekularnych (DNA, RNA, białka) w przenoszeniu informacji (Arthur Peacocke, brytyjski teolog i naukowiec).

Bernd-Olaf Koppers (fizyk): problem powstania życia co do istoty swej jest najwyraźniej równoznaczny z problemem powstania informacji biologicznej Manfred Eigen (biofizyk): naszym zadaniem jest znalezienie algorytmu, prawa przyrody, które prowadzi do powstania informacji Kurt Gödel: cechą organizmów żywych jest jakiś rodzaj zachowania informacji - złożoność organizmów żywych musi być obecna w materiale, z którego powstały albo w prawach rządzących ich powstawaniem; mechanizm (ewolucyjny) w biologii to wypaczony pogląd naszych czasów, a wytłumaczy to matematyka (jest nieprawdopodobny!) Albert Einstein: nie fakty zmuszają nas do wysunięcia nowej teorii, ale nade wszystko twórczy akt wyobraźni; każde poznanie rzeczywistości za pomocą myśli - odbiór informacji - musi zdać się na metafizyczny impuls

Instykt

Instynkty to silne, wrodzone wzorce zachowań, których nie musimy się uczyć. Są dziedziczne. Uaktywniają się niemal niezależnie od naszych zachowań. To nie tylko domena zwierząt, ale dla człowieka wydają się czymś wstydlwym, gorszym (przeciwieństwo działań rozumnych ujarzmiających nasze instynkty). To są nieświadome automaty przetwarzające informacje bez wysiłku, głęboko wbudowane w nasze mózgi. Takie nieświadomione zdolności można także w sobie rozwijać (doskonalenie instynktów): przykład sekserów kurcząt, czy też obserwatorów samolotów różnicujących maszyny wrogie i 'nasze' bez wyróżnienia świadomych cech. Przykłady: fascynacja seksualna, strach przed ciemnością, współczucie, zazdrość, rozpoznawanie nastroju po mimice.

Inteligencja

To cecha umysłu obejmująca świadomość i emocjonalność; to zdolność rozwiązywania praktycznych problemów, radzenia sobie w życiu (i po?).

Z inteligencją związane są szczególne zdolności: samoświadomość, pojmowanie, rozumienie, postrzeganie, uczenie się, analiza, rozpoznawanie, adaptacja do zmian, operowanie symbolami, posługiwanie się językiem, rozwiązywanie problemów, twórczość. Na inteligencję składa się wiedza emocjonalna, kreatywność, myślenie krytyczne, zdolność planowania.

To umiejętność pozyskiwania i wykorzystania informacji, kreatywne ich przetwarzanie w wiedzę (weryfikacja, obiektywizacja), umiejętność trafnej oceny i adaptacji. Informacja daje poznanie, a inteligencja pozwala na jej skuteczne wykorzystanie. Inteligencja wyrasta z codziennych doświadczeń, jest dynamiczna.

Inteligencja to podstawowa zdolność umysłu, decydująca o sukcesie jednostki w "walce o byt" (F. Galton, 1883).

Inteligencja to zdolność do wydawania trafnych sądów, zwł. w odniesieniu do problemów i sytuacji dnia codziennego; siedem pierwotnych zdolności umysłowych, związanych ze sprawnością w zakresie: rozumienia słów, używania słów (płynność słowna), posługiwania się liczbami, zapamiętywania, szybkości spostrzegania, rozumowania przez indukcję, wyobraźni przestrzennej (A. Binet, 1905).

Inteligencja maszynowa (sztuczna) naśladuje inteligentne działania/zachowania człowieka. Może być słaba (do realizacji określonego, ograniczonego w złożoności i poziomie abstrakcji zadania) lub silna (kopiując ludzkie zdolności poznawcze służy rozwiązywaniu realnych, złożonych problemów bez jego udziału, tak samo sprawnie). Ponadto, kognitywiści zakładają możliwość opracowania maszynowego umysłu (teoria umysłu) posiadającego wszystkie atrybuty ludzkiego umysłu, o zdolnościach do marzeń, refleksji, odczuwającego, myślącego, oczekującego, snującego refleksje itp. John Searle sugerował, że komputer może być rzeczywistym

umysłem¹

Inteligentny projekt

Informacje semantyczne to nie tylko określona treść. Ta treść odnosi się zarówno do formalnej formy, ciągu słów nad określonym alfabetem, ale też do obiektywnych zasobów wiedzy, dokonań naukowych, ich pragmatycznego znaczenia itd. Uwzględnia także swobodne, moralne, racjonalne działanie i do niego się odnosi. Inteligentny agent może "zaprogramować" zwierzęta -by reagowały na źródła pożywienia, roboty-by szukały źródeł doładowującej ich energii. Konieczne do tego inteligentnie zaprojektowane czujniki organizmów (robotów) mają wstępnie zaprogramowane reakcje. Inteligentni agenci mogą tworzyć złożone dzieła, samochody, statki kosmiczne, może nawet prostą muzykę. Inteligentni agenci mogą wypełniać dokumenty, korzystając z "rozsądku" i zasad logiki.

Intuicja

Intuicja (łac. intuitio – wejrzenie; wewnętrzne przekonanie, że masz rację) to zdolność do nagłego przeblysku myślowego, w którym dostrzega się rozwiązanie problemu lub znajduje odpowiedź na nurtujące pytanie. Pozwala na szybkie dopasowanie rozwiązania problemu do zaistniałych uwarunkowań. To nie emocje, a raczej proces podświadomy, którego nie sposób kontrolować, to przewidywanie, domyślanie się nie wynikające z klarownych reguł wnioskowania. Jest procesem bardziej kreatywnym i działającym na wyższym poziomie abstrakcji w porównaniu do myślenia logicznego. W praktyce, można jedynie dopuszczać lub odrzucać sugerowane przez intuicję rozwiązania.

Intuicja (*rozum, który się śpieszy*) jest cechą aktywnej inteligencji, której zwykle nie towarzyszy świadoma myśl. To zdolność bezpośredniego pojmowania, dotarcia do bezpośredniej wiedzy bez udziału obserwacji czy rozumu. Myślenie intuicyjne jest podobne do percepcji, czyli błyskawiczne i bez wysiłku. Wgląd, olśnienie pojawiające się w trakcie rozwiązywania problemu, przecucie, szybkie rozpoznanie, pojawiające się przekonanie sięgające do istoty problemu, którego nie jesteśmy w stanie uzasadnić. Przyjmuje się, że myślenie intuicyjne wykorzystuje 3 zasadnicze zjawiska: – mimowolne uczenie się, automatyczne, nieuświadomione zdobywanie wiedzy; – automatyzm zachowania, sztywne sposoby reagowania wynikające z doświadczenia życiowego (twórcze wykorzystanie doświadczenia); – markery somatyczne ostrzegające organizm (odczucia generowane na podstawie wtórnych emocji, wcześniejszych doświadczeń, to fizjologiczny sygnał ostrzegawczy o przewidywanych skutkach podjęcia decyzji).

¹John Searle, Minds, brains, and programs w Behavioral and Brain Sciences, vol. 3 Behavioral and Brain Sciences 3(3):417 –424, 1980

Matematyka

Gödel udowodnił, że w matematyce całość jest zawsze większa niż suma części (Freeman Dyson).

Metafora

Metafora, inaczej przenośnia, czyli przeniesienie znaczenia słów/pojęć/znaczeń na inne, by poszerzyć/uwydatnić ich rozumienie/doświadczenie, np. "wychodzimy z dołka", "nie wyrabiam się na zakrętach", "zbijać bąki", "przenosić góry", "nie opuszczę cię aż do śmierci"; metafora to sens(temat) i nośnik (tak jak informacja); to konkret wnioskujący abstrakcję. Metafora zawsze towarzyszy myśleniu abstrakcyjnemu, stanowi kluczowy mechanizm kognitywny.

Schemat metafory: struktura (nośnik) – forma (multimodalność) – zespoły funkcjonalne (pojęcia i relacje)–symulacje (wyobrażenia)-parametryzacja– obliczenia (neuralne)

Zastosowanie metafory: dziedzina źródłowa – dziedzina docelowa, przykład: "argumentowanie to wojna".

Nauka

Nauka tłumaczy prawa fizyki i biologii rządzące przyrodą, nie wyjaśnia jednak, skąd się one wzięły, kto je spisał, natchnął. Stoi więc jakby na glinianych nogach, nie sięga do źródeł informacji.

Gilbert Keith Chesterton pisał w "Eugenika i inne zło" o szczególnych działaniach naukowców – badaczy dziedziczości: rozumieją wszystko, co dotyczy ich dziedziny, ale nie znają jej zakresu. Jak przypuszczam urodzili się i wychowali w tej gęstwinie faktów i naprawdę ją badają, nigdy jednak nie dochodząc do jej granic. Innymi słowy, zgłębiają wszystko nie wiedząc, co zgłębiają. Platon, za Pitagorasem sugerującym trójkąt ojciec–matka–dziecko² wierzył w geometryczną liczbę panującą nad rodzeniem jednostek lepszych i gorszych. Arystoteles poszedł dalej, odkrył informację, która zamienia się w materię, a materia znów w informację, która rodzi materię ... i tak powstają pokolenia. Mężczyzna przekazuje kobiecie komunikat (instrukcje budowy), który natrafia na surowiec dostarczony kobiecie. Dziedziczość to przekazywanie informacji, która pozwala budować organizm od zera. Organizm zaś dojrzewa przekształcając materię w komunikat. Dziś już wiemy o DNA, to kod tej informacji życia. Informacja jest niematerialną abstrakcją, kod (model plus reguła) ją urealnia tworząc komunikat (określoną reprezentację). Gen to jednostka informacji biologicznej zapisana w żywych komórkach (chromosomy). Aby poznać świat, trzeba więc poznać elementarne jednostki

²S.Mukherjee, The gene: an intimate history, Scribner 2017, po polsku: Gen.Ukryta historia, Wydawnictwa Czarne 2017

informacji, odczytać kod i zrozumieć komunikat. nie można przy tym zapomnieć, że to abstrakcja rodzi materię przez komunikat. Tam trzeba szukać granic.

Poznanie

Poznanie, czyli odkrywanie istoty bytu lub zjawiska, to najwyższa wartość ludzkiego życia, powołanie dla starożytnych Greków (żyć to poznawać). Różnorakie czynności i akty poznawcze prowadzą do zdobywania wiedzy o rzeczywistości. Poznanie to rezultaty tego procesu wyrażane w postaci systemu zdań (sądów), twierdzeń jednostkowych i ogólnych, hipotez i teorii. Prawdziwość poznania to zgodność z rzeczywistością według pewnych kryteriów.

Redukcjonizm

Redukcjonizm metodologiczny to wyjaśnianie, polegające na rozłożeniu problemu na mniejsze części, prostsze składowe, które można prościej zbadać/zanalizować i wyjaśnić. David Hilbert stosował matematyczną kompresję. Sądził, że całą matematykę można zamknąć w kilku formalnych twierdzeniach wyrażonych za pomocą skończonego zbioru symboli, wykorzystując skończony zbiór aksjomatów i reguł wnioskowania. Chodziło o rozwiązanie wszystkiego "od dołu do góry".

Rozum

Rozum to zdolność do: –operowania pojęciami abstrakcyjnymi, –analitycznego myślenia i wyciągania wniosków (na podstawie dostępnych danych, informacji, wiedzy, eksperymentów), –uczenia się, –używania zdobytych doświadczeń i posiadanej wiedzy do radzenia sobie w sytuacjach życiowych. Jest przede wszystkim związany z myśleniem logicznym, dedukcyjnym.

Rozumienie to widzenie oczyma duszy.

Semantyka

Przyczynowość jest zaprogramowana przez umysł, trudno jej szukać w nieukierunkowanym systemie fizycznym. **Semantyka** odwołuje się do umysłu istoty "myślącej", czy to na poziomie intuicyjnych odruchów i zachowań, czy też ludzkim umyśle o zdolnościach dedukcji, wnioskowania, kreatywności, a przede wszystkim odnoszącym się do zasad moralnych i dorobku pokoleń.

Niech nikt nie mówi, że nie powiedziałem nic nowego: rozmieszczenie treści jest nowe; kiedy się gra w piłkę, obaj zawodnicy grają tą samą piłką, ale jeden robi to lepiej. To tak samo, jakby mi ktoś powiedział, że posługuję się starymi

słowami. Jak gdyby te same myśli nie tworzyły przez odmienny układ innej treści; tak samo jak te same słowa tworzą przez swój układ inne myśli (Blaise Pascal).

Świadomość

Świadomość to 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). To stan przytomności, czuwania, odbierania bodźców, zdolność do celowej orientacji i odczuwania (przeżywania doznań i stanów emocjonalnych). Stanowi podstawę tworzenia wiedzy i zapamiętywania. Jest zawsze intencjonalna, nakierowana na jakiś przedmiot materialny lub abstrakcyjny i powiązana z odczuciem własnego "ja". Cztery zasadnicze poziomy świadomości to: – percepcja (elementarne mechanizmy wzrokowej świadomości), – sfera pamięci, – sfera wyobraźni, – sfera samo-refleksji, myśli (w schemacie rozpoznanie, interpretacja, treść, wnioski).

Świadomość łączy się z narodzinami języka, samoświadomości, wyraźnego poczucia przeszłości i przyszłości. Jako samoświadomość jest specyficzną gatunkową cechą człowieka, który jest istotą czasową. Z upływu przeżywanego czasu rodzi się ciągłość tematyczna i osobowa, czyli tożsamość.

Świadomość wydaje się nieustannie ciągła, bez przerw, pęknięć czy podziałów, nigdy nie da się jej rozebrać na kawałki; nieustanna zmiana treści (William James, 1890).

Jesteśmy wyłącznie kolekcją chwil, która wpływają jedna w drugą, jak rzeka (Proust).

Wg. Dawida Hume'a umysł to nic innego niż wiązka czy zbiór różnych percepcji, które następują po sobie z niepojętą szybkością i znajdują się w nieustannym stanie płynnym i ruchu tworząc strumień świadomości.

Bardziej formalnie świadomość jest procesem i efektem wielu złożonych oddziaływań neuronalnych (mi.in. kora, podwzgórze). Jeden neuron może mieć do 10 tys. synaps, mózg w całości ma ich ponad 100 bilionów. Możliwości ich modyfikacji przez doświadczenia płynącego czasu (różne procesy myślowe, emocje itd.) są więc praktycznie nieograniczone (wielkości hiperastronomiczne). Można mówić o myśleniu populacyjnym – wielkie populacje neuronów, możliwość różniczkowej zmiany siły powiązań między nimi, formowania się funkcjonalnych grup/konstelacji neuronów; ich wzajemne oddziaływanie kategoryzuje doświadczenia... Nowe doświadczenia kształtują łączliwość różnych sfer mózgu i zmianę ich funkcji (doświadczeniowa selekcja – jak dobór naturalny w teorii ewolucji) itd. itp.

Współczesna nauka nie zna jeszcze sposobu, by zbadać problem świadomości u zwierząt. Dwa intuicyjne wnioski; pierwszy – świadomość zapewne nie jest cechą zero-jedynkową, lecz stopniową; drugi – stopień świadomości danego

gatunku będzie wprost proporcjonalny do jego elastyczności intelektualnej. Im więcej podprogramów, tym bardziej potrzebny jest dyrektor wykonawczy, który trzyma w garści organizację podsystemów ³.

Umysł

Umysł z natury swojej wierzy, a wola kocha; tak iż w braku prawdziwych przedmiotów muszą się czepiać fałszywych (Blaise Pascal)

Umysł – ogół aktywności mózgu – to rdzeń świadomości, który rodzi myśl. Myśl zaś to ruch świadomości, to ”duch, który tańczy” (Alan Moore).

Umysł sięga nieba, koncentrując się na analizie nieskończonych doświadczeń różnego poziomu i odmiennej specyfice, jest zdolny wyciągać kreatywne, niemal niczym nieograniczone wnioski.

Ksenofanes (570- 470 p.n.e.) wierzył w jednego Boga panującego nad światem. Pisał, że Bóg jest niepodobny do śmiertelników formą i myślą, lecz zupełnie bez wysiłku wstrząsa wszystkim mocą *umysłu*.

Epikur (341- 270 p.n.e.) twierdził, że wyjaśniając trzeba trzymać się jak najdalej od mitów, a idąc wiernie za zjawiskami trzeba na ich podstawie wnioskować o rzeczach niewidocznych.

Postuluję Boga po to, by wyjaśnić to, co wyjaśnia nauka (nie jest to Bóg luk, zapchajdziura) – inaczej nie da się tego wyjaśnić (Richard Swinburne).

Nie ma żadnej wątpliwości, że istnieje ogromna różnica pomiędzy *umysłem człowieka* znajdującego się najniższym poziomie rozwoju (kultura, matematyka, uczucia, rozumowanie, ciekawość, marzenia itd.), a umysłem najwyżej uorganizowanego zwierzęcia. Jest to różnica stopnia, a nie rodzaju (materia podobna!); (Karol Darwin).

Zdolności umysłu:

- subitacja, czyli szybka i precyzyjna ocena liczebności niewielkich zbiorów, liczby obiektów; oszacowania mają charakter percepcyjno-przestrzenny, to wzrokowa percepcja wzorców/prototypów, które dominują w naszym doświadczeniu.
- matematyczne i inne – zależą od inkulturacji, która w dużym stopniu dokonuje się poprzez język (syntaktyka, semantyka oraz nośnik, np. fonologia -nauka o systemach dźwiękowych języków), czyli przekaz
- wrodzone oznaczają gotowość na przyjęcie określonej sprawności.

Model umysłu – umysł ucieleśniony (George Lakoff, Rafael Nunez) to system poznawczy człowieka, który jest kształtowany przez osobiste doświadczenia w

³David Eagleman, *Mózg incognito. Wojna domowa w twojej głowie*. Zysk i S-ka, 2021

kontakcie ze środowiskiem (światem); podstawowe struktury pojęć takiego systemu (język kognitywny będący efektem procesów poznawczych jest jak żywy organizm) odnoszą się do relacji przestrzennych i są ściśle związane z możliwościami motorycznymi (cielesnymi); w interakcji ze środowiskiem umysł tworzy konkretne pojęcia (tego języka), przy czym dzięki złożonym mechanizmom poznawczym tworzy następnie, na podstawie konkretnych pojęć, pojęcia abstrakcyjne (pojęcia matematyczne kształtowane były na podstawie pojęć potocznych).

Umysł ucieleśniony jest w dużym stopniu umysłem nieświadomym. Wiele procesów nieświadomych jest zasadniczych dla działania umysłu, m.in. uczenie się i działanie w sytuacjach społecznych, komunikacja, postępowanie zgodnie z normami moralnymi, poznawanie języka itd.

Poziom nieświadomy umysłu obejmuje schematy wyobrażeniowe (język-rozumowanie-percepcja świata), system aspektualny (logiczne wnioskowanie działań motorycznych) i metafory pojęciowe!

Mózg charakteryzuje współzawodnictwo specjalistów, z których każdy jest przekonany, że to właśnie on zna właściwy sposób rozwiązania danego problemu. Zupełnie jak w dobrym dramacie: u podstaw działania mózgu leży konflikt (David Eagleman, neurobiolog ze Stanford University).

Wiedza ludzka

Rozum ma swój porządek płynący z zasad i dowodów. Matematyka zachowuje go, ale jest bezużyteczna w swojej głębi. Żadna **wiedza ludzka** nie może zachować porządku. Serce bowiem ma inny porządek - nie można udowodniać, że należy nam się miłość wyszczególniając kolejne powody do miłości; to byłoby śmieszne. Porządek miłości polega głównie na dygresjach w każdym punkcie odnosząc się do celu, tak by ukazywać go ciągle (Blaise Pascal).

Wyjaśnianie

Wyjaśnianie to zadanie myślowe polegające na wskazaniu racji (natury i funkcji) stojących za określonym problemem/stwierdzeniem/zjawiskiem (tj. przedmiotem wyjaśnienia).

Prowadzi do niego to ocena czy interpretacja obserwowanej rzeczywistości (sensu czegoś), określenie jej przyczyny i sugerowanie skutków. To wykorzystanie, przetworzenie informacji na bazie doświadczeń własnych, wiedzy, intuicji. Wyjaśnianie to przypisywanie znaczeń.

Wyjaśnianie rodzi nigdy nie zrealizowane zapośredniczenia między człowiekiem a światem (rozumiemy coś jako to coś); (Hans-Georg Gadamer, filozof).

Wyjaśnianie stanowi zemstę intelektu na sztuce (nazwać niewyraźne?) – Susan Sontag (pisarka).

Wyjaśniać, to zastępować jakiś wyraz innym na określony sposób (Ludwig Wittgenstein, filozof).

Spis treści

Wprowadzenie	iii
Oznaczenia	v
Kompendium	vii
Spis treści	xvii
1 Fundamentalna teoria informacji	1
1.1 Wprowadzenie	5
1.1.1 Reprezentacja danych	7
1.1.2 Reprezentacja informacji	7
1.2 Nośniki informacji	14
1.2.1 Wyrażanie informacji	14
1.2.2 Podstawowe przestrzenie opisu sygnałów	14
1.3 Opis informacji	22
1.3.1 Teoria informacji według Shannona	22
1.3.2 Kodowanie, czyli usuwanie nadmiarowości	30
1.3.3 Semantyczna teoria informacji	52
1.3.4 Indeksowanie, czyli znakowanie treści	56
2 Rozszerzone koncepcje informacji	83
2.1 Poprawianie Shannona, czyli semantyczna teoria informacji	84
2.2 Kompresja z selekcją informacji	88
3 Sygnały rozszerzone	91
3.1 Wprowadzenie	91
3.2 Reprezentowanie modeli sygnałów	93
3.3 Modele sygnałów	94
3.4 Liniowy problem odwrotny	96
3.5 Aproksymacja informacji z sygnału	98

3.5.1	Podstawowe schematy aproksymacji sygnałów	98
3.6	Reprezentacja rzadka	102
3.7	Strategie	105
3.8	Oszczędne próbkowanie – <i>compressive sampling</i>	108
3.8.1	Próbkowanie z selekcją informacji	109
4	Wybór klasycznych metod obrazowych	111
4.0.2	Stochastyczne modele obrazów	113
4.0.3	Geometryczne modele obrazów	123
4.0.4	Obiektowe modele obrazów	123
4.0.5	Semantyczne modele obrazów	123
4.0.6	Wieloskalowe modele treści obrazów	124
4.1	Komputerowe przetwarzanie obrazów	126
4.1.1	Operacje wykonywane na obrazach	129
	Bibliografia	147

Rozdział 1

Fundamentalna teoria informacji

Informacja syntaktyczna definiowana jest na najniższym poziomie rozumienia tego tak ważnego, a jednocześnie złożonego i trudnego w definiowaniu pojęcia. Odnosi się do składni strumienia symboli nad określonym alfabetem. Można też myśleć o tym inaczej, jak o języku alfabetem, słowami, zdaniami, strukturą, gramatyką, idiomami i metaforami itd. Rozkład wystąpień poszczególnych symboli definiuje wtedy informację, która w żadnym stopniu nie odnosi się do treści, znaczeń, przekazu wiedzy itp.

Nad nią jest informacja semantyczna, powstała z treści zapisów czy ciągu symboli, słów, obiektów, prymitywów, fraz itp. Znaczenie takiego przekazu odnosi się do człowieka, ale też do maszyny (np. ciąg komend sterujących), do typowego obserwatora (np. zdjęcia z pięknymi widokami) lub specjalisty (np. skany radiologiczne).

Najwyższy, kluczowy poziom przekazu informacji ma znaczenie pragmatyczne, skuteczne w konkretnych zastosowaniach. Odbiorca dzięki nowemu poznaniu może rozwiązać trudny problem, odkryć nowe zjawisko, znaleźć odpowiedź na frapujące wszystkich pytanie. Trafniej postawiona diagnoza, gwiazda nowoodkryta, wyjaśnienie istoty zagrożenia, wskazanie źródła/przyczyny konfliktu.

Zbiór zagadnień użytecznych w konkretyzacji i realizacji kognitywnego przekazu informacji, służących realnej konstrukcji kanału przepływu informacji od źródła do odbiorcy to przede wszystkim:

- elementy teorii sygnałów rozumianych przede wszystkim jako nośnik czy też forma urealnienia przekazu informacji w naturalnych warunkach akwizycji danych, transmisji kodowanej reprezentacji i kształtowania jej odczytu (najlepsza jakość, czytelność, zrozumienie, wyjaśnienie, poznanie);
- podstawy teorii informacji formułowane na trzech wspomnianych poziomach celem uzyskania synergii poznawczej służącej skutecznemu rozwiąza-

niu realnych problemów, kształtowaniu nowej wiedzy, odkrywaniu prawdy o człowieku i wszechświecie;

- wybrane obszary teorii aproksymacji, inteligencji obliczeniowej, wyszukiwania i opisu informacji, elementy matematyki stosowanej, fizyki, biologii i chemii itp.; także przetwarzania i analizy sygnałów; logiki, psychologii itd.
- modele odbiorcy, zdolności percepcji treści, rozumienia, poznania, np. ludzkiego systemu widzenia (*Human Visual System*), schematów decyzyjnych, miar i form obiektywizacji ocen, zachowań i decyzji, zasad rozpoznawania i interpretacji treści, reguł podejmowania decyzji itd.;
- metody inteligencji obliczeniowej, przede wszystkim w zakresie wyszukiwania, ekstrakcji czy rozpoznawania informacji, a przede wszystkim jej interpretacji.

Informacja służy odbiorcy w realizacji określonego celu. Przekaz informacji za pomocą strumienia danych (lub też inaczej – sygnału) dokonuje się w kontekście określonych znaczeń, języka, reguł, wiedzy, różnych form wyrażania treści. Mówimy o funkcji semantycznej określającej sposób rozumienia danych. To z niej wynika wartość przekazu informacji dla odbiorcy, czyli jej użyteczność. Precyzyjnie definiując i odczytując warstwę semantyczną przekazu, ustalając referencyjny zakres znaczeń, śledząc zmiany semantyki przy selektywnej lub zakłóconej komunikacji danych, ustalając właściwy porządek przesyłanej treści wpływamy na użyteczność przekazu informacji. Znając cele odbiorcy, ekstrahując informację w sposób klarowny i zwarty, zwiększamy jednoznaczność tego przekazu spodziewając się większych korzyści pragmatycznych. Istotna jest wartość przesyłanej informacji mierzona skalą sukcesu konkretnych zastosowań.

Podstawowy zbiór komponentów kształtujących przekaz informacji jest więc następujący:

- a) sygnał, sekwencja, strumień danych – wynikające z reprezentacji źródłowej, zależnie od metody obserwacji/pomiaru, ale też właściwości kanału transmisji (form kodowania i kompresji), specyfiki urządzeń służących odtwarzaniu informacji;
- b) interpretacja danych, czyli opis elementarnych znaczeń numerycznej reprezentacji informacji (określamy zestaw symboli nad alfabetem, kształtujących słowa danego języka, zdania, terminy, wzajemne relacje, struktury itp.),
- c) treść, czyli znaczenie ciągu symboli interpretowanych względem reguł języka, konstrukcje obiektów ze zbiorem specyficznych cech, ich znaczenia, wymowa zespołu obiektów, ich wzajemnej relacji przestrzennej, metrycznej,

odniesień do tła, cech podobieństwa, rozpoznawalności względem specjalistycznej wiedzy dziedzinowej itp.),

- d) znaczenie treści w kontekście użytkowym, wobec odniesień do realnych zadań, celów, doświadczeń i zamiarów odbiorcy.

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 doskonaleniu przekazu 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;
- formalne koncepcje teorii informacji, zarówno w jej probabilistycznej analizie składniowej (teoria Shannona) jak i funkcjonalnej (Kolmogorow); rozszerzenia semantyczne (słownikowe, językowe itp.);
- podstawy teorii aproksymacji, jako poszukiwanie przybliżeń treści istotnych przekazu informacji, wykorzystywanie metod ekstrakcji istoty przekazu, znaczeniowej charakterystyki sygnału w kontekście predefiniowanych słowników kluczowych komponentów, bezpośrednio nawiązujących do warstwy znaczeniowej, odwołującej do określonego modelu odbiorcy i konkretnych, realnych zdań;
- charakterystyki odbiorcy informacji, zdolności percepcji treści, np. ludzkiego systemu widzenia – ang. *human visual system* czy też pracy odbiorcy ze źródłami informacji, rozpoznawanie treści, ich interpretacji cele podejmowania decyzji – metoda ROC (ang. *Receiver Operating Characteristic*);

- metody inteligencji obliczeniowej, głównie w zakresie wyszukiwania, ekstrakcji czy identyfikacji kluczowych wartości informatywnych, wnioskowania i uogólnień interpretacyjno-decyzyjnych.

W tym podstawowym rozdziale przedstawiono podstawy teorii informacji, zarówno w sensie matematycznej koncepcji informacji rozumianej jako poziom niepewności odbiorcy, jak też jej semantycznych i pragmatycznych rozszerzeń. Tzw. matematyczna teoria informacji abstrahuje od specyfiki zastosowań, semantyki, modeli użytkownika (odbiorcy informacji), specjalistycznej wiedzy dziedzinowej. Dzięki temu pozwala konstruować formalne modele źródeł informacji, wszędzie stosowalne, podatne na optymalizację względem kryteriów formalnych, uniwersalne, wygodne przy testowaniu i obiektywnych ocenach. Jednak tak opisany obraz informacji wielokrotnie nie przystaje do realiów, praktycznych zastosowań i ocen źródeł informacji. Semantyczne rozszerzenia modeli stochastycznych umożliwiają selekcję informacji w metodach kompresji, ekstrakcję treści istotnej za pomocą dobranych metod przetwarzania danych/sygnałów. Pozwalają na definiowanie skutecznych deskryptorów treści, opracowanie selektywnych mechanizmów wyszukiwania informacji w rozległych zasobach, archiwach danych, zgodnie z oczekiwaniami użytkownika.

1.1 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)

Definicja 1.1 *Informacja*

Informacją nazywamy to wszystko, co przekazane – okazuje się użyteczne dla odbiorcy, co 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 stosownej formie (reprezentacji), jest podstawową funkcją szeroko rozumianych multimediiów. Zarówno sposób – bezpośredni przekaz (komunikacja) lub pośrednicząca archiwizacja, jak i forma uzupełniających się strumieni danych o charakterze zróżnicowanym w sensie sposobu percepcji przekazywanej treści służy wiarygodnej i użytecznej komunikacji nadawcy z odbiorcą.

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, multimediiów, 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, sztamkowych treści i pseudo-treści, 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.

1.1.1 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.

1.1.2 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 urządzeniach rejestracji obrazów. Rejestracja dźwięku przy typowej częstotliwości 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 dziedzin, dającej reprezentację upakowaną, skalowalną, a nawet naturalnie uporządkowaną w sensie przyjętego kryterium progresji jakości danych (dziedzina falkowa w algorytmie kodowania standardu JPEG2000¹). 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 JPIP²).

Rozumienie danych, czyli treść

Treść przypisana do ciągu, zbioru lub strumienia 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 przy czytaniu treści odgrywa intuicja czy intelekt odbiorcy, nie została opisana 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 rolę 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 [89].

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 zobjektywizowanego opisu, algorytmicznej procedurze ustalania warunków przekazu np. multimedialnego³. 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 ujednoczenie opinii dotyczących wartości przesyłanych danych, bazuje na fakcie, że pozyskanie informa-

³Przekaz 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 [95], 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 licznosci 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 [90].

Proces ustalania dobrze uformowanej, tj. skutecznej w danym zastosowaniu reprezentacji informacji powinien odwoływać się więc zarówno do znaczenia, jak i prawdziwości, czy też 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 wspomaganie 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 składowej 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. Znaczący, 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.

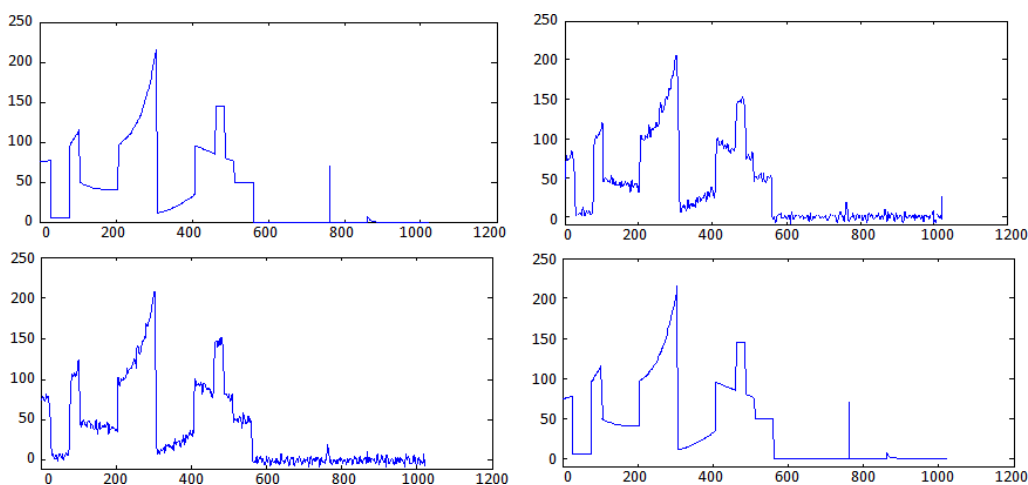
Przyjmując bardziej formalnie pewną złożoność rozważanego problemu, można założyć, że sygnał f składa się z K składników o różnej morfologii $f = \sum_{i=1}^K f_i$, np. obraz składa się z kilku obiektów o różnej morfologii (charakterystyce teksturowej, kształcie itp.), a ponadto istnieje słownik baz 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 $\min \sum_{i=1}^K \|\varphi_i^T f_i\|_0$ 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. 1.6 pokazano przybliżenia obrazów testowych, uzyskane za pomocą upakowanej reprezentacji w kilku różnych bazach – funkcji falkowych, falek geometrycznych - kliników, falek kierunkowych - krzywek oraz funkcji szere-

gu fourierowskiego. Zalety bazy falkowej widać na rys. 1.2, gdzie uzyskano bardzo wierny obraz sygnału za pomocą falkowej reprezentacji o wymiarze stanowiącym zaledwie 15% wymiaru dziedziny źródłowej.



Rysunek 1.1: Efekty opisu dwóch obrazów testowych (barbara i goldhill) za pomocą upakowanej reprezentacji; od lewej kolejno obrazy źródłowe o rozmiarze $512 \times 512 \times 8$ bitów oraz ich przybliżenia z 13,6% współczynników obrazów, uzyskanych za pomocą bazy falkowej, wedgetowej (kliników), curveletowej (krzywek) oraz fourierowskiej.



Rysunek 1.2: 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 gradientcie, w punktach nieciągłości - efekt Gibbsa.

1.2 Nośniki informacji

Nośniki informacji, czyli ogólnie sygnały są bardzo ważnym elementem konstruowanych aplikacji multimedialnych. Sposób ich definiowania, opisu, kształtowania, modulacji treścią są nierozdzielnie związane 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 multimediiów.

1.2.1 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 niedostępnej 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 jednoznacznie 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.

1.2.2 Podstawowe przestrzenie opisu sygnałów

Chcąc przybliżyć 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.

Pokrótkie 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 1.2 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 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{L}$ oraz $a, b \in \mathbb{R}$ (lub ogólniej $a, b \in \mathbb{C}$) spełnione są następujące aksjomaty:

1. przemienność dodawania
 $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$,
2. łączność dodawania
 $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$,
3. rozdzielność dodawania względem mnożenia przez liczbę
 $a(\mathbf{x} + \mathbf{y}) = a\mathbf{y} + a\mathbf{x}$,
4. element neutralny dodawania
 $\forall \mathbf{x} \in \mathcal{L}$: istnieje element $\mathbf{0} \in \mathcal{L}$ taki, że $\mathbf{x} + \mathbf{0} = \mathbf{x}$
5. element przeciwny dodawania
 $\forall \mathbf{x} \in \mathcal{L}$: istnieje element $-\mathbf{x} \in \mathcal{L}$ taki, że $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$,
6. element neutralny mnożenia
 $\forall \mathbf{x} \in \mathcal{L}$: $\mathbf{1} \cdot \mathbf{x} = \mathbf{x}$.

□

Niech liniowa \mathcal{L} będzie przestrzenią ciągów $\mathbf{x} = (x_1, x_2, \dots)$, $\mathbf{y} = (y_1, y_2, \dots)$, $\mathbf{z} = (z_1, z_2, \dots)$ itd. Wtedy dodawanie definiowane jest jako:

$$\mathbf{x} + \mathbf{y} = (x_1, x_2, \dots) + (y_1, y_2, \dots) = (x_1 + y_1, x_2 + y_2, \dots),$$

$$\text{a mnożenie przez liczbę jako: } a\mathbf{x} = a(x_1, x_2, \dots) = (ax_1, ax_2, \dots).$$

Ustalmy podzbiór \mathcal{B} składający się z n elementów liniowej przestrzeni \mathcal{L} , tak że $\mathcal{B} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ i $\forall_{i=1, \dots, n} \mathbf{x}_i \in \mathcal{L}$. Elementy te są **liniowo niezależne**, jeśli $\sum_{i=1}^n a_i \mathbf{x}_i = \mathbf{0}$ jedynie wtedy, gdy $\forall_{i=1, \dots, n} a_i = 0$. 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} :

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

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 $\mathbf{y} \in \mathcal{L}$ mamy więc:

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

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 $\mathcal{L} = \text{span}(\mathcal{B}_{\mathcal{L}})$. 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 1.3 *Przestrzeń unormowana*

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

- a) $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- b) $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
- c) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

□

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

- a) $\|\mathbf{x} - \mathbf{y}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$
 b) $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$
 c) $\|\mathbf{x} - \mathbf{z}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\|$

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 $\{\mathbf{x}_n\}$ jest **ciągiem podstawowym**, inaczej **ciągiem Cauchy'ego**, jeśli odległość $\|\mathbf{x}_n - \mathbf{x}_m\| \rightarrow 0$ dla $n, m \rightarrow \infty$.

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 średnicach⁴ mniejszych lub równych ϵ).

Definicja 1.4 Przestrzeń Banacha

Przestrzeń unormowaną zupełną nazywamy przestrzenią Banacha.

□

W przestrzeni Banacha norma określa metrykę pozwalającą sprawdzić warunek zupełności. W analizie sygnałów ważną rolę odgrywa szczególnie 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 1.5 Przestrzeń unitarna

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

- a) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
 b) $\langle a\mathbf{x}, \mathbf{y} \rangle = a\langle \mathbf{x}, \mathbf{y} \rangle$ dla $a \in \mathbb{C}$
 c) $\langle \mathbf{x}, \mathbf{y} \rangle^* = \langle \mathbf{y}, \mathbf{x} \rangle$
 d) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$
 e) jeśli $\langle \mathbf{x}, \mathbf{x} \rangle = 0$, to $\mathbf{x} = \mathbf{0}$

□

⁴Średnica zbioru to supremum odległości wszystkich par elementów tego zbioru

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

Definicja 1.6 *Przestrzeń Hilberta*

Przestrzeń unitarna zupełna, której norma określona jest przez iloczyn skalarny, nazywana jest przestrzenią Hilberta.

□

Przestrzenie Hilberta są podstawowym pojęciem wykorzystywanym w analizie funkcjonalnej⁵.

Szczególnym przypadkiem przestrzeni Hilberta są: **przestrzeń funkcji całkownych z kwadratem** $L^2(\mathbb{R})$ opisująca sygnały ciągłe o skończonej energii oraz **przestrzeń funkcji sumowalnych z kwadratem** $l^2(\mathbb{Z})$ 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 $x(t) \in L^2(\mathbb{R})$, jeśli $|x(t)|^2$ jest całkowne, czyli:

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

W przestrzeni tej iloczyn skalarny jest zdefiniowany jako

$$\langle x(t), y(t) \rangle_{L^2} = \int_{t \in \mathbb{R}} x(t)y(t)^* dt, \text{ norma } \|x(t)\|_{L^2} = \sqrt{\langle x(t), x(t) \rangle} = \sqrt{\int_{t \in \mathbb{R}} |x(t)|^2 dt},$$

$$\text{a metryka } \|x(t) - y(t)\|_{L^2} = \sqrt{\int_{t \in \mathbb{R}} |x(t) - y(t)|^2 dt}.$$

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

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

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

Bazy przestrzeni Hilberta

Ważnym przypadkiem wzajemnej relacji dwóch wektorów jest ich zerowy iloczyn skalarny. Mówimy, że wektor \mathbf{x} jest **ortogonalny** do wektora \mathbf{y} (co zapisujemy jako $\mathbf{x} \perp \mathbf{y}$), jeśli zachodzi warunek: $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Jeśli dodatkowo długość każdego z ortogonalnych wektorów jest równa 1, czyli $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, wówczas wektory te nazywamy *ortonormalnymi*. Zbiór dowolnej liczby wektorów $\{\mathbf{x}_i | i = 1, 2, \dots\}$ 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

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

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

Rozszerzając pojęcie ortogonalności (i analogicznie ortonormalności), \mathbf{y} jest ortogonalny do podprzestrzeni unitarnej $\mathcal{P} = \{\mathbf{x}_i | i = 1, \dots, n\}$, czyli $\mathbf{y} \perp \mathcal{P}$, jeśli $\forall_{i=1, \dots, n} \mathbf{y} \perp \mathbf{x}_i$. 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 $\{x_i(t) = x_i\}_{i=1,2,\dots}$ 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 $\{x_i(t) = x_i\}_{i=1,2,\dots,n}$ 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 nieskończeniowym wymiarowej \mathcal{H} .

Twierdzenie 1.1 *O zbieżności rzutów ortogonalnych*

Jeśli zbiór funkcji ortonormalnych $\{x_i(t) = x_i\}_{i=1,2,\dots}$ stanowi bazę przestrzeni Hilberta \mathcal{H} , to dla każdego sygnału $y(t) = y \in \mathcal{H}$ ciąg $(\tilde{y}_n)_{n \rightarrow \infty}$ 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 $\{x_i(t) = x_i\}_{i=1,2,\dots,n}$ 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) - nieznannej, poszukiwanej, docelowej funkcji (*target function*) maskowanej w analizowanym sygnale.

⁶Symbol $\delta_{i,j}$ 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 $t \in (-\infty, \infty)$, całkownych 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:

$$\tilde{y}(t) = a_0 + \sum_{i=1}^{\infty} \hat{a}_i \cos(i\omega_0 t) + \sum_{i=1}^{\infty} \hat{a}_i \sin(i\omega_0 t) = a_0 + \sum_{i=1}^{\infty} a_i \cos(i\omega_0 t + \vartheta_i) \quad (1.1)$$

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

1. $y(t)$ jest bezwzględnie całkowna, czyli $\int_{|T|} |y(t)| dt < \infty$;
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 $x_i = e^{j i \omega_0 t}$ (przyjęto oznaczenie $j = \sqrt{-1}$) z częstotliwością podstawową (pierwszej harmonicznej) równą $\omega_0 = 2\pi/T$, wygląda następująco:

$$y(t) = \sum_{i=-\infty}^{\infty} b_i e^{j i \omega_0 t} \quad (1.2)$$

gdzie współczynniki $b_i = 1/T \int_{-T/2}^{T/2} y(t) e^{-j i \omega_0 t} dt$, przy czym w odniesieniu do współczynników szeregu trygonometrycznego (1.1) jest następująca: $b_0 = a_0$, $b_i = (\hat{a}_i - \hat{a}_i)/2$, a $b_{-i} = b_i^*$ dla sygnałów rzeczywistych.

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

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

$$y_n = \sum_{i=0}^{N-1} a_i e^{j2\pi i/Nn} \quad (1.3)$$

Reprezentacja sygnału w bazach fourierowskich według równań (1.2) lub (1.3), 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 bazy 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

$$x_{i,k} = g(t - k\tau) e^{j\omega_0(t-k\tau)} \quad (1.4)$$

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 przestrzeń 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.

1.3 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 znaczenie 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).

1.3.1 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 $\Sigma(W)$ u nadawcy N zgodną z wartością semantyczną u odbiorcy O :

$$\Sigma_N(W) \simeq \Sigma_O(W) \quad (1.5)$$

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 $A_S = \{a, b, c\}$ generuje ciąg informacji: $s(A_S) = (a, a, c, a, b, b, b, c, c, a, c, b, \dots)$, 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 $\mathbf{s} = (s_1, s_2, \dots)$ 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, \dots) traktowana jako proces stochastyczny (dokładniej łańcuch). Właściwości źródła określone są wtedy przez parametry procesu stochastycznego (prawdopodobieństwa łączne, charakterystyka stacjonarności itd.). Stacjonarność źródła w naszych

⁷Łańcuchem stochastycznym nazwiemy proces stochastyczny z argumentem ziarnistym

rozważaniach jest rozumiana w kontekście procesu, którego realizacją jest dana 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 $T(s_1, s_2, s_3, \dots) = (s_2, s_3, s_4, \dots)$. 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 [94].

Warto przypomnieć, że w źródle będącym realizacją procesu ergodycznego każda generowana sekwencja symboli ma te same właściwości statystyczne [95]). 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) $\{s_i\}_{i=1,2,\dots}$, 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 $A_S = \{a_1, a_2, \dots, a_n\}$, oraz zbiór wartości prawdopodobieństw występowania poszczególnych symboli alfabetu: $P_S = \{p_1, p_2, \dots, p_n\}$, gdzie $\Pr(s = a_i) = P(a_i) = p_i$, $p_i \geq 0$ 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: $A_S^N = \underbrace{A_S \times A_S \times \dots \times A_S}_N$, czyli $A_S^N = \{(a_{i_1}, a_{i_2}, \dots, a_{i_N}) :$

$\forall_{j \in \{1, \dots, N\}} a_{i_j} \in A_S\} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_{n^N})$. Prawdopodobieństwo i -tego elementu 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 $\mathbf{s}^t = (s_1, s_2, \dots, s_t)$, 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 $A_S = \{a_1, a_2, \dots, a_n\}$,
- zbiór kontekstów C dla źródła S postaci $A_C^S = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k\}$,
- prawdopodobieństwa warunkowe $P(a_i | \mathbf{b}_j)$ dla $i = 1, 2, \dots, n$ oraz $j = 1, 2, \dots, k$, często wyznaczone metodą częstościową z zależności

$$P(a_i | \mathbf{b}_j) = \frac{N(a_i, \mathbf{b}_j)}{N(\mathbf{b}_j)} \quad (1.6)$$

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 pewnych 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 $\mathbf{s}^t = (s_1, s_2, \dots, s_l, \dots, s_t)$, gdzie $s_l \in A_S$. Sekwencja kontekstów wystąpienia tych symboli jest określona przez funkcję $f(\cdot)$ oraz A_C^S i przyjmuje postać: $\mathbf{c}^t = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_l, \dots, \mathbf{c}_t)$, gdzie $\mathbf{c}_l = f(s_1, s_2, \dots, s_{l-1}) \in A_C^S$ dla $l = 2, \dots, t$ (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 (1.6).

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 : $\mathbf{c}_l = f(s_1, s_2, \dots, s_{l-1}) = s_{l-1}$ dla $l = 2, \dots, t$ oraz $\mathbf{c}_1 = a_r \in A_S$ dla pewnego $1 \leq r \leq n$. 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).

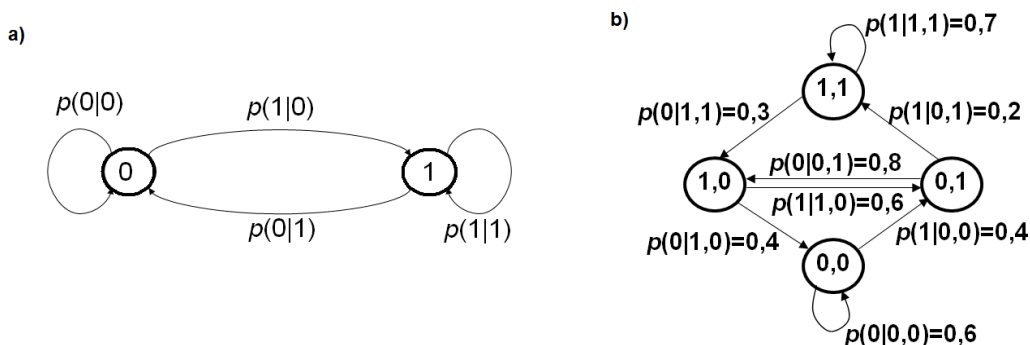
Model źródła Markowa Źródło Markowa rzędu m jest źródłem, w którym kontekst $C^{(m)}$ wystąpienia kolejnych symboli s_l generowanych przez źródło S stanowi skończona liczba m poprzednich symboli $\mathbf{c}_l^{(m)} = (s_{l-1}, s_{l-2}, \dots, s_{l-m})$, czyli dla dowolnych wartości l oraz $t \geq m$ $P(s_l | s_{l-1}, s_{l-2}, \dots, s_{l-t}) = P(s_l | \mathbf{c}_l^{(m)})$. 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):

$$P(a_i | \mathbf{b}_j) = P(a_i | a_{j_1}, a_{j_2}, \dots, a_{j_m}) \quad (1.7)$$

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

Często źródło Markowa analizowane jest za pomocą diagramu stanów, jako znajdujące się w pewnym stanie, zależnym od skończonej liczby występujących

poprzednio symboli - zobacz rys. 1.3. Dla źródła Markowa pierwszego rzędu jest n takich stanów, dla źródła rzędu m mamy n^m stanów.



Rysunek 1.3: Diagramy stanów prostych modeli Markowa z alfabetem binarnym: a) ogólny model rzędu 1 z przejściami między poszczególnymi stanami, b) model rzędu 2 z przykładowymi wartościami prawdopodobieństw warunkowych. Stany opisane są wszystkimi możliwymi kombinacjami kontekstów, zaś odpowiednie przejścia między 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 $i = 1, 2, \dots$, dostarczanych ze źródła informacji S o

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

$$H(S) = \lim_{m \rightarrow \infty} \frac{1}{m} I_m \quad (1.8)$$

gdzie

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

oraz (s_1, s_2, \dots, 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 (1.8) 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 (1.8) 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 werniejszy 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 (1.8). Entropia modelu źródła bez pamięci, uzyskana przez uśrednienie ilości informacji własnej po wszystkich symbolach alfabetu źródła wynosi:

$$H(S_{\text{DMS}}) = - \sum_{i=1}^n P(a_i) \log_2 P(a_i) \quad (1.9)$$

gdzie n oznacza liczbę symboli a_i w alfabecie. Dla $P(a_i) = 0$ wartość $0 \cdot \log_2 1/0 \equiv 0$, gdyż $\lim_{\phi \rightarrow 0^+} \phi \log_2 1/\phi = 0$. 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 (1.9) 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:

$$H(C, S) = H(S|C) + H(C) \quad (1.10)$$

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 $(a_{j_1}, a_{j_2}, \dots, a_{j_m})$ jako:

$$H(S|a_{j_1}, a_{j_2}, \dots, a_{j_m}) = - \sum_{i=1}^n P(a_i|a_{j_1}, \dots, a_{j_m}) \log_2 P(a_i|a_{j_1}, \dots, a_{j_m}) \quad (1.11)$$

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

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

$$H(S|C^{(m)}) = \sum_{A_{C^{(m)}}^S} P(a_{j_1}, \dots, a_{j_m}) H(S|a_{j_1}, \dots, a_{j_m}) \quad (1.12)$$

czyli

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

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 (1.8)). Zależność pomiędzy postaciami entropii związanymi z przedstawionymi modelami źródeł informacji, opisującymi z większym lub mniejszym przybliżeniem informację zawartą w konkretnym zbiorze danych, jest następująca:

$$H(S) \leq H(S|C^{(m)}) \leq H(S_{\text{DMS}}) \quad (1.13)$$

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. Wiarygodnie 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.

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

Definicja 1.7 *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 1.8 *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 informacji⁸ 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.

⁸Inaczej 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 najmłodszyimi bitami.

Zgodnie z klasycznym paradygmatem kompresji stratnej dane wejściowe transformuje się w nową przestrzeń pośrednią, w której zredukowana 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 skompresowanym 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.

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 archiwizatorów 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 odtwarzania danych źródłowych w szerokim zakresie skal, możliwość kodowania 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 1.3.1). 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 nieodwracalne, 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 zadawalającej wiarygodności modelu probabilistycznego w określonym kontekście lub przy maksymalnym dopasowaniu mechanizmu deterministycznego. Przy selekcji informacji ważne jest porządkowanie, maksymalny 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, doskonałe są poprzez rozszerzanie obszaru przeszukiwań zależności 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 stworzeniu 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.

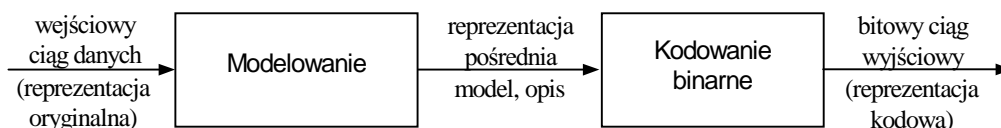
Modelowanie można zrealizować na dwa zasadnicze sposoby:

- 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 wiarygodności 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. 1.4, zaś w przykładzie 1.3.2 opisano proste jego realizacje.



Rysunek 1.4: 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 38). Niech sekwencja danych wejściowych będzie następująca:

$s_{we} = (5, 5, 5, 2, 2, 11, 11, 11, 11, 11, 8)$. Jeśli wyznaczymy model opisujący dane źródłowe seriami jednakowych symboli według schematu: (liczba powtórzeń, symbol), wówczas opis s_{we} źródła za pomocą takiego modelu jest następujący:

$\mathcal{M}_{s_{we}}^{\text{serie}} = ((3, 5), (2, 2), (5, 11), (1, 8))$. 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 $A_S = \{0, \dots, 15\}$, na kolejnych czterech bitach kodu dwójkowego reprezentowany jest symbol tej serii.

Binarna postać sekwencji kodowej \mathbf{s}_{wy} wygląda wówczas następująco: $\mathbf{s}_{wy} = (0100101, 0010010, 1001011, 0001000)$. 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 wystąpień symboli za pomocą wag $w(\cdot)$, równych sumie wystąpień poszczególnych symboli alfabetu. W przypadku rozważanej \mathbf{s}_{wy} mamy więc: $\mathcal{M}_{\mathbf{s}_{we}}^{\text{wagi}} = \{w(5) = 3, w(2) = 2, w(11) = 5, w(8) = 1\}$. Jeśli na podstawie tych wag zróżnicujemy słowa kodowe w sposób następujący: $\zeta(5) = 10, \zeta(2) = 110, \zeta(11) = 0, \zeta(8) = 111$, wówczas otrzymamy ciąg wyjściowy: $\mathbf{s}_{wy} = (10, 10, 10, 110, 110, 0, 0, 0, 0, 0, 111)$ o długości 20 bitów. Druga forma modelowania okazała się skuteczniejsza, chociaż wymaga jeszcze szeregu dookreśleń (w zakresie metody ustalania słów kodowych o zmiennej długości oraz konieczności przekazania w nagłówku pliku skompresowanego parametrów modelu $\mathcal{M}_{\mathbf{s}_{we}}^{\text{wagi}}$).

Przykładem wstępnej dekompozycji danych w celu uproszczenia modelu jest wykorzystanie prostego mechanizmu liczenia różnic pomiędzy wartością kodowaną i wartością bezpośrednio ją poprzedzającą:

$\mathcal{D}^{\text{różnice}} = \{r_i : r_i = s_i - s_{i-1}, i = 1, \dots, 11, s_0 = 0\}$. Dla rozważanej \mathbf{s}_{we} mamy wtedy $\mathcal{D}_{\mathbf{s}_{we}}^{\text{różnice}} = \{5, 0, 0, -3, 0, 9, 0, 0, 0, 0, -3\}$. Dalej stosując modelowanie z wagami ustalamy: $\mathcal{M}_{\mathbf{s}_{we}}^{\text{wagi różnic}} = \{w(5) = 1, w(0) = 7, w(-3) = 2, w(9) = 1\}$. Przypisując tym symbolom zróżnicowane długością słowa kodowe: $\zeta(5) = 110, \zeta(0) = 0, \zeta(-3) = 10, \zeta(9) = 111$, 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 $i \in \{0, \dots, n-1\}$ kolejnych symboli w $A_S = \{a_0, \dots, a_{n-1}\}$). Prosty algorytm zakodowania symboli A_S metodą przyrostową za pomocą kodu dwójkowego wygląda następująco (algorytm 1.1):

Algorytm 1.1 *Kodowanie symboli źródła z wykorzystaniem kodu dwójkowego stałej długości*

1. Pobierz kolejny symbol do zakodowania $s_i = a_j \in A_S = \{a_0, \dots, a_{n-1}\}$ 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 1.2):

Algorytm 1.2 *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 [96].

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} (dobrej 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 $s_i = a_j \in A_S = \{a_1, \dots, a_n\}$. Zwykle stosuje się kody dwójkowe o różnej precyzji lub też, do zapisu liczby powtórzeń, kody zmiennej długości. Alfabet słów kodowych wygląda wtedy następująco:

$$A_{\text{RLE}} = \{(\mathcal{K}_l(k), \mathcal{K}_s(a_j)) : k = 1, 2, \dots, k_{\max}, j = 1, \dots, n\}$$

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 ziar-nistym można opisać za pomocą ciągów liczbowych, np. w systemie dwójkowym. Urządzenia pomiarowe, systemy akwizycji różnego typu danych rejestrują za po-mocą 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 rejestro-wanego procesu, konieczną dokładność rozróżnienia informacji szczegółowych, za-leżności 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 reguły języka polskiego wynika, że statystycznie rzecz biorąc znacznie czę-ściej po literce 't' występuje literka 'a' niż 'x', a po 'z' prawie nigdy nie występuje drugie 'z' czy 'ż' itd. Natomiast w typowych fragmentach tekstu literka 'a' wy-stępuje znacznie częściej niż 'ą' czy 'w'. W obrazach przedstawiających obiekty o rozmiarach większych od pojedynczego piksela wartości sąsiednich pikseli są ze sobą skorelowane⁹, a cały obraz można zazwyczaj scharakteryzować poprzez określenie dominującego koloru.

Zależności te można wyznaczać np. za pomocą statystycznego rozkładu war-tości 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 poszcze-gólnych 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 wielo-wymiarowego, warunkowego z kontekstem przyczynowym. Odpowiada mu model prawdopodobieństw warunkowych źródła z pamięcią.

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

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

Definicja 1.9 *Nadmiarowość źródłowa stochastyczna*

Nadmiarowość stochastyczna sekwencji danych źródła informacji określana jest jako różnica pomiędzy entropią tego źródła i średnią bitową reprezentacji danych.

□

Definicja 1.10 *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.

□

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 wiarygodności 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, \dots, s'_t oraz $s''_1, s''_2, \dots, s''_r$ o wartościach z alfabetu $A_S = \{a_1, \dots, a_n\}$ przyporządkowano bitową sekwencję kodową $z \in Z$ (Z - zbiór wszystkich sekwencji bitowych generowanych przez \mathcal{K}):

$$\mathcal{K}(s'_1, \dots, s'_t) = \mathcal{K}(s''_1, \dots, s''_r) = z \Rightarrow t = r \quad \text{oraz} \quad s'_i = s''_i, \quad i = 1, \dots, t \quad (1.14)$$

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 1.3.2 na stronie 37) 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: $A_{B_4} = \{00, 01, 10, 11\}$, 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 , $s_i \in A_S$, $i = 1, 2, \dots$ odpowiednich binarnych słów kodowych $B_4(s_i)$, takich że $B_4(s_i = a_1) = 00$, $B_4(s_i = a_2) = 01$, $B_4(s_i = a_3) = 10$ i $B_4(s_i = a_4) = 11$, dołączając je do wyjściowej sekwencji bitów (konkatenacja bitów słów kodowych symboli źródła S według porządku ich występowania na wejściu koder). 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 s^t .

Kodowanie według B_k jest przykładem tzw. *kodowania przyrostowego*: $\mathcal{K}(s_1, s_2, \dots, s_t) = \mathcal{K}(s_1)\mathcal{K}(s_2) \cdots \mathcal{K}(s_t)$, gdzie $s_i \in A_S$, a $\mathcal{K}(s_i) \in A_{\mathcal{K}} = \{\varsigma_1, \varsigma_2, \dots, \varsigma_n\}$ ($A_{\mathcal{K}}$ nazwiemy alfabetem kodu \mathcal{K}). Ogólniej kodowanie przyrostowe polega na kolejnym dołączaniu sekwencji bitowych, przypisywanych czytanej sekwencji 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. 48).

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: $\mathcal{K}_1^{-1}(\beta) = a_1 \vee \dots \vee \mathcal{K}_1^{-1}(\beta) = a_2 \vee \mathcal{K}_1^{-1}(\beta) = a_3$.

Innym przykładem kodu niejednoznacznie dekodowalnego przy dwuelementowym alfabetcie źródła jest: $\mathcal{K}_2(a_1) = 0$ i $\mathcal{K}_2(a_2) = 00$, kiedy to sekwencję 00 można zdekodować jako ciąg symboli (a_1, a_1) lub też jako pojedynczy symbol a_2 . Przy większym alfabetcie źródła, np. $A_S = \{a_1, a_2, a_3, a_4\}$ określimy \mathcal{K}_3 jako: $\mathcal{K}_3(a_1) = 0$, $\mathcal{K}_3(a_2) = 1$, $\mathcal{K}_3(a_3) = 01$ i $\mathcal{K}_3(a_4) = 10$. Dekodowanie sekwencji bitowych $\beta_1 = 01$ lub $\beta_2 = 10$ 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: $A_{\mathcal{K}_4} = \{\mathcal{K}_4(a_1), \mathcal{K}_4(a_2), \mathcal{K}_4(a_3)\mathcal{K}_4(a_4)\} = \{\varsigma_1^{(\mathcal{K}_4)}, \varsigma_2^{(\mathcal{K}_4)}, \varsigma_3^{(\mathcal{K}_4)}, \varsigma_4^{(\mathcal{K}_4)}\} = \{0, 01, 001, 0011\}$. Rozpatrzmy krótką sekwencję

kodową $\beta = 001$ utworzoną ze słów kodowych alfabetu $A_{\mathcal{K}_4}$, która może być interpretowana jako połączenie słów 0 i 01 ($\varsigma_1\varsigma_2$) lub też 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 $A_{\mathcal{K}_5} = \{1, 01, 001, 0001\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_6} = \{0, 10, 110, 111\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_7} = \{0, 10, 11, 111\}$ nie jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_8} = \{0, 01, 11\}$ jest jednoznacznie dekodowalny,
- kod $A_{\mathcal{K}_9} = \{001, 1, 100\}$ 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 konkatencją 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 ciągu 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 1.3 *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 $i = 0, \dots, n - 1$ 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 najstarszego). 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

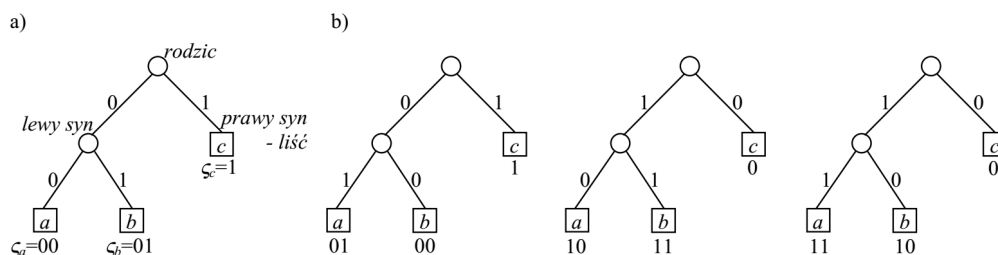
kodu \mathcal{K}_8 postaci: $\beta = 001110 = \varsigma_1\varsigma_2\varsigma_3\varsigma_1$. Zdekodowanie pierwszego symbolu $a_1 = \mathcal{K}_8^{-1}(\varsigma_1)$ trójelementowego alfabetu A_S wymaga odczytania dwóch pierwszych bitów (drugi bit sekwencji kodowej równy 0 wskazuje, że pierwsze 0 to ς_1 , 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. 1.5a, 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.

Przy etykietowaniu drzewa binarnego gałęzie łączące wierzchołek rodzica z dziećmi muszą mieć różne etykiety. Dwa możliwe sposoby etykietowania gałęzi



Rysunek 1.5: Etykietowanie drzew binarnych: a) drzewo o dwóch wierzchołkach wewnętrznych i trzech liściach z symbolami a, b, c , z zaznaczoną relacją rodzic-dzieci oraz ustalonymi słowami kodowymi $\zeta_a, \zeta_b, \zeta_c$; b) trzy inne sposoby etykietowania tego samego drzewa.



Rysunek 1.6: Efekty opisu dwóch obrazów testowych (barbara i goldhill) za pomocą upakowanej reprezentacji; od lewej kolejno obrazy źródłowe o rozmiarze $512 \times 512 \times 8$ bitów oraz ich przybliżenia z 13,6% współczynników obrazów, uzyskanych za pomocą bazy falkowej, wedgetowej (kliników), curveletowej (krzywek) oraz fourierowskiej.

każdego elementarnego poddrzewa (rodzic-dwójka dzieci) to 0 dla połączenia z lewym synem i 1 - z prawym lub odwrotnie (zobacz rys. 1.5a). 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. 1.5.

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 1.5 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.

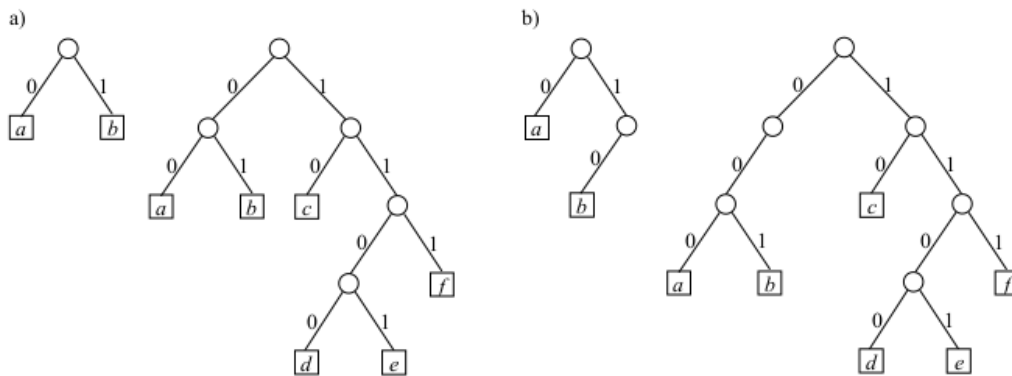
Drzewo lokalnie pełne o zadanej liczbie liści można zbudować, zaczynając np. od struktury z jednym wierzchołkiem, korzeniem, dodając nowe liście zgodnie z poniższą regułą.

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. 1.7.

Aby uzyskać postać drzewa, które utraci cechę lokalnej pełności można zmodyfikować 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 liczbą węzłów wewnętrznych przy niezmienniej 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 $K > 2N - 1$.



Rysunek 1.7: Przykłady drzew: a) lokalnie pełnych; b) nie będących lokalnie pełnymi.

W mniejszym drzewie na rysunku 1.7a liczba wierzchołków $N = 2$, $K = 3$. Wersja tego drzewa po utracie cechy lokalnej pełności z rys. 1.7b 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. 1.7a ma $N = 6$, $K = 11$, jego modyfikacja na rys. 1.7b 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 $\alpha 0$ i $\alpha 1$. 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 $\alpha 0$ i $\alpha 1$ będące przedrostkami tego słowa oraz innego $\varsigma_j \neq \varsigma_i$. Kod drzewa lokalnie pełnego nazywany jest *kodelem przedrostkowym pełnym*.

W większym drzewie z rys. 1.7a 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 $\varsigma_a = 0$ i $\varsigma_b = 10$ mniejszego drzewa z rys. 1.7b świadczą o tym, że nie jest on przedrostkowo pełny: przedrostek właściwy $\alpha = 1$ słowa ς_b nie ma przedłużenia $\alpha 1$, a jedynie przedłużenie $\alpha 0$ 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 *kodech 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 taktie (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 $L_i = |\zeta_i|$, dla których zachodzi następująca zależność:

$$\sum_{i=1}^n 2^{-L_i} \leq 1 \quad (1.15)$$

Odwrotnie, mając dany zbiór n dodatnich liczb całkowitych $\{L_1, \dots, L_n\}$ spełniających (1.15) istnieje jednoznacznie dekodowalny kod symboli o długościach kolejnych słów kodowych równych L_i (formalny dowód można znaleźć np. w [97]).

Nierówność (1.15) 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 długości $L_1 = 1$, czyli spełniona jest nierówność (1.15). 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. 47 stosowanej do dowolnego liścia. Powoduje to następującą zmianę w alfabecie słów kodowych: słowo $\zeta_i = \alpha$ o długości L_i zastępo-

wane jest dwoma słowami postaci $\zeta'_i = \alpha 0$ i $\zeta''_i = \alpha 1$ o długości L_{i+1} . Odpowiedni składnik sumy z wyrażenia (1.15) 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 $\sum_{i=1}^n 2^{-L'_i} < \sum_{i=1}^n 2^{-L_i} \leq 1$, 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 (1.15) jest warunkiem koniecznym jednoznacznej dekodowalności kodu, nie jest jednak warunkiem wystarczającym (dostatecznym). Potwierdzeniem jest np. kod symboli określony słowami $A_{\mathcal{K}} = \{0, 10, 110, 101\}$ (niewielka modyfikacja kodu $A_{\mathcal{K}_6}$) spełniający (1.15). Przykładowe kody $A_{\mathcal{K}_3}$ i $A_{\mathcal{K}_7}$ nie spełniają nierówności (1.15), 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 (1.15) 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 (1.15) jako $\sum_{i=1}^n p_i \cdot 2^{-\epsilon_i} \leq 1$. Jeśli założymy stałą wartość $\epsilon_i = \epsilon$ dla wszystkich słów, wtedy $\epsilon \geq 0$. 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 danych o dowolnie małej długości przy zachowaniu pełnej informacji dostarczonej ze źródła. Okazuje się, że entropia łączna $H(S)$, wyznaczona według równań (1.8) 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 ogranicze-

nia liczby bitów) (*noiseless source coding theorem*) [95, 98], 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, osiągananej 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 koderu prowadzi w kierunku realizacji kodu strumieniowego, uwzględniając przy tym w możliwie wydajny sposób zależności pomiędzy poszczególnymi symbolami kodowanej sekwencji.

Twierdzenie 1.2 *O bezstratnym kodowaniu źródła*

Niech S będzie ergodycznym źródłem z alfabetem o t wygenerowanych elementach i entropii $H(S)$. Ponadto, niech bloki po N ($N \leq t$) 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 \bar{L}_S spełnia równanie:

$$H(S) \leq \bar{L}_S < H(S) + \delta \quad (1.16)$$

Ponadto, nierówność $H(S) \leq \bar{L}_S$ 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 określenia 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 (przybliżyć) 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 $H(S_{DMS})$.

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.

1.3.3 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 [91] 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 wyko-

rzystanie analizy funkcjonalnej, które pozwala w większym stopniu 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 [92, 93] 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 sygnałów/obrazów w celu bardziej zwartego (oszczędnego) 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: $\Sigma_{\mathcal{R}(o_1, o_2, \dots)}(\mathcal{W}) = \Sigma_{\mathcal{R}(\mathcal{O}_{\mathcal{I}})}(\mathcal{W})$, 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 synerгии, całościowa treść zawarta w danych obrazowych:

$$\Sigma_{\mathcal{I}}(\mathcal{O}, \mathcal{R}, \mathcal{W}) = \Sigma(\mathcal{I}, \mathcal{O}_{\mathcal{I}}, \Sigma_{\mathcal{O}_{\mathcal{I}}}, \Sigma_{\mathcal{R}(\mathcal{O}_{\mathcal{I}})}, \mathcal{W}) \quad (1.17)$$

gdzie wektor znaczeń obiektów $\Sigma_{\mathcal{O}_T} = [\Sigma_{o_1}, \Sigma_{o_2}, \dots]$. 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.

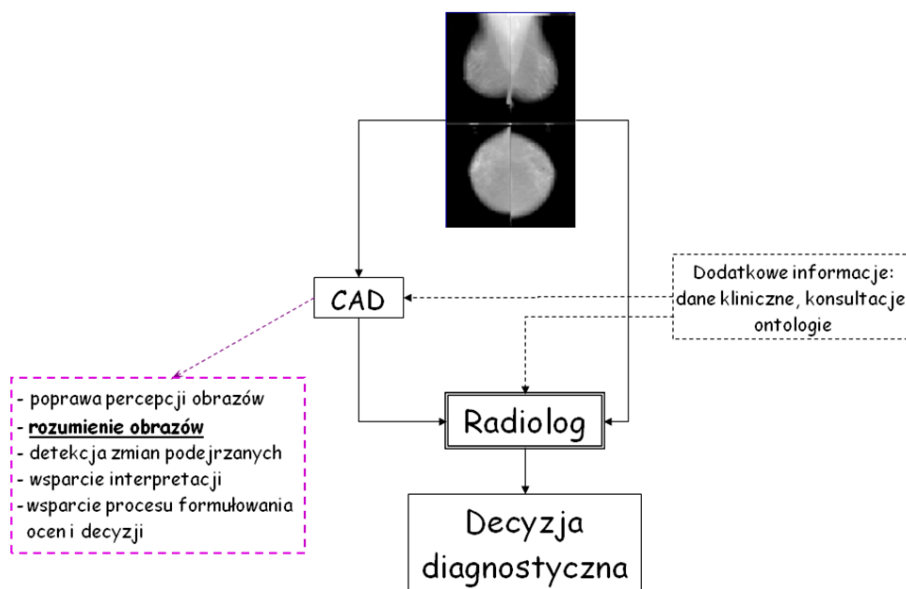
Bezpośrednim owocem rozumienia (operator rozumienia oznaczmy przez \mathcal{U}) treści obrazowej $\Sigma_{\mathcal{I}}(\mathcal{O}, \mathcal{R}, \mathcal{W})$ jest ocena odczytanej treści, czyli wskazanie przedmiotu zainteresowania (SOI, czyli *subject of interests*) lub też stwierdzenie jego braku (brak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

$$\text{SOI} = \mathcal{U}\{\Sigma_{\mathcal{I}}(\mathcal{O}, \mathcal{R}, \mathcal{W})\} \quad (1.18)$$

Skuteczna segmentacja i rozpoznanie obiektów, określenie relacji pomiędzy nimi tak w zakresie porównania ich właściwości za pomocą numerycznych dekskryptorów, 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 różnorodnych 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 [122].

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. 1.8.



Rysunek 1.8: Przykład roli, jaką pełnią metody rozumienia obrazów w realizacji koncepcji komputerowego wspomagania obrazowej diagnostyki medycznej – CAD.

1.3.4 Indeksowanie, czyli znakowanie treści

Gwałtownie rosną cyfrowe zasoby danych multimedialnych, pęcznieją przepelnione archiwa, palące stają się potrzeby sprawnego dostępu do stale rozbudowywanych hurtowni danych. Według najnowszego raportu IDC Digital Universe Study¹⁰ 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,

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

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.

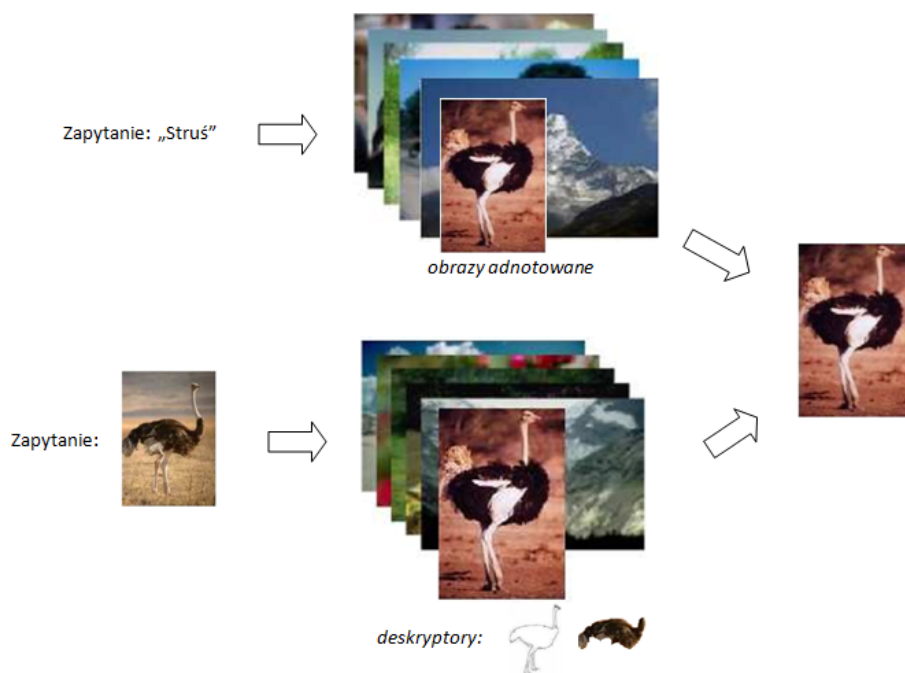
Szczególne znaczenia nabierają systemy selektywnego i możliwie szybkiego wyszukiwania pożądanej treści. Przeszukiwanie baz danych o rozległych, lawinowo rosnących zasobach wymaga obok efektywnych obliczeniowo struktur danych, wygodnych mechanizmów formułowania zapytań i prezentacji odpowiedzi, osadzonych w rzetelnie realizowanych systemach teleinformatycznych, także semantycznych technologii opisu treści. Szczególnie istotnym wymaganiem jest tutaj 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. W przypadku obrazów konsekwencją są dwie metody wyszukiwania: TBIR (*text-based image retrieval*) oraz CBIR (*content-based image retrieval*), przedstawione schematycznie na rys. 1.9.

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



Rysunek 1.9: Zestawienie koncepcji wyszukiwania obrazów z wykorzystaniem koncepcji opisu tekstowego TBIR (u góry) oraz obliczeniowej charakterystyki zawartości CBIR (na podstawie rysunku zaczerpniętego z [123]).

nie 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 multimedialnych 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*)[124].

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 określo-

ny 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 $\{a_k\}_{k=1,\dots,K}$, tj. właściwości istotnych np. ze względu na sposób użytkowania. W przypadku danego obiektu właściwości przyjmują określoną postać tworząc wektor cech obiektu: $\mathcal{C}_a = [a_1(o), \dots, a_K(o)] = [c_1, \dots, c_K]$, 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 [86].

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 - $2h < \text{czas} \leq 3h$, bardzo długi - $\text{czas} > 3h$).

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 szczegółowo: *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 $\rho_a: \mathcal{C}_a \times \mathcal{C}_a \leftarrow [0, 1]$. 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 $\rho_a(c1, c2) \geq t$. Z kolei jeśli ze zbioru obiektów $\{o\}_{1,2,\dots}$ chcemy wybrać najbardziej podobny do o (w sensie określonego a), wtedy

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

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

$$\rho(c1, c2) \triangleq 1 - \frac{\delta(c1, c2)}{\max_c |c|} \quad (1.19)$$

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 (1.19), 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 $c1 = \text{Mama}$ oraz $c2 = \text{Matka}$ 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

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

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

$$\rho(c1, c2) \triangleq 1 - \frac{|r(c1) - r(c2)|}{\max_c r(c) - \min_c r(c)} \quad (1.20)$$

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

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

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

$$\rho(c1, c2) \triangleq 1 - \frac{\min(|r(c1) - r(c2)|, \max_c r(c) - |r(c1) - r(c2)|)}{\max_c r(c) - \min_c r(c)} \quad (1.21)$$

Wtedy podobieństwo *czerwca* do *października* wynosi

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

zaś *stycznia* do *grudnia*

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

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 $\tau_a : \mathcal{C}_a \rightarrow \mathcal{C}_a^\tau$, 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 wyznaczone 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 odległość¹¹ 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*,

¹¹Odległość 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.

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 multimedialnych 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*) [99]. 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 ι_1, ι_2, \dots (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: $\mathcal{L}_{c^\tau} = \{l_{c^\tau}; \iota_1, \iota_2, \dots, \iota_{l_{c^\tau}}\}$.

Liczba obiektów l_{c^τ} na liście zależy od przyjętego kryterium podobieństwa, przy czym zwykle przyjmowana jest progowa definicja zbioru obiektów podobnych jako $\mathcal{O}_{c^\tau} = \{o \in \mathcal{O}: \rho(a(o), c^\tau) \geq \rho_{min}\}$, 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 licznosci, tj. $L_{min} \leq l_{c^\tau} \leq L_{max}$. 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 licznosc 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 multimedialnych \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

$$\rho(c_{query}, c^\tau) \geq \rho_{min} \quad (1.22)$$

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

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

dobór wielkości ρ_{min} powinien być podyktowany względami merytorycznymi, zależnie od rodzaju danych oraz celów wyszukiwania, przy czym możliwy 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}_{cquery} .

W realizacji powyższego scenariusza użyteczne są następujące struktury danych realizujące indeks [86]:

- słownikowa *DictionaryOfFeatures* cech reprezentatywnych atrybutu, zwykle w pamięci operacyjnej; w kolejnych pozycjach słownika obok c^T umieszczone są adresy przypisanych im list obiektowych $\iota_{\mathcal{L}_{c^T}}$;
- 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 $\iota_{\mathcal{L}_{c^T}}$ 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 im 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 [125]. 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 pożądanego, 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 – zapytaniem 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 [131], AltaVista Photofinder, Amor, Berkeley Digital Library Project, Blobworld i in. [132]). Pfund i Marchand-Maillet [125] 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 [126, 127, 128, 129].

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*) [133]. 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. 1.10. Na rys. 1.11 pokazano sytuację odwrotną, kiedy to podobne treściowo obrazy (gatunek tygrysa) są do siebie wizualnie mało podobne lub wręcz niepodobne na poziomie podstawowych cech obrazowych (kolorystyka, tekstury, nawet kształt).



Rysunek 1.10: Luka semantyczna procedur wyszukiwania – przykład zaczerpnięty z [134]; dwa obrazy pomimo dzielących je, oczywistych różnic treściowych zostały mylnie określone przez CBIR jako podobne; podobieństwo koloru i do pewnego stopnia kształtu oraz rozmiar wielu drobnych obiektów nie przekłada się w tym przypadku na wspólną semantykę.

Ważnym powodem luki semantycznej jest **polisemia**, czyli wieloznaczność wyrazu treści obrazowej. Na rys. 1.12 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 [135].

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 1.11: Ograniczenia skuteczności CBIR powodowane trudnym do opisu numerycznym podobieństwem prezentowanych obrazów – przykład zaczerpnięty z [134]; dwa podobne znaczeniowo obrazy mają wyraźnie odmienne wizualne cechy podstawowe.



Rysunek 1.12: Wieloznaczność treści obrazów – przykład zaczerpnięty z [134]; 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 górską, krajobraz), obraz dość jednoznaczny (kwiaty).

rodzaj badania, np. badanie echokardiografii, czyli USG serca. We wspomaganie 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. 1.13 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 pożądaných atrybutów, które pozwolą właściwie określić podobieństwo treści.

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. 1.14 przedstawiono efekt ekstrakcji cech ukrytego symptomu choroby udaru niedokrwiennego (symptom uwi-



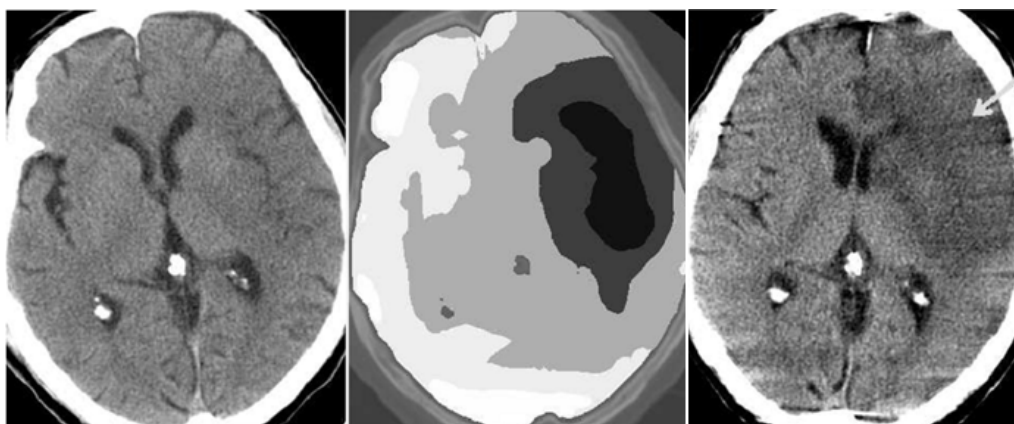
Rysunek 1.13: Wieloznaczność obrazów medycznych – obrazy podobne w sensie modalności i rodzaju zobrazowania (rentgen płuc), jednak różne w interpretacji diagnostycznej, po lewej obraz bez patologii, po prawej z guzem nowotworowym (zaczepnięty z bazy JSRT - *Japanese Society of Radiological Technology database* - utworzonej przez Japońskie Towarzystwo Radiologiczne w 1998: http://www.jsrt.or.jp/web_data/english03.php).

doczniono 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 metodą CBIR pozwala zwiększyć szansę trafnej diagnozy [136].

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 tyl-



Rysunek 1.14: 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 [136]; 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.

ko 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 dobranych 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 [133], adaptacyjne – np. poprzez automatyczną segmentację obiektów zainteresowania [139] lub też interaktywnie, poprzez wybór użytkownika [137, 138].

Opis koloru Charakterystyka kolorów treści obrazowej związana jest w pierwszej kolejności z wyborem przestrzeni barw. Choć obrazu 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. W szczególności korzystnej reprezentacji kolorów jedna składowa opisuje poziom jasności (sygnał luminancji), a dwie pozostałe dotyczą cech koloru (chrominancja). 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 [141]) i inne. Różnice barw liczone w tych przestrzeniach lepiej wyrażają obserwowane zróżnicowanie barw, w sposób bardziej niezależny od uwarunkowań procesu akwizycji [142, 143, 144, 145]. Możliwa jest optymalizacja przestrzeni kolorów wiernie odzwierciedlającej warunki oświetlenia czy też niezmienniczej względem cieni [146, 147]. 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 istnieją wiele definicji określających poszczególne przestrzenie barw, np. CIE RGB, Adobe RGB, NTSC RGB, itd., przy czym szczególną pozycję zajmują normy międzynarodowego komitetu CIE¹².

Przykładowo, wzajemne konwersje wybranych przestrzeni barw opisują proste zależności:

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

$$\begin{aligned} Y &= 0,2989 \cdot R + 0,5866 \cdot G + 0,1145 \cdot B \\ Cr &= 0,5 \cdot R - 0,4183 \cdot G - 0,0816 \cdot B \\ Cb &= -0,1687 \cdot R - 0,3312 \cdot G + 0,5 \cdot B \end{aligned} \quad (1.23)$$

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

$$\begin{aligned} Y &= 0,2974 \cdot R + 0,6273 \cdot G + 0,0753 \cdot B \\ X &= 0,5767 \cdot R + 0,1856 \cdot G + 0,1882 \cdot B \\ Z &= 0,0270 \cdot R + 0,0707 \cdot G + 0,9911 \cdot B \end{aligned} \quad (1.24)$$

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 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 [149]. 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 [150]).

¹²International Commission on Illumination (CIE).

Rozważmy histogram 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: $f(k) \in A_{\mathbf{f}} = \{a_0, \dots, a_{M-1}\}$. 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

$$h(m) = \#\{k | f(k) = a_m\} \quad (1.25)$$

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 reprezentujących 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 $Q_f : \mathbb{Z} \rightarrow \mathbb{Z}$ przekształcającego $Q_f(f = a(m)) = (\tilde{f} = b(n))$ z alfabetem kolorów reprezentatywnych $\tilde{f}(k) \in A_{\tilde{\mathbf{f}}} = \{b_0, \dots, b_n, \dots, b_{N-1}\}$, $N \ll M$ oraz rozkładem kolorów reprezentatywnych

$$\tilde{h}(n) = \#\{k | \tilde{f}(k) = b_n\} \quad (1.26)$$

Tak uproszczony histogram spróbujemy przedstawić w konwencji histogramu rozmytego. Po normalizacji skwantowanego histogramu $p(n) = \tilde{h}(n)/H_{\tilde{\mathbf{f}}}$, gdzie $H_{\tilde{\mathbf{f}}} = \sum_{m=0}^{M-1} h(m) = \sum_{n=0}^{N-1} \tilde{h}(n)$, histogram możemy zinterpretować w kategoriach prawdopodobieństwa [151], rozumiejąc $p(n) = p(b_n) = Pr(b_n) = Pr(\tilde{f} = b_n)$ jako

$$p(n) = \sum_k^{H_{\tilde{\mathbf{f}}}} p(n|k)p(k) = \frac{1}{H_{\tilde{\mathbf{f}}}} \sum_{k=1}^{H_{\tilde{\mathbf{f}}}} p(n|k), \quad \forall b_n \in A_{\tilde{\mathbf{f}}} \quad (1.27)$$

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 [150], warunkowe prawdopodobieństwo przynależności piksela (ze względu na jego wartość) do przedziału danego koloru zastępowane jest funkcją przynależności: $\mu : \mathbb{Z} \rightarrow [0, 1]$ określającą relację wartości każdego piksela do wszystkich przedziałów skwantowanych kolorów b_n w sposób mniej jednoznaczny, z możliwością

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 różnicowanie 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 $b_n - b_{n-1} = \text{constant}$. Głównym zamysłem jest bowiem przypisanie każdemu z pikseli percepcyjnego znaczenia (istotności) jego koloru poprzez odniesienie do wartości reprezentatywnych. Spodziewanym efektem jest rozkład kolorów, który odzwierciedla różnicowanie cech percepcji kolorów obrazu, niezależnie od innych czynników (zmian oświetlenia, szum, itd.).

Przez analogię do zapisu histogramu klasycznego (1.27), otrzymujemy

$$r(n) = \sum_k^{H_f} \mu(n, k) p(k) = \frac{1}{H_f} \sum_{k=1}^{H_f} \mu(n, k), \quad \forall b_n \in A_{\tilde{f}} \quad (1.28)$$

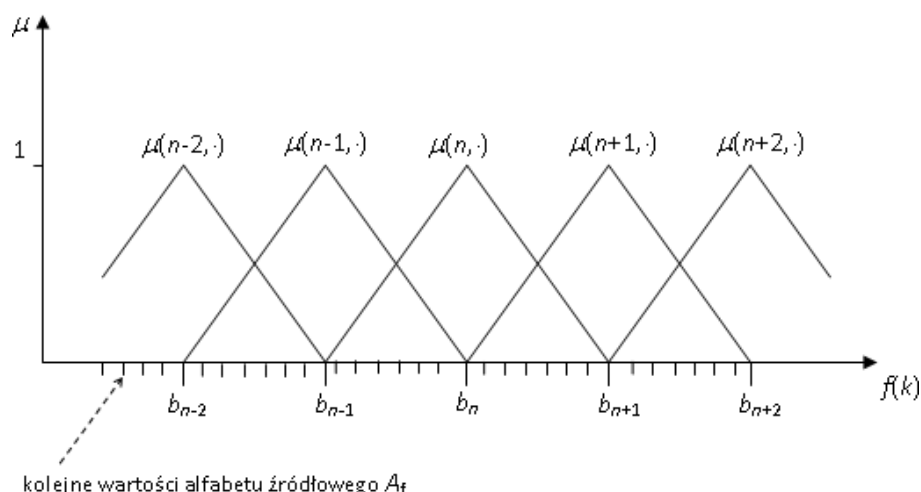
$r(n)$ jest zbiorem rozmytym na przestrzeni wartości pikseli obrazu \mathbf{f} , z funkcją przynależności $\mu(n, k)$ do przedziałów kolorów reprezentatywnych b_n , o następujących cechach:

- 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 ciągłą, monotonicznie rosnącą na przedziale $[b_{n-1}, b_n]$ i malejącą na przedziale $[b_n, b_{n+1}]$, symetryczna względem b_n
- spełniona jest zależność

$$\sum_n \mu(n, k) = 1, \quad \forall k = 1, \dots, H_f \quad (1.29)$$

Na rysunku 1.15 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).

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

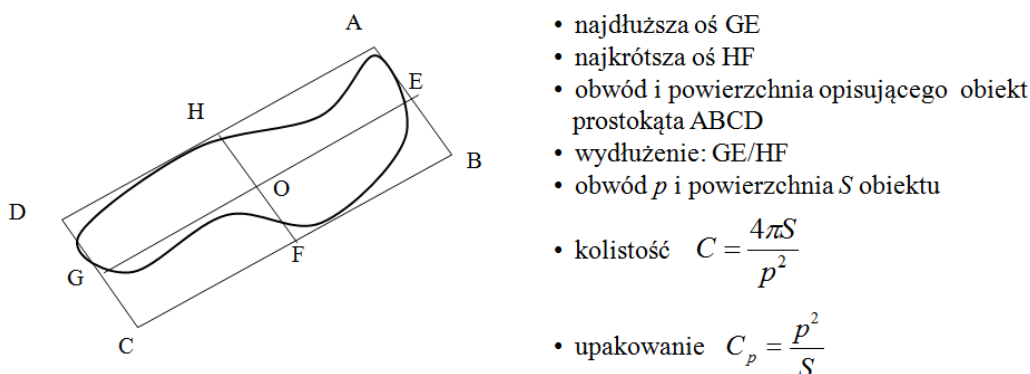


Rysunek 1.15: Ilustracja trójkątnej funkcji przynależności wykorzystywanej przy wyznaczaniu rozmytego histogramu obrazu.

wręcz nierozwiązywalnym [153, 131, 139]. 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. 1.16 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.



Rysunek 1.16: 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 (zróznicowanie 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 pól Markowa, funkcji autokorelacji sygnału czy cechach częstotliwościowych (fourierowskich) (np. w [158]). Często stosowane są też cechy wyznaczone na podstawie macierzy powinowactwa [159, 160, 161], takie jak kontrastowość, zmienność, jednorodność, energia, korelacja.

Niestacjonarne tekstury (zawierające przynajmniej dwa dające się wyróżnić wzory) opisywane są najlepiej za pomocą falek [154, 155] i ich dwuwymiarowych uogólnień (*curvelety*, falki zespolone), innych transformacji typu czas częstotliwość, dwuwymiarowych sygnałów analitycznych, funkcji i kierunkowych filtrów Gabora [142, 156, 157].

Selektywność wyszukiwania

Selektywność wyszukiwania według zadanego scenariusza możemy oceniać, jeżeli znana jest semantyczna relacja równoważności pomiędzy obiektami w kolekcji [86]. 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 wyszukiwarke 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*), przywołanie (*recall*), odniesienie precyzji do przy-

wołania, stopa sukcesu (*success rate*) czy średnia ranga (*average rank*) [162, 86]. 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 oznaczymy zbiór wszystkich testowych, możliwie licznych zapytań $q \in Q$. Jeśli wśród wszystkich $K(q)$ odpowiedzi na każde q znajdzie się dokładnie $T(q)$ odpowiedzi poprawnych (co oznacza również $K(q) - T(q)$ odpowiedzi niepoprawnych), wtedy precyzję obliczamy jako

$$pr = \frac{1}{|Q|} \sum_q \frac{T(q)}{K(q)} \quad (1.30)$$

Wygodniej jest ustalić stały scenariusz wyszukiwania odpowiedzi kładąc stałe $K(q) = K$, niezależnie od zapytania. Często jest to koniecznością przy testowaniu bazy o nieznannej zawartości (nie wiemy, ile obrazów danej treści jest w bazie). Jeśli jednak znana i zróżnicowana jest reprezentacja odmiennych treściowo obiektów bazy, rzetelniesze wyniki można uzyskać ustalając odpowiednio duże $K(q)$. W szczególności, by nie ograniczać liczby możliwych odpowiedzi poprawnych warto przyjąć przynajmniej $K(q) \geq N(q)$, gdzie $N(q)$ jest liczbą wszystkich obiektów bazy semantycznie równoważnych q .

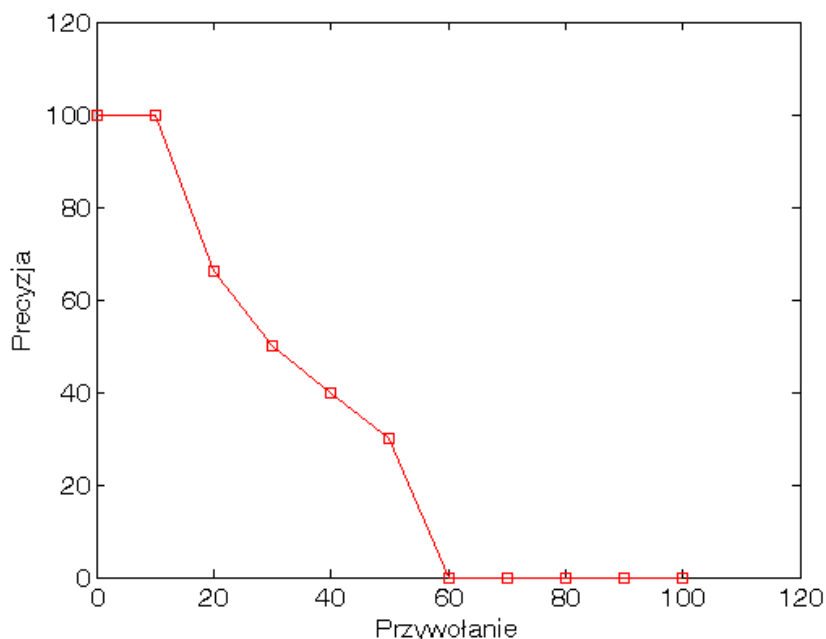
Przywołanie wskazuje, ile z wszystkich potencjalnych odpowiedzi poprawnych na zapytanie q (ich liczba jest równa liczbie wszystkich obiektów semantycznie równoważnych q , które znajdują się w bazie) znalazło się wśród obiektów zwróconych przez wyszukiwarkę. Mamy więc przywołanie zdefiniowane jako

$$rec = \frac{1}{|Q|} \sum_q \frac{T(q)}{N(q)} \quad (1.31)$$

Miara ta zależy więc jeszcze silniej od scenariusza wyszukiwań testowych (tj. sposobu ustalania wartości $K(q)$). Z (1.30) oraz (1.31) wynika własność $pr, rec \in [0, 1]$.

Precyzja odniesiona do wartości przywołania jest częstym miernikiem efektywności procesu indeksowania – zobacz rys. 1.17. 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 grupie zapytań testowych. W takich sytuacjach można zastosować pojedynczą wartość precyzji (dla każdego zapytania), traktując jako uśrednioną dla danego poziomu przywołania, zgodnie w

podjęciem przedstawionym powyżej. Możliwe jest także zaproponowanie innych podejść do wyznaczania precyzji, które mogą być bardziej użyteczne [162], takie jak średnia precyzja dla widzianych dokumentów adekwatnych czy R-precyzja [128]. Średnia precyzja dla widzianych dokumentów adekwatnych polega na wy-



Rysunek 1.17: Przykładowy wykres precyzji w funkcji przywołania. Standardowo precyzja jest wyznaczana dla 11 poziomów przywołania (źródło: [128]).

znaczaniu precyzji poprzez uśrednianie precyzji wyznaczonej po zaobserwowaniu w zbiorze wyników adekwatnego dokumentu. Przykładowo, dla danych z rysunku 1.17 wartość precyzji po pojawieniu się w zbiorze wyników 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

$$sr = \frac{1}{|Q|} \sum_{q \in Q} \sigma(q) \quad (1.32)$$

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

$$ara = \frac{1}{|Q|} \sum_{q \in Q} ra(q) \quad (1.33)$$

gdzie $ra(q)$ jest średnią pozycją wszystkich obiektów równoważnych q znajdujących się w bazie, w odniesieniu do konkretnej realizacji zapytania q . Tak więc

$$ra = \frac{1}{N(q)} \sum_{n=1}^{N(q)} pos(n, q) \quad (1.34)$$

$pos(n, q)$ określa tę pozycję jako

$$pos(n, q) = \begin{cases} k & \text{jeżeli obiekt } n \text{ zajmuje w odpowiedzi pozycję } k \leq K(q) \\ K(q) + 1 & \text{w p. p.} \end{cases}$$

Rozdział 2

Rozszerzone koncepcje informacji

Uzupełnienie użytkowej koncepcji informacji o aspekty semantyczne i pragmatyczne prowadzi do pełniejszego obrazu informacji i jej roli w rozwoju świata na kluczowych dziś kierunkach. Problem jest wieloaspektowy i bardzo różnorodny. Poniżej zamieszczono pewne przyczynki i impresje. Także w wielu przypadkach codziennych zastosowań semantyka przekazywanych reprezentacji czy sygnałów odgrywa na tyle znaczącą rolę ze względu na ich wykorzystanie, że powinna stanowić kluczowy element w procesie optymalizacji algorytmów wstępnego przetwarzania danych pomiarowych, transmisji czy prezentacji informacji. Wymaga to "uzupełnienia" stochastycznej (matematycznej) teorii informacji.

Dwa zasadnicze zadania teorii Shannona ,czyli a) obliczanie ilości informacji dostarczanej przez źródła możliwie wiernie opisujące przesyłane dane, b) opracowanie efektywnych kodów bazujących na właściwych modelach statystycznych źródeł informacji wymagają rozszerzeń znaczeniowych nad alfabetem symboli, uwzględnienia kontekstu pojęciowego oraz natury 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 [91] 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 ułatwiają semantyczną selekcję w algorytmach rozpoznawania. Chodzi tu o wykorzystanie analizy funkcjonalnej, które pozwala w większym uolnić się od niereali-

stycznych 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 [92, 93] 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 sugestie 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).

2.1 Poprawianie Shannona, czyli semantyczna teoria informacji

Matematyczna, formalna definicja informacji zaproponowana przez C.Shannona dała istotny impuls gwałtownego rozwoju informatyki i telekomunikacji, otworzyła drzwi systemom transmisji sygnałów, umożliwiła dobór efektywnych reprezentacji i konstrukcję narzędzi przekazu informacji, zaowocowała optymalizacją złożonych algorytmów kompresji danych. Jakkolwiek przydatna w praktycznych im-

plementacjach, ma jednak istotne ograniczenia w zakresie wartości poznawczych otaczającej nas rzeczywistości. Nawet odpowiednio rozszerzając i komplikując formalną definicję Shannona (np. za pomocą złożonych modeli prawdopodobieństw warunkowych, czy nawet łącznych), nie da się za jej pomocą opisać informatywnej wartości zjawisk realnego świata, ich znaczenia, użyteczności i sensu.

Dużo później David Wolpert z Instytutu Santa Fe sformułował problem przepływu informacji w znacznie szerszej skali i zakresie, w odniesieniu do realiów fizycznego wszechświata¹. Ma to bezpośredni związek z ograniczeniami automatycznej algorytmiki. I tak D.Wolpert i W.Macready wykazali, że żaden algorytm wyszukiwania, w tym algorytmy ewolucyjne, nie jest lepszy od ślepego wyszukiwania ("nic za darmo")². W.Dembski powiązał to z prawem zachowania informacji - pokazał, że nie można rozsądnie oczekiwać, że wyszukiwanie ślepe (losowe) znajdzie określony, złożony cel bez pomocniczych *informacji spoza systemu*³. By przyspieszyć ślepe wyszukiwanie, jakaś reguła (wiedza) musi być dostarczona z zewnątrz. Widać tu podobieństwo do twierdzeń Gödla o niezupełności, gdzie system matematyczny nie może się sprawdzić za pomocą własnych zasobów (zbiór wszystkich twierdzeń arytmetyki nie może być wygenerowany przez żaden system formalny; czyli zbiór zdań dowodzonych formalnie nigdy nie będzie równy ze zbiorem zdań prawdziwych określonej teorii).

Rzucając kostką, można oszacować prawdopodobieństwo każdej z możliwych liczby oczek. Mamy znany, skończony zbiór rozwiązań (zdarzeń elementarnych), potrzeba jedynie czasu i eksperymentalnej pracy. Jednak w przypadku biosfery, technosfery, kultury, ludzkiej myśli nie znamy zbioru możliwych rozwiązań⁴, nie sposób mówić o prawdopodobieństwach zachodzących zdarzeń. Biorąc więc życie człowieka, jak mamy sobie radzić nie znając tego, co może się wydarzyć? To dylemat każdej i każdego, który można rozwiązać za pomocą zdobywanych informacji. Tu nie pomoże żaden formalizm, a zasoby najgłębszej wiedzy tam nie sięgają. Nie sposób matematycznie opisać rozwoju kultury, sztuki, humanizmu (najbardziej do tego pasuje chaos deterministyczny, przy niestety nieznanym warunkach początkowych; albo mechanika kwantowa mówiąca: cóż, rzeczy się zdarzają ...). Alternatywą jest pojęcie z nauki o życiu, czyli gen, odcinek DNA kodujący białka, który jest aperiodycznym kryształem, ale w swej istocie jest jedną z podstawowych form osadzonej w rzeczywistości informacji. Bo informacja to złamane symetrie, które umożliwiają zajście we Wszechświecie specyficznych procesów, zależnie od warunków brzegowych. Ale jak to się dzieje, nie wiadomo - potrzebne są wciąż nowe informacje, by przesunąć granice wiedzy. A potem, gdy

¹https://evolutionnews.org/2015/11/a_mechanical_ph/

²<https://evolutionnews.org/2018/12/trends-in-philosophy-of-science-what-does-semantic-information-mean/>

³W.A. Dembski, *No free lunch: Why specified complexity cannot be purchased without intelligence*. Rowman & Littlefield Publishers, Inc., 2002.

⁴Stuart Kauffman w 'Heretycy, buntownicy, wizjonerzy. 22 podróże z największymi umysłami naszych czasów', Copernicus Center Press, 2019.

już się niemal wie, przychodzi rozczarowanie i potrzebne są nowe informacje. Stuart Kauffman⁵ opowiada: sądzono (realizując projekt poznania ludzkiego genomu), że poznając genom, będziemy wiedzieć, jak powstają organizmy. Jednak nie dowiemy się. Organizmy to nieliniowe, dalekie od stanu równowagi systemy termodynamiczne - dalej nie wiemy w jaki sposób radzą sobie z najprostszymi procesami, nie potrafimy matematycznie opisać tej siedziby życia! Na nic potężna moc obliczeniowa olbrzymich zasobów obliczeniowych . . . A życie trwa, kwitnie . . . bo prawda jest zbyt złożona, bo cała komórka to system, który się organizuje, dynamicznie stwarza nową komórkę . . . Proste formy życia są bogatsze od złożonych modeli, form optymalizacji czy globalnych mocy obliczeniowych . . . Czym wobec tego jest umysł, czy jest kwantowy? Jak zdefiniować umysł względem mózgu, jak formalnie wytłumaczyć, czym jest świadomość, wolna wola . . .

W ten sposób informacja sięga nieznannej, nieskończonej głębiny pozwalając aproksymować w jakimś stopniu skuteczne rozwiązania. Zwiększając poznanie, możemy poruszać się w kierunku nieskończoności zwiększając wymiar świata. Tu pomaga inteligencja podsuwając trafne rozwiązania życiowych (praktycznych) problemów. Informacja temu służy, zwiększa szanse sukcesu, rozwija człowieka i świat, jakby "dotwarzając" brakujące elementy, kontynuując poznawanie świata w nieznanych fragmentach jego nieskończoności.

Idąc dalej w tę stronę, na końcu możemy spotkać Boga. Realny wymiar semantyki staje się wtedy nieograniczony. Polkinghorne⁶ uważa, że wszystkie twierdzenia bazujące na wiedzy bazują w jakiś sposób na osobistych, subiektywnych sądach, a metody poznania (prawdy) nie da się zmechanizować. Nauka potrzebuje więc aktów dobrze umotywowanej wiary. Rozsądne zaufanie jest nierzadko warunkiem koniecznym, który napędza dalszy rozwój, uzasadnia konieczność dopracowania szczegółów/poprawienia niektórych elementów kształtującej się teorii. Budujemy mity, w których prawdziwość wierzymy, jednocześnie wiedząc (biorąc także pod uwagę, dopuszczając), że mogą być fałszywe. Leibniz pytał, dlaczego istnieje raczej coś niż nic - dlaczego nasze formalne coś nie obejmuje wszystkiego? Zawsze jest coś dalej poza fizyką . . . idąc dalej, pojawia się samoświadomość, rodząca świadome pytania o Boga, czyli przyczynę świata.

Skuteczniejsza definicja informacji "z zewnątrz" odwołuje się więc do semantyki, a za jej pośrednictwem do innych wymiarów poznawczych otaczającej rzeczywistości. Do komunikacji pomiędzy podmiotami (nadawca-odbiorca) dodaje się 'znaczenie' przesyłanych danych czy struktur. Źródła tych koncepcji są niewiele młodsze od definicji Shannona. Lata doświadczeń nie tyle przyznały rację intuicji wspomnianych Bar-Hillela i Carnapa, co raczej uczyniły je realnymi w kontekście imponującego rozwoju technologii informatycznych (kolejne rewolucje). Na nowo zdefiniowano więc informację odwołując się do znaczeń kształtujących

⁵tamże

⁶tamże

treść przekazu. Wymaga to odwołania do natury środowiska, w którym powstaje przekaz (fizycznej, biologicznej) oraz do natury odbiorcy (użytkownika informacji), z uwzględnieniem realnego czasu i warunków egzystencji, jego istotnych potrzeb i zainteresowań (problem przydatności: to co jest niechciane, nieprzydatne, obiektywnie nie mające znaczenia-przynajmniej według aktualnej wiedzy- nie kwalifikuje się jako informacja semantyczna). Według założeń semantycznej teorii informacji, przekaz danych, sygnałów może określać taką użyteczność/model użytkownika w sposób obiektywny (lokalizacja pożywienia dla zwierząt według ogólnej wiedzy na ten temat, celem przetrwania) lub subiektywny (znajomość specyficznych preferencji naszych czworonożnych ulubieńców, ale też wcześniejsza negocjacja zamówienia z odbiorcą lub też inteligentne przewidywania potrzeb na podstawie wnioskowania rzetelnych obserwacji). Konieczne jest odwołanie do przyczynowo- skutkowych obserwacji względem upływającego czasu.

Zgodnie z tą definicją, teoria informacji, odwołując się do wiedzy aktualnej lub stanowionej obejmuje każdą realną interakcję w biologii i fizyce -informację zapisaną jest w naszym DNA, w skałach i huraganach, dotyczy reguł przebiegu reakcji chemicznych czy też zasad wiarygodnie opisujących zjawiska fizyczne. Odkrywać zjawiska znaczy zdobywać informację semantyczną służącą głębszemu poznaniu środowiska, systemu fizycznego czy biologicznego. Z drugiej strony, każdy system fizyczny lub biologiczny, który trwa, wykorzystuje informacje semantyczne zapisane w regułach tworzących je zjawisk, struktur, zasad regulujących zachodzące procesy itd.

W dziedzinie biologii odczytanie i zrozumienie informacji semantycznej zapisanej w poznawanej naturze zjawisk może pomóc odpowiedzieć na najbardziej intrygujące pytania - np. jak ewoluowały najwcześniejsze, pierwotne formy życia, jak istniejące organizmy przystosowują się do zmian w środowisku, do nowych zagrożeń itd. Jeśli mówimy o rozwoju, doskonaleniu organizmów, o ich większej sprawności czy lepszej adaptacji, czy to znaczy, że w jakiś sposób docierają do nich semantyczne informacje, ich ilość rośnie w czasie ewolucji, jak to się dzieje, skąd pobierane są te informacje i jaka jest ich natura? Jak organizmy potrafią dotrzeć do tych najbardziej znaczących? Skąd wreszcie bierze się przekaz służący doskonaleniu organizmów, jakie jest źródło informacji zapisanej w organizmach, zjawiskach czy całych systemach? W tym kontekście istotne okazuje się pytanie: czy i jak semantyka zapisu/przekazu informacji naprawdę może być, i na ile, znacząca w odniesieniu sfery nieświadomych bytów? Jak świadome nadanie znaczeń jest w stanie kształtować rozwój nieświadomej rzeczywistości? A może i świadomej, na ile świadomej? Co z wolną wolą?

By sensownie odpowiedzieć na te pytania, konieczne jest rozszerzenie semantycznej teorii poza pojęcie znaczenia przekazu informacji zapisanego w strukturach biologicznych czy zjawiskach fizycznych. Znaczenie bowiem odnosi się do określonej wiedzy, reguł oraz procesów poznawczych. Określić je może funkcja

semantyczna nad określonych alfabetem, zestawem pojęć podstawowych i złożonych, regułami wnioskowania, aż po intuicyjne poznanie. Powtórzmy to, przypisane znaczenia wraz z regułami, z wykorzystaniem pełnego formalizmu nie da nam pełnej prawdy o określonej teorii, zjawisku czy rzeczywistości.

Semantyka idzie więc dalej od automatów, wymaga myślącego podmiotu, inaczej "agenta". Tylko świadome istoty moralne rozpoznają znaczenie, dokonują wyboru, uczą się. Biorą w tym udział określone procesy myślowe, instynkt, wolny czynnik moralny podejmowania decyzji, a nawet kreatywność, dojrzewanie świadomości itp. Semantyka odnosi się więc do umysłu myślącej istoty jest zdolności decydowania, wyboru, interpretacji i rozwoju. Semantyka zjawisk fizycznych czy prymitywnych form biologicznych jest jedynie innym przejawem formalizmu, który nie daje możliwości poznania prawdy.

Uogólniając, osiągnięcia nauki, zdobyta wiedza zasadniczo służą przewidywaniu zdarzeń, zjawisk, rozwoju. Informacja zaś to przede wszystkim wyjaśnianie, to rozwój, postęp poprzez pozyskiwanie nieznanego, odkrywanie tajemnic czy rozjaśnianie ciemności. To skok do przodu, konieczny wobec mozolnej pracy nad tym co już było, wyczerpującej wszystkie możliwe do odkrycia wnioski. To odkrywanie/przekazywanie nowego, to przełom, który dedukcyjnie rozrasta się w nową wiedzę.

2.2 Kompresja z selekcją informacji

Ograniczona skuteczność redukcji długości reprezentacji danych źródłowych, osiągnięta za pomocą metod kodowania, spowodowała rozwój alternatywnych metod tzw. kompresji stratnej. Zakładają one możliwość zmiany postaci danych źródłowych w procesie kompresji/dekompresji. Zmiana ta nawiązuje do podstaw teoretycznych stworzonych w początkowym okresie rozwoju teorii informacji, a dotyczących realiów transmisji sygnałów ze stratami powodowanych uwarunkowaniami ówczesnych mediów transmisyjnych.

Zgodnie z **klasycznym paradygmatem kompresji stratnej** dane wejściowe transformuje się w nową przestrzeń pośrednią, w której zredukowana 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

pozwole ustalić wartości dopuszczalnych stopni kompresji w określonych zastosowaniach.

O możliwościach formułowania nowych paradygmatów kompresji, gdzie znikają granice pomiędzy kompresją stratną i bezstratną, można przeczytać w [163].

Charakterystyka metod kompresji obrazów

Dla obrazów stosuje się szereg metod eliminacji nadmiarowości przestrzennej (zasadniczy typ nadmiarowości - zobacz p.1.3.2) 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 [177]. Poważnym ograniczeniem ich efektywności jest niedoskonałość predykcji powodowana m.in. brakiem możliwości uształtowania kontekstu (sąsiedztwa) predykcji otaczającego kodowany piksel z każdej strony (konieczność zachowania warunku przyczynowości).

Modyfikacją metod predykcyjnych są algorytmy interpolacyjne HINT (*Hierarchical Interpolation*) [178, 179, 180], w których zastosowano kodowanie/dekodowanie kolejnych wersji obrazu o coraz większej 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 zazwyczaj kodowane z wykorzystaniem kodu Huffmana lub arytmetycznego.

Nowsze rozwiązania zawierają także rozbudowaną fazę modelowania statystycznych zależności danych w określonym kontekście (np. metoda CALIC [181]), które bezpośrednio sterują pracą koderów tworzących sekwencję kodową na podstawie statystycznych modeli źródeł informacji (tzw. kodery entropijne). 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 [182], JPEG2000 [183], JBIG [185], JBIG2 [186] oraz AVC z MPEG-4 [184].

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 [188]) 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 [189] i PNG [190]);
- 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 [191]), 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).

Rozdział 3

Sygnaly rozszerzone

Rozszerzono metodykę reprezentowania informacji o przybliżenia informacji w sygnale rzeczywistym, przy uwzględnieniu specyfiki i realnych ograniczeń systemów akwizycji, estymacji modelu funkcji aproksymowanej (docelowej - ang. *target function*) na bazie możliwych form odwołań do semantyki rejestrowanych danych oraz pojęcia użyteczności sygnału względem wymagań odbiorcy. Istotne jest tutaj głębsze rozumienie adaptacyjności metod wyznaczania reprezentacji informacji, jeśli to możliwe - w zintegrowanym schemacie od akwizycji po prezentację odbiorcy. Już na etapie próbkowania, zamiast reguł klasycznych, zgodnych z twierdzeniem Nyquista-Shannona, rozważana jest możliwość oszczędnego pomiaru reprezentacji rzadkiej względem dobranej formy przekształcenia przestrzeni mierzonej. Postać metody rekonstrukcji naturalnej reprezentacji sygnałów rejestrowanych może być zależna od istotnych cech sygnału oraz realnych wymuszeń. Nierzadko wprowadza się możliwość interakcji z odbiorcą w celu ustalenia bardziej użytecznej reprezentacji przekazywanej informacji.

3.1 Wprowadzenie

Przyjmijmy ogólnie, że **aproksymacja** to proces określania rozwiązań przybliżonych na podstawie charakterystyk zadanych (zmierzonych, sugerowanych, ustalonych wymuszeniem modelu itp.). Rozwiązania przybliżone są przy tym bliskie zadany ustaleniom w ściśle sprecyzowanym sensie. Przyjmowane kryterium aproksymacji, tak w sensie jakościowym jak i ilościowym, wynika zwykle z celu takich przybliżeń odnoszącego się do zasadniczych walorów użytkowych. Regułą jest, że obiekty (modele, konstrukcje) złożone aproksymowane są za pomocą form prostszych. Formy te stanowią zestaw definiujący (rozpinający) przestrzeń przybliżeń (przyjmijmy – wektorową), który w sensie formalnym może być bazą, ramą czy ogólnie **układem wektorów**.

Przybliżenie (modelowanie) źródła informacji poprzez oszczędny, inaczej nie-nadmiarowy opis treści z pominięciem szumów i zakłóceń charakterystycznych dla realnego sygnału mierzonego jest istotą zastosowań teorii aproksymacji. Układ ustalonej, oszczędnej przestrzeni opisu sygnału narzuca sposób pomiaru czy reprezentowania informacji w dziedzinie stosownego przekształcenia oraz warunki wyznaczenia postaci przybliżenia.

Te dwa aspekty stają się niezwykle ważne przy reprezentowaniu przybliżanej postaci informacji:

- syntetyczny: przybliżenie sygnału źródłowego z kryterium reprezentowania całej informacji w sposób możliwie jednoznaczny, przejrzysty i przede wszystkim użyteczny;
- analityczny: wyrażenie przybliżanej informacji za pomocą szeregu składników, odpowiadających najlepiej poszczególnym, istotnym właściwościom przekazywanej informacji.

Reprezentacja sygnałów za pomocą przestrzeni liniowych oraz szeregów funkcji (sygnałów) bazowych służy w pierwszej kolejności analizie tych sygnałów. Komponenty sygnału zawierające informację stają się łatwiej separowalne, przez co powstaje możliwość syntezy przybliżonej postaci sygnału, reprezentującego istotną informację. Proces syntezy przybliżonej wersji sygnału o bardziej pożądanych cechach jest głównym zadaniem aproksymacji. Celem jest utworzenie przybliżonej reprezentacji, selektywnie zawierającej pełną informację dla odbiorcy, bez maskujących treści nieistotnych, niejednoznacznych, bez zakłóceń, artefaktów, ograniczających czytelność obrazu efektów szumowych. Aproksymacja informacji obrazowej ma ułatwić detekcję obszarów zainteresowania, analizę ich cech, rozpoznanie potencjalnych obiektów – istotnych elementów odczytu treści, segmentację i opisanie tych zmian cechami różnicującymi, wreszcie ich klasyfikację w kategoriach zastosowania.

Wydobyta czy uwydatniona treść może być w dalszej kolejności efektywnie przetwarzana, analizowana i interpretowana przez odbiorcę, zwiększając użyteczność dostarczanych danych. Celem doskonalonych schematów aproksymacji obrazów jest zwykle ekstrakcja obiektów ukrytych, słabo widocznych, klarowne wydzielenie tekstur i krawędzi obiektów, selekcja i lepsza charakterystyka obiektów istotnych, redukcja artefaktów, szumu czy też nieużytecznych elementów obrazu.

W wielu przypadkach opisywania informacji ukrytej, mającej istotny wpływ na interpretację treści obrazowej, subtelna zmiana jest lokalnym zaburzeniem typowego, określonego szerokim kontekstem przebiegu sygnału. Przy funkcjonalnym modelu sygnału, gdzie sygnał opisywany jest za pomocą pewnej klasy funkcji, związane jest to z występowaniem osobliwości, tj. podzbiorów dziedziny, gdzie funkcja zachowuje się nietypowo, czyli inaczej niż w pozostałych obszarach dziedziny. Przykładowo, funkcja nie jest w ogóle określona, jest nieciągła, nie jest

różniczkowalna itp. W przypadku sygnałów 1W są to punkty osobliwe (wymiarowości 0), dla obrazów są to gładkie krzywe konturów (osobliwości wymiarowości 1). Stosując terminologię funkcjonalnej analizy sygnałów możemy mówić o aproksymacji **funkcji celu** w sygnale opisanym jako funkcja źródłowa.

3.2 Reprezentowanie modeli sygnałów

Rozważania dotyczą naturalnych nośników informacji mierzonych czy rejestrowanych za pomocą realnych systemów akwizycji. Ze względu na typowe ograniczenia tych systemów, wynikające m.in. z fizycznych uwarunkowań samej metody pomiarowej, ze skończonej precyzji i wiarygodności czujników, prawdopodobnych źródeł błędu systematycznego i losowego, chwilowych zakłóceń, niestabilności (niestacjonarności) samego procesu mierzzonego, zasadne jest pytanie o reprezentatywność źródłowego zapisu sygnału; na ile zebrana informacja jest reprezentatywna w odniesieniu do rejestrowanej rzeczywistości. Źródłowa reprezentacja sygnału analizowana jest także pod kątem jej użyteczności w kontekście przekazu informacji do odbiorcy.

Uzasadnia to wysiłki zmierzające do wytworzenia reprezentacji bardziej przydatnej użytkowo ze względu na a) wiarygodność informacji, b) kompresowalność. Służą temu modele sygnału konstruowane na bazie wiedzy *a priori* (semantyka), charakterystyki użytkownika (pragmatyka) oraz wymagań technologicznych (syntaktyka). Wiarygodność opisu modelu weryfikują kryteria pragmatyczne, jego konstrukcję warunkuje semantyka, natomiast kryterium maksymalnej redukcji złożoności modelu kształtuje jego cechy syntaktyczne.

Kompresowalność sygnału oznacza możliwość redukcji liczby jego źródłowych próbek - reprezentantów (przyjmijmy, że stanowią one zbiór dyskretny, ziarnisty). Informacja z wystarczającą wiarygodnością jest wtedy reprezentowana przez mniejszy zbiór (ciąg, zestaw) parametrów modelu, inaczej współczynników. Kompresja, inaczej kompresowalne kodowanie wykazuje więc podatność pomierzonej reprezentacji źródłowej na modelowanie ze zredukowaną liczbą parametrów. Analizując źródłową postać sygnału nie kwestionuje jej, nie podważa. Krok dalej idzie koncepcja oszczędnego próbkowania (*compressed sensing* - CS) wykorzystując kompresowalność sygnałów już na etapie akwizycji. Wymaga to rozszerzenie procesu optymalizacji schematu kompresji o etap pomiaru. Pomiar staje się "skompresowany", czyli małowyczący w stosunku do warunków określonych klasycznym twierdzeniem o próbkowaniu (Shannona-Nyquista). Można to nazwać kompresującą recepcją sygnału.

Zderzają się tutaj dwie koncepcje. W klasycznym CS pomiar jest niespójny (niekoherentny) z przekształceniem kompresującym, o losowej lokalizacji w dziedzinie pomiaru. Pomiar jest wtedy rozumiany jako odczytywanie kolejnych próbek za pomocą wektora losowo zestawionej kombinacji parametrów modelu.

Małoliczny zbiór takich pomiarów pozwala zrekonstruować wiarygodny model sygnału.

3.3 Modele sygnałów

Jedną z podstawowych form optymalizacji przekazu informacji jest jej modelowanie i ekstrakcja na podstawie zwartego opisu sygnałów dyskretnych, tj. takiego, w którym nośnik sygnału, rozumiany jako dziedzina jego określoności w modelu funkcyjnym, jest domknięty (jego dopełnienie jest otwarte) i ograniczony (nie rozciąga się do nieskończoności). Zwarty nośnik sygnału staje się bardziej oszczędny, jeśli energię sygnału udaje się upakować w stosunkowo niewielkiej względem rozmiaru nośnika liczbie jego próbek (ziaren, reprezentantów). Mówimy wtedy o reprezentacji rzadkiej (*sparse*), w której liczba niezerowych próbek jest mała w stosunku do rozmiaru nośnika. Jeśli dodatkowo

Biorąc pod uwagę zasadnicze koncepcje teorii informacji, zwarty opis sygnału zakłada w najlepszym przypadku zupełną niezależność elementów (inaczej: parametrów modelu, współczynników) jego dyskretniej reprezentacji. Współczynniki te zwykle nie są bezpośrednimi próbkami pomiarowymi sygnału źródłowego. Są zaś często wagami rozwinięcia sygnału w zbiorach atomów przestrzeni czas-częstotliwości. Rozwinięcia realizowane są często w procesie kodowania/kompresji pomierzonych, dyskretnych sygnałów źródłowych na etapie ich archiwizacji czy w ramach formowania strumienia danych przesyłanych/wymienianych.

Niezależność wektora współczynników może być rozumiana w sensie probabilistycznym, jako brak statystycznej zależności pomiędzy kolejnymi wartościami współczynników. Można to weryfikować jedynie na podstawie wielokrotnych opisów takimi współczynnikami sygnałów możliwie szerokiej przestrzeni zupełnej. Liniową niezależność można badać określając współczynnik korelacji lub też algebraicznie weryfikując warunek liniowej niezależności wektorów.

Odmiennej charakter pojęcia niezależności może być brany pod uwagę w kontekście przekazu informacji – jako niezależność semantyczna. Wtedy każdy z komponentów reprezentuje całkowicie odmienną (niezależną) treść semantycznie złożonego przekazu. Niezależność semantyczna może być konstruowana na bazie niezależności prostszej (probabilistycznej), niejako na wyższym poziomie abstrakcji z ziarnem właściwym podstawowemu poziomowi abstrakcji (probabilistycznej niezależności wektorów wartości).

Występują trzy podstawowe formy zwartości reprezentacji: a) rzadka, b) upakowana, c) uporządkowana (zagnieżdżona).

Rozważając odpowiednią złożoność problemu można założyć, że sygnał f składa się z K składników o różnej morfologii $f = \sum_{i=0}^K f_i$, np. obraz składa się z kilku obiektów o różnej "budowie" (charakterystyce teksturowej, kształcie itp.), a ponadto istnieje słownik baz, tj. zbiorów wektorów bazowych, efektywnej re-

prezentacji sygnałów. Znaczący to, że każdy ze składników f_i może mieć możliwie rzadką reprezentację za pomocą określonej bazy φ_i 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 do poszczególnych składników bazy. 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.

Przybliżenia źródła informacji poprzez upakowany, czyli nie nadmiarowy opis treści z pominięciem szumów i zakłóceń charakterystycznych dla sygnału rzeczywistego jest istotą teorii aproksymacji. Baza ustalonej przestrzeni opisu sygnału narzuca sposób reprezentowania informacji w dziedzinie stosowanego przekształcenia oraz warunki wyznaczenia postaci przybliżenia. Szczególne miejsce w poszukiwaniu efektywnej reprezentacji sygnału niosącego informację zajmuje koncepcja nieliniowej aproksymacji, polegająca na przybliżeniu najbardziej istotnych cech sygnału za pomocą liniowej kombinacji wybranych elementów bazy w ustalonej przestrzeni aproksymacji. Funkcje (ogólniej wektory) bazowe, z których składamy to przybliżenie, najlepiej reprezentują określone właściwości sygnału. Doskonalenie schematu nieliniowej aproksymacji polega na doborze efektywnej bazy przybliżeń, która za pomocą możliwie niewielkiej liczby wyselekcjonowanych elementów pozwala dokładnie opisać zbiór wszystkich istotnych właściwości sygnału.

Reprezentacja sygnałów za pomocą szeregu Fouriera ma ograniczone zastosowanie w analizie lokalnych cech sygnału, które są decydujące w interpretacji obrazów medycznych. Dlatego też szeroko stosowane są skalowalne bazy funkcji rozpinające przestrzeń kolejnych przybliżeń z zachowaniem lokalności opisu sygnału. Przykładem są bazy funkcji falkowych.

Modele są następujące:

- kompresowalny (*compressible*) – typowy dla sygnałów naturalnych, w trzech postaciach (działanie – efekt):
 - działanie: układ atomów dający rzadki zbiór parametrów modelu
 - efekt: **model rzadki**, tj. reprezentacja rzadka względem dziedziny sygnału
 - opis: rzadki wzorzec (*sparsity pattern*) plus rozkład wartości
 - działanie: pakowanie wzorca rzadkiego, czyli koncentracja niezerowych danych w wąskich regionach dziedziny
 - efekt: **model upakowany**, tj. reprezentacja modelu rzadka i upakowana (zagęszczona) w centrach dziedziny sygnału (aspekt semantyczny lub nie); upakowanie rzadkiego sygnału - *compacted sparsity pattern*

- działanie: porządkowanie wzorca rzadkiego
- efekt: **model zagnieżdżony**, tj. reprezentacja modelu rzadka i uporządkowana według kryteriów użyteczności (aspekt semantyczny)
- niekompresowalny (*incompressible*)
 - podział na kompresowalne *patches* (regiony, obiekty, kawałki)
 - reprezentacje aproksymacji nieliniowej (selektywne kryterium istotności)
- rzadki (*sparse*)
 - niezależność w sensie probabilistycznym (PCA, ICA)
 - niezależność w sensie semantycznym (komponenty odmiennej treści - wyższy poziom abstrakcji, może być na bazie niezależności prob.)
 - struktura

Przyjmijmy, że rozważany sygnał jest elementem **przestrzeni Hilberta** \mathcal{H} . Syntetycznie, jest to przestrzeń a) unitarna, czyli liniowa z określonym iloczynem skalarnym $\langle \cdot, \cdot \rangle$ o odpowiednich właściwościach oraz normie $\|\cdot\|_{\mathcal{H}}$, b) zupełna z bazą ortonormalną $\{\varphi_i\}_{i=1,2,\dots}$.

Każdy sygnał $f \in \mathcal{H}$ może być opisany w tej bazie jako:

$$f = \sum_{i=1}^{+\infty} a_i \varphi_i \quad (3.1)$$

Parametrami takiego modelu są współczynniki ortogonalnego rozwinięcia $a_i = \langle f, \varphi_i \rangle$. Liniową reprezentację (aproksymację) sygnału stanowi rzut f na podprzestrzeń $\mathcal{H}_N \in \mathcal{H}$, rozpiętą na skończonej liczbie N pierwszych elementów bazy: $\mathcal{H}_N := \text{span}\{\varphi_i : 1 \leq i \leq N\}$.

Wartość N dobierana jest *a priori* i służy regulacji dokładności przybliżenia. \mathcal{H}_N jest więc liniowa, zaś liniowe przybliżenie $f \in \mathcal{H}$ wygląda następująco:

$$\tilde{f}_{(N)} = \sum_{i=1}^N a_i \varphi_i \quad (3.2)$$

3.4 Liniowy problem odwrotny

Częstym zagadnieniem opisującym ograniczenia procesu rejestracji sygnałów jest problem odwrotny, formułowany dla systemów liniowych w postaci:

$$y = \mathcal{M}f_t + \eta \quad (3.3)$$

gdzie $f_t \in \mathbb{R}^N$ jest poszukiwanym wektorem danych, którego nie można zaobserwować (pomierzyć) bezpośrednio, zaś $y \in \mathbb{R}^M$, $M \leq N$ jest wektorem obserwacji

(pomiarów) źródłowych, maskowanych addytywnym η o ograniczonej wariancji (energii) σ_η^2 . Macierz \mathcal{M} opisuje ograniczony operator liniowy $\mathbb{R}^N \rightarrow \mathbb{R}^M$, który modeluje ograniczenia związane z realną akwizycją bądź jej symulacją (np. w warunkach, gdy zarejestrowany wcześniej zbiór danych musi posłużyć dodatkowym celom analizy, nieprzewidywanym na etapie rejestracji). W zależności od zastosowań, rozumienie roli i charakteru \mathcal{M} może być różnorakie, zaś składnik η opisuje szum, artefakt lub inny element mierzonego sygnału złożonego, który nie jest informatywny, czyli nie jest nośnikiem informacji.

Przykładowo, przy modelowaniu ograniczeń pasma przenoszenia systemu akwizycji, macierz \mathcal{M} charakteryzuje rozmycie szczegółów docelowego sygnału f_t , np. metodą splotu z filtrem wygładzającym. Wtedy $M = N$. Innym razem jest to rozkład wartości określających zależność pomiędzy rejestrowanymi projekcjami y , a poszukiwanym wektorem rekonstruowanej f_t o zadawalającej rozdzielczości, przy $M < N$, a nieraz $M \ll N$.

Schemat (3.3) pozwala opisać szczególnie interesujące przypadki ograniczonej rozdzielczości oraz maskowania pola widzenia. Problem aproksymacji f_t o wysokiej rozdzielczości w stosunku do limitowanej rozdzielczości zarejestrowanej y , nazywany problem super-rozdzielczości – SR (ang. *superresolution*), może być opisany poprzez s -krotne podpróbkowanie, bądź też złożenie operacji rozmycia (filtrem o szerokości s) i podpróbkowania. Wtedy $M = N/s$.

Z kolei problem uzupełniania danych, czyli tzw. wmalowania (ang. *inpainting*) sprowadza się do \mathcal{M} jako maski binarnej (1 – piksel rejestrowany, 0 – piksel przesłonięty).

Ogólnie można wyróżnić kilka kategorii liniowego problemu odwrotnego dotyczącego rejestrowanych obrazów:

- ekstrakcji sygnału docelowego f_t na podstawie niedokładnych pomiarów pikselowych (orientowanych na pojedyncze piksele), dla $M = N$;
- uzupełnienia sygnału f_t na podstawie niepełnych pomiarów pikselowych (ograniczona rozdzielczość, martwe pola rejestracji), a więc przy $M < N$;
- rekonstrukcji sygnału f_t na podstawie ograniczonej liczby pomiarów projekcyjnych (uśrednionych, nie orientowanych na pojedyncze piksele), przy $M \ll N$;
- rejestracji kompresowalnego, czyli potencjalnie nadmiarowego f_t za pomocą niewielkiej liczby pomiarów losowych, przy $M \ll N$; jest to przypadek *compressive sensing*, zmieniającego typowy schemat próbkowania (pozawala np. za pomocą niewielkiej liczby próbek wyrazić sygnały bez ograniczeń pasma, jak w klasycznej metodzie próbkowania Nyquista-Shannona).

Zasadniczym sposobem rozwiązywania takich problemów jest aproksymacja, przybliżanie rozwiązania z kryterium jakościowym dopasowanym do zastosowa-

nia. Szczególnego znaczenia nabiera tutaj wykorzystywane narzędzie (kryteria, bazy przestrzeni, założenia dotyczące klasy i cech f_t itp.).

3.5 Aproksymacja informacji z sygnału

Reprezentacja sygnałów za pomocą przestrzeni liniowych i szeregów funkcji (sygnałów) bazowych służy w pierwszej kolejności analizie tych sygnałów. Komponenty sygnału zawierające informację stają się łatwiej separowalne, przez co powstaje możliwość syntezy przybliżonej postaci sygnału, reprezentującego istotną informację. Proces syntezy przybliżonej wersji sygnału o bardziej pożądanym cechach nazywamy *aproksymacją*.

Chodzi o przybliżoną reprezentację, selektywnie zawierającą pełną informację dla odbiorcy, bez maskujących treści nieistotnych, niejednoznacznych, bez zakłóceń, artefaktów, ograniczających czytelność obrazu efektów szumowych. Aproksymacja informacji obrazowej ma ułatwić detekcję obszarów zainteresowania, analizę ich cech, rozpoznanie potencjalnych obiektów – istotnych elementów odczytu treści, segmentację i opisanie tych zmian cechami różnicującymi, wreszcie ich klasyfikację w kategoriach zastosowania.

Przybliżenia źródła informacji poprzez upakowany, czyli nie nadmiarowy opis treści z pominięciem szumów i zakłóceń charakterystycznych dla sygnału rzeczywistego jest istotą teorii aproksymacji. Baza ustalonej przestrzeni opisu sygnału narzuca sposób reprezentowania informacji w dziedzinie stosowanego przekształcenia oraz warunki wyznaczenia postaci przybliżenia.

Reprezentacja sygnałów za pomocą szeregu Fouriera ma ograniczone zastosowanie w analizie lokalnych cech sygnału, które są decydujące w interpretacji obrazów medycznych. Dlatego też szeroko stosowane są skalowalne bazy funkcji rozpinające przestrzeń kolejnych przybliżeń z zachowaniem lokalności opisu sygnału.

3.5.1 Podstawowe schematy aproksymacji sygnałów

Podstawowe rozwiązania w zakresie aproksymacji sygnałów wykorzystują schemat liniowy lub nieliniowy. Poniżej przedstawiono formalne podstawy tych metod, które przekładają się na realne efekty przybliżania istotnych cech sygnału.

Pojęcie aproksymacji sygnału definiowane jest za pomocą **przestrzeni Hilberta** \mathcal{H} . Ogólnie jest to przestrzeń a) unitarna, czyli liniowa z określonym iloczynem skalarnym $\langle \cdot, \cdot \rangle$ o odpowiednich właściwościach oraz normie $\|\cdot\|_{\mathcal{H}}$, b) zupełna z bazą ortonormalną $\{\varphi_i\}_{i=1,2,\dots}$. Każdy sygnał $f \in \mathcal{H}$ może być reprezentowany w tej bazie jako: $f = \sum_{i=1}^{+\infty} a_i \varphi_i$, gdzie współczynniki ortogonalnego rozwinięcia $a_i = \langle f, \varphi_i \rangle$. Liniową aproksymację sygnału stanowi rzut f na podprzestrzeń $\mathcal{H}_N \in \mathcal{H}$, rozpiętą na skończonej liczbie N pierwszych elementów bazy: $\mathcal{H}_N := \text{span}\{\varphi_i : 1 \leq i \leq N\}$. Wartość N dobierana jest *a priori* i służy regulacji

dokładności przybliżenia. \mathcal{H}_N jest więc liniowa, zaś liniowe przybliżenie $f \in \mathcal{H}$ wygląda następująco:

$$\tilde{f}_{(N)} = \sum_{i=1}^N a_i \varphi_i \quad (3.4)$$

W przypadku aproksymacji nieliniowej wykorzystywana jest nieliniowa podprzestrzeń rzutowania sygnału \mathcal{A}_N , dająca przybliżenie $f \in \mathcal{H}$ jako rozwinięcie:

$$\tilde{\tilde{f}}_{(N)} = \sum_{i \in \Lambda} a_i \varphi_i \quad (3.5)$$

$\Lambda \subset \mathbf{N}$ jest zbiorem indeksów, przy czym liczba kardynalna tego zbioru $\#\Lambda = M \leq N$. Specyficzny dla nieliniowej (M elementowej) aproksymacji jest określony wybór bazy rozpinającej \mathcal{A}_N , która zależy od f . Nieliniowość tej operacji wynika z faktu, iż dla operatora aproksymacji nieliniowej $\tilde{\tilde{F}}_{(N)}$ można znaleźć $a, b \in \mathcal{H}$, dla których $\tilde{\tilde{F}}_{(N)}(a + b) \neq \tilde{\tilde{F}}_{(N)}(a) + \tilde{\tilde{F}}_{(N)}(b)$. W typowej realizacji aproksymacji nieliniowej dokonujemy dekompozycji f w \mathcal{H} (lub przy praktycznych ograniczeniach w \mathcal{H}_N), a następnie wybieramy M funkcji φ_i o największych współczynnikach $|a_i|$ jako przybliżenie według (3.7). Uzyskujemy wtedy minimalny błąd przybliżenia ($\|f - \tilde{\tilde{f}}\|$). Rozwiązywany praktycznie problem dotyczy zwykle reprezentacji sygnału opisywanego w przestrzeni funkcji całkowalnych z kwadratem $L^2(\mathbb{R})$, czyli szczególnej postaci przestrzeni Hilberta, gdzie iloczyn skalarny $\langle f, g \rangle = \int_{t \in \mathbb{R}} f(t)g^*(t)dt$, a norma $\|f\| = \sqrt{\int_{t \in \mathbb{R}} |f(t)|^2 dt}$.

Jeśli elementy bazy lepiej wpasowują się w sygnał, tj. w sposób bardziej reprezentatywny opisują istotne właściwości sygnału, wtedy na tym samym poziomie dokładności sygnał w zakresie interesujących cech opisywany jest za pomocą mniej licznych zestawu elementów bazy. Mówi się wtedy, że f jest lepiej reprezentowany w danej bazie, co pozwala efektywniej identyfikować sygnał. Teoria aproksymacji zajmuje się badaniem błędów tych przybliżeń przy różnych schematach aproksymacji budowanych na podstawie baz ortonormalnych.

Biorąc pod uwagę uniwersalność oraz użyteczność metod aproksymacji wydaje się, że najlepszym rozwiązaniem byłby schemat aproksymacji liniowej, który gromadzi współczynniki w porządku nierosnącym ich amplitud. Chcąc zachować lokalny opis sygnału w dziedzinie współczynników rozwinięcia można domagać się jedynie maksymalnego upakowania nieliniowej reprezentacji, czyli maksymalnego skupienia, najlepiej bezpośredniego sąsiedztwa współczynników w możliwie małej liczbie centrów ich koncentracji. Wśród miar należy wymienić entropię, średnią odległość skupisk względem centroidu itp.

Błąd aproksymacji

Błąd aproksymacji liniowej można określić jako: $\tilde{\varepsilon}_N^2(f) = \inf_{\tilde{f} \in \mathcal{H}_N} \|f - \tilde{f}\|_{\mathcal{H}}$, a nieliniowej: $\tilde{\tilde{\varepsilon}}_N^2(f) = \inf_{\tilde{\tilde{f}} \in \mathcal{A}_N} \|f - \tilde{\tilde{f}}\|_{\mathcal{H}}$. Przy ortogonalnym rzucie f na \mathcal{H}_N można zapisać zależność na energię błędów aproksymacji jako $\tilde{\varepsilon}_N^2(f) = \sum_{i=N+1}^{+\infty} |a_i|^2$,

przy czym błąd zbiega do zera dla dużych N : $\lim_{N \rightarrow +\infty} \tilde{\varepsilon}_N^2(f) = 0$. Analogicznie przy procedurze nieliniowej mamy błąd $\tilde{\varepsilon}_N^2(f) = \sum_{i \notin \Lambda} |a_i|^2$. Błąd ten zanika przy rosnącym $\#\Lambda$, jeśli zbiór Λ określimy jako zbiór indeksów bazy $\{\varphi_i\}_{i=1,2,\dots}$ uporządkowanych według monotonicznie nierosnących wartości odpowiadających im współczynników $|a_i|$ rozwinięcia f w \mathcal{H} . Jeśli więc uporządkowany zbiór $\Lambda_r = \{i_k\}_{k=1,2,\dots,M}$, gdzie $|a_{i_k} = \langle f, \phi_{i_k} \rangle| \geq |a_{i_{k+1}}|$, $0 < k < M$ mamy więc nieliniowe przybliżenie: $\tilde{f}_{(M)} = \sum_{k=1}^M a_{i_k} \varphi_{i_k}$ oraz błąd $\tilde{\varepsilon}_N^2(f) = \sum_{k=M+1}^{+\infty} |a_{i_k}|^2$. Jakość aproksymacji danego schematu zależy od szybkości zanikania błędu aproksymacji przy rosnącym M , czyli *de facto* od szybkości zanikania $|a_{i_k}|$. Zamiast porządkowania współczynników, często stosowana jest metoda progowania modułu ich wartości. Obniżanie wartości progów, zaczynając od maksymalnej bezwzględnie wartości współczynnika, powoduje stopniowe dołączanie następnych $|a_{i_k}|$ według informacji przez nich reprezentowanej.

Niekiedy potrzebna jest bardziej złożona reguła porządkowania współczynników ze względu na ich znaczenie w opisie istotnych cech sygnału. Można na przykład uwzględnić rozkład wartości współczynników w bezpośrednim sąsiedztwie, rozpoznając obszar istotny lub też jedynie zaszumiony region o niewielkim znaczeniu. Takie rozwiązanie wymaga lokalności opisu cech sygnału przez dany współczynnik (czego nie zapewnia np. baza fourierowska) oraz zdefiniowanych śladów istotnych zmian (struktur, obiektów) w rzucie na daną podprzestrzeń.

Przybliżenie źródła informacji poprzez nienadmiarowy, upakowany opis treści z pominięciem szumów i zakłóceń charakterystycznych dla sygnału rzeczywistego jest istotą teorii aproksymacji. Baza ustalonej przestrzeni opisu sygnału narzuca sposób reprezentowania informacji w dziedzinie stosowanego przekształcenia oraz warunki wyznaczenia postaci przybliżenia. Informacja reprezentowana przez sygnał jest tutaj rozumiana jako znacząca wartość współczynnika, wyrażająca udział określonej funkcji bazowej w dekomponowanym sygnale. Stanowi to o dużej wadze wyboru bazy takiego przekształcenia. Kryteria doboru funkcji bazowych przestrzeni powinny wykorzystywać miary dopasowania elementów bazy do istotnych właściwości sygnału, czyli do semantyki informacji reprezentowanej w analizowanym sygnale (obrazie). Dobierając bazę i ustalając metodę porządkowania współczynników, można zwiększyć skuteczność aproksymacji, czyli poprawić jakość modelu opisującego istotne właściwości sygnału (tj. dokładniej opisać reprezentatywne cechy sygnału, uzyskując dopuszczalny błąd aproksymacji przy mniejszej złożoności modelu).

Estymacja sygnału ukrytego

Problem estymacji sygnału ukrytego lub ogólniej - ekstrakcji sygnału użytecznego f ze źródłowego sygnału s można opisać prostym schematem: $s = f + \eta$, gdzie η jest składnikiem maskującym (szum addytywny, artefakt lub inny element mierzzonego sygnału złożonego, który nie jest informatywny, czyli nie jest nośnikiem

informacji). Aproksymacja s sprowadzona do wiernego przybliżenia jedynie składowej f z pominięciem η nie jest zadaniem łatwym. Wymaga przede wszystkim dobrej charakterystyki właściwości f , która różnicuje sygnał użyteczny w stosunku do nieinformatywnego tła, a charakterystyka ta w przypadku kiedy nie ma stabilnych wzorców f , może być bardzo trudna. Wyznaczenie upakowanej reprezentacji poprzez selekcję najlepiej przydatnych elementów bazy, z ewentualną modyfikacją ich procentowego udziału w przybliżanym sygnale, służy takiemu właśnie wydzielaniu f w procesie estymacji informacji trudno dostępnej dla odbiorcy. Nieliniowa aproksymacja sygnału jest sposobem takiej estymacji.

Aproksymacja nieliniowa realizowana jest zwykle poprzez progowanie modułu wartości współczynników $a_i^{(s)}$ rozwinięcia s w bazie $\{\varphi_i\}$ nowej dziedziny, opisane zależnością $\hat{f} = \tilde{s} = \sum_{i=1}^N d(a_i^{(s)})\varphi_i$, gdzie $d(\cdot)$ jest funkcją progowania. Chcemy, aby $\hat{f} \approx f$.

Odrzucenie współczynników o module mniejszym od wartości progu τ realizowane jest poprzez tzw. progowanie twarde: $d_{\tau}^{(hard)}(a_i^{(s)}) = \begin{cases} a_i^{(s)} & \text{dla } |a_i^{(s)}| \geq \tau \\ 0 & \text{dla } |a_i^{(s)}| < \tau \end{cases}$ Wśród innych funkcji progowania warto wymienić progowanie półmiękkie, wykorzystujące dwie wartości progów τ_1 i τ_2 :

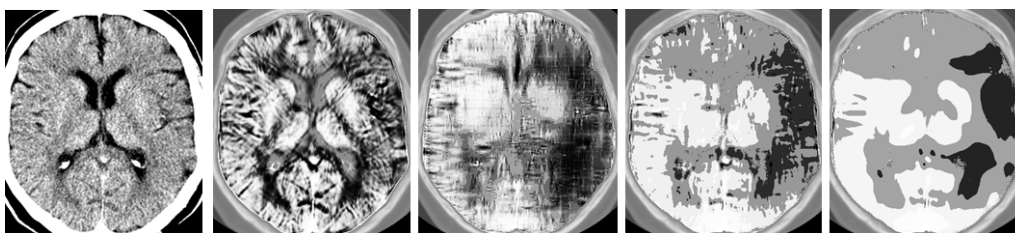
$$d_{\tau_1, \tau_2}^{(semisoft)}(a_i^{(s)}) = \begin{cases} a_i^{(s)} & \text{for } |a_i^{(s)}| > \tau_2 \\ \text{sign}(a_i^{(s)}) \frac{\tau_2(|a_i^{(s)}| - \tau_1)}{\tau_2 - \tau_1} & \text{for } \tau_1 \leq |a_i^{(s)}| \leq \tau_2, \text{ co daje bar-} \\ 0 & \text{for } |a_i^{(s)}| < \tau_1 \end{cases}$$

dziej naturalną postać \hat{f} . Dobór progów powinien uwzględniać zróżnicowanie wartości modułów współczynników f (potencjalnie większych od progu) oraz η (potencjalnie mniejszych od progu). Jest to możliwe przy założeniu, że dekompozycja sygnału źródłowego s w dobranej bazie ma własności różnicujące. W przypadku obrazów dobiera się często wartość progu bliską poziomowi $\tau = 3\sigma$, gdzie σ oznacza szacowaną wartość odchylenia standardowego części nieinformatywnej sygnału, traktowaną jako szum. W przypadku wykorzystania bardziej złożonej, semantycznej charakterystyki informacji i składników nieużytecznych wymagane są zaawansowane, specyficzne, bazujące na analizowanych wzorcach i bazach treningowych metody selekcji i modyfikacji współczynników $a_i^{(s)}$.

Przykładowe wydobycie informacji ukrytej w obrazie źródłowym pokazano na rys. 3.1.

Aproksymacja sygnałów bez względu na mechanizm czy algorytm definiujący przybliżenie obiektu docelowego (funkcji docelowej, rozkładu docelowego etc.), bazuje zwykle na następujących założeniach:

- sygnał wejściowy (pomierzony i podlegający procedurze aproksymacji) jest określony w przestrzeni Hilberta (z iloczynem skalarnym definiującym normę), zwykle $L^2(\mathbb{R})$ lub $l^2(\mathbb{R})$;
- poszukiwana jest dyskretna, często skończona, reprezentacja zawierająca



Rysunek 3.1: Estymacja ukrytego sygnału użytecznego w specjalistycznym zastosowaniu; wybrana warstwa badania tomografii komputerowej głowy nie wskazuje przy tradycyjnej obserwacji symptomów patologii (pierwszy z lewej obraz) - kolejne obrazy są efektem przetwarzania w dziedzinie falkowej metodami aproksymacji nieliniowej, kiedy to estymowany jest obszar hipodensyjny (o obniżonej gęstości tkanki mózgowia, widoczny na obrazie jako region nieco ciemniejszy w obszarze istoty szarej lub białej), widoczny w obrazie końcowym jako ciemne plamy, wyekstrahowane z obrazu źródłowego.

wartości w \mathbb{C} ; wartości te są określone w dziedzinie sygnału (przykładem tradycyjne próbkowanie według reguły Nyquista-Shannona nakładające ograniczenia na maksymalną częstotliwość sygnału) lub też ogólnie w dziedzinie przekształcenia (transformacji) – przykładem jest reprezentacja sygnału w postaci widma fourierowskiego;

- jakość reprezentacji aproksymującej zasadniczo charakteryzowana jest poprzez:
 - błąd aproksymacji w relacji do liczby danych reprezentacji przybliżającej,
 - podatność reprezentacji na kodowanie,
 - spełnienie wymagań użytkownika, zgodnie z charakterystyką jego pracy.

Nierzadko istotna jest też złożoność obliczeniowa algorytmu rekonstrukcji sygnału.

3.6 Reprezentacja rzadka

Wektor danych $x \in \mathbb{R}^N$ jest postaci rzadkiej (*sparse*), jeśli $\#(\text{sup } x) \leq M$ dla pewnych $M \ll N$.

Niech dyskretny sygnał \mathbf{f} w skończenie wymiarowej podprzestrzeni \mathbb{R}^N będzie opisany za pomocą N pomiarów (próbek) jako wektor: $\mathbf{f} = [f_1, f_2, \dots, f_N]$. Wektor jest **ściśle rzadki**, jeśli liczebność (moc, liczba kardynalna) skończonego zbioru próbek znaczących, tj. niezerowych o indeksach $\Lambda = \{1 \leq i \leq N : f_i \neq 0\}$ wynosi $\#\Lambda = M \ll N$ [?]. W tym przypadku

możemy powiedzieć, że \mathbf{f} jest M -rzadki i dla wygody zapisać: $\|\mathbf{f}\|_0 = M \ll N$, wykorzystując pseudonormę l_0 określającą liczbę niezerowych współczynników x jako $\|x\|_0$.

Większość rejestrowanych czy próbkowanych sygnałów rzeczywistych nie jest rzadka w źródłowej przestrzeni akwizycji, jednak często dość proste przekształcenie pozwala uzyskać kompresowalną postać sygnału w nowej dziedzinie. Obok usunięcia składowej stałej czy zastosowania mniej lub bardziej złożonych modeli predykcji, stosowana jest dekompozycja (analiza) sygnału za pomocą specyficznej klasy funkcji o elementarnych, przybliżających podstawowe lub bardziej złożone formy kształtach, zwanych atomami.

Poszukiwana jest odpowiednia dziedzina reprezentacji sygnału wykorzystując kryterium możliwie rzadkiej reprezentacji przekształconego sygnału. Ogólnie jest to realizowalne przez liniowe rozwinięcie sygnału po atomach słownika, czyli nadmiarowego zestawu $\mathcal{D} = \{\varphi_i\}_{i=1}^T$ jednostkowych wektorów: $\forall_i \|\varphi_i\|_2 = 1$, takich że $\text{span}\{\varphi_i\} = \mathbb{R}^N$ (w \mathcal{D} mogą być reprezentowane wszystkie rzeczywiste sygnały). Jeśli $\{\varphi_i\}$ są liniowo zależne, słownik jest nadmiarowy.

Dobór postaci rozwinięcia w \mathcal{D} prowadzi do upakowanej reprezentacji dużej klasy sygnałów według koncepcji selektywnego próbkowania. Sygnał \mathbf{f} jest liniową syntezą M odpowiednio dobranych (metodą ad-hoc lub też za pomocą bardziej złożonych, mniej lub bardziej formalnych procedur) atomów słownika \mathcal{D} , zgodnie z zależnością

$$\mathbf{f} = \sum_{i=1}^T a_i \varphi_i \quad (3.6)$$

Dekompozycja \mathbf{f} w \mathcal{D} jest wektorem współczynników $\mathbf{a} = [a_1, a_2, \dots, a_M] \in \mathbb{R}^N$ oraz $\mathbf{a} = \langle \mathbf{f}, \varphi_{i_j} \rangle$. Dąży się, by reprezentacja współczynników \mathbf{a} była możliwie rzadka.

Zapewniające dużą elastyczność reprezentacji atomy rozwinięcia w \mathcal{D} mogą być adaptacyjnie dobierane do morfologicznej treści sygnału, zgodnie z wiedzą *a priori* oraz przyjętymi założeniami i ustaleniami, dotyczącymi poszukiwanej formy reprezentacji rzadkiej. Rzadkie rozwinięcie sygnału w specyficznym $\Phi = \{\varphi_{i_j}\}_{j=1}^M$ oznacza, że tylko M dobranych atomów \mathcal{D} jest aktywnych w opisie \mathbf{f} . Inaczej, sygnał został zamodelowany za pomocą niewielkiej liczby wybranych atomów, najlepiej przybliżających cechy sygnału, które stanowią przedmiot zainteresowania (tj. stanowią informację).

Formalnie, problem dokładnego wyznaczenia reprezentacji rzadkiej sygnału (tj. interpolacji M -rzadkiej) sprowadza się to wyznaczenia

$$\min_{\mathbf{a}} \|\mathbf{a}\|_0 \text{ przy ograniczeniu } \mathbf{f} = \sum_{j=1}^M a_j \varphi_{i_j} \quad (3.7)$$

co oznacza wyznaczenie najbardziej rzadkiej reprezentacji \mathbf{f} względem \mathcal{D} [?]. Celem stosowanej procedury optymalizacji jest wyszukiwanie odpowiedniego zbioru

składników-aproksymant, jeśli znana jest funkcja aproksymowana.

W praktyce wiele interesujących sygnałów nie ma ściśle rzadkiej reprezentacji w dowolnym rozwinięciu ze względu na ograniczenia akwizycji (obecność szumów), specyfikę zawartej informacji itp. Jednak sygnały te mogą być ”w przybliżeniu” rzadkie, tj. prawie rzadkie lub z rzadką informacją, gdy uporządkowane nierosnąco amplitudy współczynników rozwinięcia szybko zanikają według ustalonej zależności potęgowej.

Rzadka reprezentacja przybliżenia (aproksymacji) \mathbf{f} z błędem $\epsilon \geq 0$ (*error sparse approximation*) jest ustalana według formuły

$$\min_{\mathbf{a}} \|\mathbf{a}\|_0 \text{ przy ograniczeniu } \|\mathbf{f} - \sum_{j=1}^M a_j \varphi_{i_j}\|_2 \leq \epsilon \quad (3.8)$$

Alternatywna procedura uzyskania najlepszej aproksymacji \mathbf{f} przy ograniczeniu M -rzadkiej reprezentacji jej rozwinięcia w \mathcal{D} definiowana jest przez

$$\min_{\mathbf{a}} \|\mathbf{f} - \sum_{j=1}^M a_j \varphi_{i_j}\|_2 \text{ przy ograniczeniu } \|\mathbf{a}\|_0 \leq M \quad (3.9)$$

Celem jest wyznaczenie reprezentacji sygnału \mathbf{f} o ustalonej ”rzadkości” (tj. M -rzadkiej aproksymacji) tak, aby M atomów dało minimalny błąd aproksymacji.

Niestety, przy założeniach ogólnych dowolnie nadmiarowej postaci słownika \mathcal{D} , wyznaczenie rzadkiej reprezentacji sygnału \mathbf{f} jest problemem NP-trudnym. Koncepcja kompresowalnego próbkowania (uzysku) ustala jednak solidne wsparcie na poziomie teoretycznym i algorytmicznym, proponując metody wyznaczania rozwiązań o rzadkiej reprezentacji. Istotną wskazówką *a priori* jest uwaga, że naturalne i najbardziej pożądane słowniki mają lepiej określoną postać. Atomy w postaci funkcji bazowych transformacji ortogonalnych dają korzystne i jednoznaczne rozwiązania problemu reprezentacji rzadkiej. Spójność i konsekwencja konstrukcji słownika jest miarą jego użyteczności. Duże, niespójne słowniki czynią rozwiązanie problemu kompresowalności trudniejszym, mniej podatnym w procedurach optymalizacji.

Reprezentacja danych jest rzadka, jeśli liczba niezerowych współczynników opisujących sygnał w ustalonej bazie jest znikomo mała (znacząco zredukowana) w stosunku do wymiaru dziedziny źródłowej, czyli że wymiar nowej dziedziny, w której reprezentowany jest sygnał, został znacząco zredukowany. Taki zwarty opis sygnału daje zwykle jedynie przybliżoną postać wersji źródłowej, najlepiej przy zachowaniu wszystkich cech sygnału istotnych dla odbiorcy.

Efektywna realizacja wspomnianych schematów aproksymacji skutkuje rzadką reprezentacją informacji wydzielonej z sygnału źródłowego. Zbiór dostępnych narzędzi, pozwalających skuteczniej przybliżyć informację w niektórych zastosowaniach warto rozszerzyć o wykorzystanie nadmiarowych słowników funkcji, łączących zalety wielu baz podstawowych o różnych właściwościach, zaczynając

od baz fourierowskich, a kończąc na skalowalnych, lokalnych bazach funkcji definiowanych w wielowymiarowych przestrzeniach danych źródłowych.

Rzadka reprezentacja informacji f wykorzystuje wtedy lokalny słownik \mathcal{D} baz funkcji (atomów) dobrze przybliżających zbiór pożądaných cech sygnału. Każdej istotnej składowej sygnału może odpowiadać dobrze ją reprezentująca baza. Mogą to być różne zestawy baz ortonormalnych, których elementy są względem siebie jedynie liniowo niezależne. Zwarte rozwinięcie w \mathcal{D} oznacza, że tylko niektóre atomy są aktywne w reprezentacji f , co z kolei może służyć automatycznemu rozpoznaniu treści przekazu, skuteczniejszej klasyfikacji poprzez identyfikację określonych składowych.

3.7 Strategie

Wspomniane zalety rzadkiej reprezentacji sygnałów wykorzystuje się w różnych strategiach. Do najbardziej przydatnych należą:

- reprezentacje kompresowalne (aproksymacja nieliniowa): rzadkie, uporządkowane

- zakładamy: sygnał $x \in \mathcal{H}$
- cechy przestrzeni Hilberta: a) unitarna, czyli liniowa z określonym iloczynem skalarnym $\langle \cdot, \cdot \rangle$ o odpowiednich właściwościach oraz normie $\|\cdot\|_{\mathcal{H}}$, b) zupełna z bazą ortonormalną $\{\psi_i\}_{i=1,2,\dots}$
- sygnał w bazie: $\mathbf{x} = \sum_{i=1}^{+\infty} a_i \psi_i$
- współczynniki ortogonalnego rozwinięcia: $a_i = \langle x, \psi_i \rangle$
- liniowa aproksymacja sygnału: rzut na podprzestrzeń: $\mathcal{H}_N \in \mathcal{H}$, rozpiętą na skończonej liczbie N kolejnych elementów bazy; rozważmy przykład próbkowania w $L^2(\mathbb{R})$ za pomocą fcy grzebieniowej - (w bazie poprzesuowanych delt Kroneckera $\mathcal{H}_N := \text{span}\{\delta_i : 1 \leq i \leq N\}$ jako $\mathbf{x}_{(\mathcal{H}_N)} = \sum_{i=1}^N x(i\Delta t) \delta(t - i\Delta t) = \sum_{i=1}^N x_i \delta_i$
- reprezentacja rzadka sygnału uproszczonego $\mathbf{x} = (x_i)_{i=1}^N \in \mathbb{R}^N$ jeśli

$$\#(\text{supp } \mathbf{x}) = \#\{i : x_i \neq 0\} = \|\mathbf{x}\|_0 \leq K \ll N$$

- liniowo mamy $\mathbf{x} = \sum_{i=1}^N a_i \psi_i \approx \sum_{i=1}^K a_i \psi_i$ with $K \ll N$
- nieliniowa aproksymacja sygnału \mathbf{x} : rzut na podprzestrzeń $\mathcal{A}_K \in \mathcal{H}_N$ tak że

$$\tilde{\mathbf{x}}_{(\mathcal{A}_K)} = \sum_{i \in \Lambda} a_i \psi_i$$

gdzie $a_i = \langle \mathbf{x}, \psi_i \rangle$ oraz $\Lambda = \{i : |a_i| > \tau\} \subset \mathbb{N}$ i $\#\Lambda = K \leq N$
w uproszczeniu $\tilde{\mathbf{x}} = \Psi^T \mathbf{a}$

- błąd aproksymacji nieliniowej

$$\epsilon_{(\mathcal{A}_K)}^2(x) = \inf_{\tilde{x} \in \mathcal{A}_K} \|x - \tilde{x}\|_2^2 \cong \inf_{\tilde{\mathbf{x}} \in \mathcal{A}_K} \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2 \quad (\mathcal{H})$$

co liczy się tak

$$\epsilon_{(\mathcal{A}_K)}^2(x) = \sum_{i \notin \Lambda} |a_i|^2$$

- rzadka reprezentacja uporządkowana z zanikaniem błędu; sprecyzujemy definicję zbioru indeksów współczynników uporządkowanych nierosnących uzyskanego poprzez bijekcją $\mathcal{P} : \Lambda \rightarrow \Lambda_u$, $\mathcal{P}(i) = k$ tak że $\Lambda_u = \{jk \in \Lambda : |a_{jk} = \langle x, \psi_{jk} \rangle| \geq |a_{j(k+1)}|; jk = 1, \dots, K\}$, wtedy $\tilde{\mathbf{x}}_{(\mathcal{U}_K)} = \sum_{jk=1}^K a_{jk} \psi_{jk}$, a błąd

$$\epsilon_{(\mathcal{U}_K)}^2(x) = \sum_{jk=K+1}^{+\infty} |a_{jk}|^2$$

- jakość aproksymacji zależy od szybkości potęgowego zanikania błędu przy rosnącym K ; określa ją parametr $p \in (0, 1)$:

$$\epsilon_{(\mathcal{U}_K)}^2(x) \leq \int_{K+1}^N O((k^{-1/p})^2) dk \leq C_p K^{1-2/p}$$

Reprezentacja jest wtedy p -kompresowalna.

- reprezentacja sygnału kompresowalnego \mathbf{x} jest rzadka przy dopuszczalnym błędzie ε , czyli p musi być odpowiednio małe

$$\|\mathbf{x} - \tilde{\mathbf{x}}_{(\mathcal{U}_K)}\|_2^2 \leq C_p K^{1-2/p} \leq \varepsilon \text{ dla } K \ll N$$

- metoda: interpolacja K -rzadka

$$\min_{\mathbf{a}(\Psi \in \mathcal{D})} \|\mathbf{a}\|_0 \text{ przy ograniczeniu } \mathbf{x} = \sum_{j=1}^K a_j \psi_{ij}$$

- metoda: rzadka reprezentacja nieliniowej aproksymacji z błędem $\varepsilon \geq 0$

$$\min_{\mathbf{a}(\Psi \in \mathcal{D})} \|\mathbf{a}\|_0 \text{ przy ograniczeniu } \|\mathbf{x} - \sum_{j=1}^K a_j \psi_{ij}\|_2^2 \leq \varepsilon$$

- metoda: minimalizacji błędu przybliżenia przy ustalonym K

$$\min_{\mathbf{a}(\Psi \in \mathcal{D})} \|\mathbf{x} - \sum_{j=1}^K a_j \psi_{ij}\|_2^2 \text{ przy ograniczeniu } \|\mathbf{a}\|_0 \leq K$$

- nadmiarowy słownik $\mathcal{D} = \{\psi_i\}_{i=1}^T$, $T > N$; synteza sygnału w nadmiarowym słowniku $\mathbf{x} = \sum_{i=1}^T a_i \psi_i$ nie jest jednoznaczna; optymalizacja rzadkiej reprezentacji dokonuje się poprzez dobór najlepszej bazy, takiej że

$$\mathbf{x} = \sum_{j=1}^K a_j \psi_{ij}, \quad K \ll N$$

- szybkość zanikania falek tensorowych $\epsilon_{(\mathcal{U}_K)}^2(x) = O(K^{-1})$, czyli $p = 1$,
zaś falek kierunkowych $\epsilon_{(\mathcal{U}_K)}^2(x) = O(K^{-2}(\log K)^3)$, czyli $p \approx 0,8$
- podstawowe twierdzenie:
 - * istnieje Φ - nieadaptacyjny zbiór $O(K \log(N/K))$ wektorów w \mathbb{R}^N ,
który można skonstruować w jednej postaci dla wszystkich baz standardowych
 - * wtedy dla ustalonego $p \in (0, 1)$ i dowolnego, p -kompresowalnego w standardowej bazie sygnału \mathbf{x}
 - * mając tylko zbiór pomiarów $\langle \mathbf{x}, \phi_i \rangle$, $\phi_i \in \Phi$
 - * można w czasie wielomianowym wyznaczyć reprezentację x , taką że

$$\|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2 = O(K^{1-2/p}) \quad (3.10)$$

- kodowanie reprezentacji rzadkich

$$\min_{\tilde{\mathbf{x}}} \|\Psi \tilde{\mathbf{x}}\|_0 + \lambda \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2$$

lub inaczej

$$\min_{\mathbf{a}} \|\mathbf{a}\|_0 + \lambda \|\mathbf{x} - \Psi^T \mathbf{a}\|_2^2$$

lub ewentualnie (z wątkiem entropii)

$$\min_{\mathbf{a}} \mathcal{E}\{\|\mathbf{a}\|_0\} + \lambda \|\mathbf{x} - \Psi^T \mathbf{a}\|_2^2$$

- oszczędne próbkowanie

$$\min_{\mathbf{a}} \|\mathbf{a}\|_0 + \lambda \|\mathbf{y} - \Phi \Psi^T \mathbf{a}\|_2^2$$

$$\min_{\mathbf{x}} \|\Psi \mathbf{x}\|_0 + \lambda \|\mathbf{y} - \mathbf{A} \mathbf{x}\|_2^2$$

gdzie $\mathbf{A} = \Phi \Psi$, $\mathbf{A} \in \mathbb{R}^{M \times N}$

- adaptacyjne próbkowanie
 - * tradycyjnie mamy
 $y_i = \langle \mathbf{x}, \phi_i \rangle$, $i = 1, \dots, M$ oraz ϕ_i są ustalone a priori
 - * adaptacyjnie zaś
 $\phi_i = f(\phi_1, y_1, \dots, \phi_{i-1}, y_{i-1})$
- relaksacje problemu minimalizacji (pseudo)normy zerowej

– l_1

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\Psi\mathbf{x}\|_1$$

$$\min_{\mathbf{a}} \|\Phi\mathbf{x} - \Phi\Psi^T\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1$$

– relaksacja całkowitej wariancji (TV)

$$\min_{\mathbf{x}} \sum_{i,j} |(\nabla x)_{i,j}| \quad \text{p.o.} \quad \|\tilde{\mathbf{y}} - \Phi\mathbf{x}\|_2 \leq \epsilon$$

$$\min_{\mathbf{a}} \|\Delta x\|_1 + \lambda \|\Phi\mathbf{x} - \Phi\tilde{\mathbf{x}}\|_2^2$$

– algorytmy zachłanne

– metody kombinatoryczne

Wykorzystano poszczególne schematy regularyzacji ograniczając liczbę parametrów modelu.

3.8 Oszczędne próbkowanie – *compressive sampling*

Kompresowalna (*compressive*, inaczej kodowalna), czyli podatna na kompresję reprezentacja sygnału nawiązuje do postaci rzadkiej wektora danych opisujących sygnał cyfrowy.

Znaczy to, że sygnał źródłowy o reprezentacji kompresowalnej może zostać przedstawiony albo aproksymowany za pomocą znikomo małej, znacząco zredukowanej w stosunku do wymiaru dziedziny źródłowej liczby niezerowych współczynników (próbek czy ogólnie danych). Stanowią one nową, rzadką reprezentację sygnału, która w pewnych okolicznościach może być bardziej użyteczna w danym zastosowaniu. Dotyczy to szczególnie tych rozwiązań problemu odwrotnego (3.3), w których $M \ll N$, a więc akwizycji z małą liczbą pomiarów, problemu superrozdzielczości czy uzupełniania obrazów.

Silnie zredukowany wymiar nowej dziedziny reprezentacji sygnału daje zwykle jedynie przybliżoną postać wersji źródłowej, najlepiej przy zachowaniu wszystkich cech sygnału istotnych dla odbiorcy. W skutecznej aproksymacji sygnałów użyteczna jest ich rzadka reprezentacja uzyskana za pomocą dobranych atomów, odzwierciedlających ich zasadnicze właściwości i cechy. W przypadku sygnałów kompresowalnych możliwe jest takie jego uproszczenie, że zaledwie niewielka liczba wartości go definiujących (tzw. pomiarów¹) w stosunku do liczby N próbek źródłowych (z próbkowania według reguły Nyquista) zawiera całą informację użyteczną. Oznacza to, że zdecydowana większość próbek źródłowych po upakowaniu informacji jest w przybliżeniu równa zeru, czyli sygnał uzyskuje zredukowaną

¹inaczej mezurandów, próbek, współczynników, ogólniej danych

reprezentację w odniesieniu do wymiaru nadmiarowej, źródłowej dziedziny określoności sygnału według reguł tradycyjnych (Nyquista-Shannona). Sygnał kompresowalny może uzyskać formę upakowaną już na etapie akwizycji – według koncepcji *compressive sampling*, czyli kompresowalnego próbkowania (lub może lepiej – selektywnego próbkowania) – bądź też w wyniku przekształcenia źródłowej reprezentacji nadmiarowej metodą selektywnego rozwinięcia w nadmiarowym słowniku atomów.

3.8.1 Próbkowanie z selekcją informacji

Problem sprowadza się do następujących założeń:

- celem jest rejestracja sygnału (obrazu) źródłowego \mathbf{f} , zawierającego N próbek (pikseli) $\mathbf{f} = [f_1, f_2, \dots, f_N]^T$ (założono kolumnowy opis wektorów) za pomocą P pomiarów $\mathbf{y} = [y_1, y_2, \dots, y_P]^T$;
- dane pomiarowe określone są za pomocą macierzy \mathcal{N} charakteryzującej uwarunkowania pomiaru (rejestracji):

$$\mathbf{y} = \mathcal{N}\mathbf{f} + \eta = \mathcal{N}_{[P \times N]}\mathbf{f}_{[N \times 1]} + \eta \quad (3.11)$$

- szukana jest liczba pomiarów P wektora \mathbf{y} , która wystarczy, by określić (zrekonstruować) wiarygodnie \mathbf{f} .

Z algebry liniowej mamy, że $P = N$, a $\mathcal{N} = I_N$ (I_N – macierz jednostkowa o N elementach przekątnej), czyli liczba pomiarów odpowiada docelowej liczbie próbek dyskretnej postaci rejestrowanego sygnału. Klasyczny, równomierny sposób próbkowania sygnałów ciągłych realizowany jest według tego właśnie schematu. Znane są jednak metody nierównomiernego próbkowania, które można opisać schematem tzw. martwych próbek, których nie musimy pobierać w zakresach dziedziny o silniejszych ograniczeniach widma lokalnego (maksimum widma lokalnego jest mniejsze od maksimum globalnego widma sygnału).

Możliwe jest też inne rozwiązanie, wykorzystujące fakt podatności na kompresję (kompresowalność) nadmiarowych sygnałów rzeczywistych rejestrowanych w zdecydowanej większości systemów akwizycji. Stosowane powszechnie transformacyjne kodowanie sygnałów (np. standardy multimedialne z rodziny JPEG oraz MPEG) bazuje na znaczącej redukcji przestrzeni gromadzącej gros energii sygnału poprzez transformacje liniowe, będące rozwinięciem sygnału w bazach ortogonalnych czy biortogonalnych. Dzięki tym przekształceniom otrzymujemy rzadkie reprezentacje sygnałów kosztem znaczącej redukcji ich nadmiarowości (rozumianej przede wszystkim w sensie statystycznym).

Zapis rozwinięcia sygnału (transformacji odwrotnej) jest następujący: $\mathbf{f} = \Phi_{[N \times N]}\mathbf{a}$, gdzie Φ zawiera wektory bazowe danego przekształcenia (jako kolumny), a \mathbf{a} jest (kolumnowym) wektorem N współczynników. Dla przypadku M -

rzadkiej reprezentacji \mathbf{f} w Φ mamy: $\|\mathbf{a}\|_0 = M \ll N$, co pozwala wykorzystać rozwinięcie \mathbf{f} jedynie w M odpowiednich wektorach bazy:

$$\mathbf{f} = \Phi_{[N \times M]} \mathbf{a}_{[M \times 1]} \quad (3.12)$$

Wracając do podstawowego opisu procesu akwizycji (3.11), aby $P < N$, \mathbf{f} musi być kompresowalne względem przekształcenia opisanego przez M . Mając wtedy sytuację jak w 3.12 można zapisać, że

$$\mathbf{y} = \mathcal{M}_{[P \times M]} \mathbf{a}_{[M \times 1]} \quad (3.13)$$

gdzie $\mathcal{M}_{[P \times M]} = \mathcal{N}_{[P \times N]} \Phi_{[N \times M]}$.

Przyjmując sensownie stopień kompresowalności rejestrowanych sygnałów, czyli stosunek $\frac{N}{M}$, można oszacować liczbę losowych pomiarów $M + 1 < P < N$ niezbędnych do wiarygodnej rekonstrukcji \mathbf{f} z dużym prawdopodobieństwem.

Kluczowe pytanie dotyczy sposobu optymalizacji NP-trudnego problemu ustalenia możliwie małej liczby pomiarów dzięki zastosowaniu dobranej do sygnału x transformacji Φ .

Rozdział 4

Wybór klasycznych metod obrazowych

Ogólnie modelem nazywany jest opis (w formie schematu, zależności matematycznych, konstrukcji, rysunku technicznego itp.) ukazujący istotę (proces, działanie, budowę, cechy, zależności) jakiegoś aspektu rzeczywistości (zjawiska, sytuacji, przedmiotu zainteresowania, itp.). Opis ten jest zwykle sparymetryzowany (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 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 ustalonym *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.

4.0.2 Stochastyczne modele obrazów

Stochastyczne modele obrazów są naturalną konsekwencją probabilistycznej koncepcji źródeł informacji (zobacz p. 1.1.2), 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): $f(k, l) \in A_f = \{0, \dots, 255\}$.

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 $i \in [1, \dots, I]$ 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 $[1, \dots, I]$ w Ω_f , tak że $(k, l) = \xi(i)$. 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 \mathbf{F}** z konkretną realizacją – obrazem: $\mathbf{f} = (f_1, f_2, \dots, f_I)$. \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 $e_i = k \in A_e = \{1, \dots, K\}$ (dla uproszczenia tło potraktujemy jako obiekt uzupełniający). Model $\mathbf{E} = \{E_i\}_{i=1}^I$ nazwijmy polem etykiet (pikseli) z wymagającą ustalenia realizacją $\mathbf{e} = (e_1, e_2, \dots, e_I)$.

Probabilistyczny opis pola losowego stanowi łączny rozkład prawdopodobieństwa $P_{\mathbf{F}}(\mathbf{f})$, będący w przypadku dyskretnym funkcją rozkładu prawdopodobieństwa (dyskretną wersją funkcji gęstości prawdopodobieństwa). Jak wspomniano w p. ??, 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.

Histogram Silnie uproszczonym, a przy tym często stosowanym opisem właściwości obrazu jest histogram, czyli zliczenie liczby wystąpień każdej z możliwych wartości piksela. Tworzona na jego bazie funkcja rozkładu prawdopodobieństwa wartości pikseli P_F 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 alfabetcie, stacjonarnego, po normalizacji sumy liczby wystąpień, uzyskujemy dyskretny rozkład prawdopodobieństw poszczególnych symboli alfabetu źródła (funkcję rozkładu prawdopodobieństwa).

Dokładniej, przy alfabetcie wartości pikseli $A_f = \{a_i\}_{i=1}^n : \forall f \in A_f$ można oszacować rozkład prawdopodobieństw wartości pikseli $P_F = \{\Pr(F = a_i)\}_{i=1}^n$,

gdzie $\Pr(F = a_i) = P(a_i) = \frac{n_i}{\eta}$; liczba pikseli obrazu $\eta = I$. Uzyskujemy więc histogramowe przybliżenie rozkładu wartości pikseli $P_f(f)$, które charakteryzuje globalne właściwości obrazu, abstrahując od przestrzennego rozkładu tych wartości w przestrzeni obrazu.

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 (ang. *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 rzędach (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 $k = 1, \dots, K$. Korzystając z prawdopodobieństwa warunkowego mamy:

$$P(f_i, R_k) = P(R_k)P(f_i|k) \quad (4.1)$$

zaś rozkład brzegowy dla $F_i = f_i \in A_f$ daje zależność:

$$P(f_i) = \sum_{k=1}^K P(f_i, k) = \sum_{k=1}^K P(R_k)P(f_i|R_k) \quad (4.2)$$

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 licznosc poszczególnych regionów. Prawdopodobieństwo wartości poszczególnych pikseli

obrazu f_i jest więc konstruowane na podstawie prawdopodobieństw warunkowanych ich przynależnością do poszczególnych regionów R_k :

$$P(f_i) = \sum_{k=1}^K \pi_k P_k(f_i) \quad (4.3)$$

Warunkowe prawdopodobieństwo przynależności piksela do regionu: $P_k(f_i) = P(f_i|R_k)$ poszukiwane jest zwykle w klasie rozkładów opisanych sparometryzowanym, uogólnionym rozkładem Gaussa:

$$P_k(f_i) = g_k(f_i | \mu_k, \sigma_k, \alpha) = g_k(f_i) = \frac{\alpha \beta_k}{2\Gamma(1/\alpha)} \exp[-|\beta_k(f_i - \mu_k)|^\alpha] \quad (4.4)$$

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 (4.4)) daje następującą postać rozkładu:

$$P_k(f_i) = g_k(f_i | \mu_k, \sigma_k, \alpha = 2) = g_k^{(2)}(f_i) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(f_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (4.5)$$

Wykorzystując prawdopodobieństwo wartości piksela (4.3) z doprecyzowanym rozkładem prawdopodobieństw przynależności do regionów według (4.4) 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 [118].

Funkcja rozkładu podobieństwa w modelu FMM z rozkładem gaussowskim, tj. modelu GMM (*Gaussian Mixture Model*) ma więc postać ogólną:

$$P(\mathbf{f}) = \prod_{i=1}^I \sum_{k=1}^K \pi_k g_k(f_i) \quad (4.6)$$

Model ten pozwala wygenerować (syntetyzować) przybliżony obraz $\tilde{\mathbf{f}} = \mathcal{G}(P(\mathbf{f})) \approx \mathbf{f}$. Oznaczmy $\tilde{\mathbf{f}} = \varkappa$.

Taka konstrukcja modelu nie zakłada spójności regionów, które należy według (4.2) 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 (ang. *independent and identically distributed*) wewnątrz K regionów (obiektów) $R_k, k = 1, \dots, K$. 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 [119, 120].

Aby określić model (4.6) dla konkretnego \mathbf{f} , należy ustalić wartości następującego zbioru parametrów:

- globalnych (inaczej strukturalnych): $\mathcal{P}_s = \{K, \alpha\}$
- lokalnych, charakteryzujących regiony: $\mathcal{P}_l = \{\mu_k, \sigma_k, \pi_k \mid k = 1, \dots, K\}$

Zasadniczym kryterium optymalizacji jest maksymalizacja funkcji wiarygodności modelu. Alternatywnie, w celu uproszczenia obliczeń, stosowane jest kryterium minimalizacji względnej entropii (odległości Kullbacka-Leiblera – K-L) pomiędzy znormalizowanymi histogramami obrazu źródłowego dającymi prawdopodobieństwa $P_F(f)$ oraz obrazu wygenerowanego z modelu $P_\varkappa(f)$. Odległość K-L określona jest zależnością:

$$H(P_F||P_\varkappa) = \sum_f P_F(f) \log \frac{P_F(f)}{P_\varkappa(f)} \quad (4.7)$$

Konstrukcja modelu GMM opisującego dany obraz bazuje przede wszystkim na trafnym wyznaczeniu parametrów regionów. Pojawiają się więc kluczowe dla analizy treści obiekty. Na tej podstawie można dokonać segmentacji obrazu opisując informacje obrazową zgodnie z wymogami aplikacji. Ustalenie przynależności pikseli do regionów odbywa się na zasadzie reguły Bayesa, przy czym wyznaczone w ten sposób regiony mają zwykle szereg ograniczeń, wynikających z braku opisu w GMM zależności lokalnych.

Przynależność pikseli do regionów Ogólnie, problem klasyfikacji polega na przewidywaniu (predykcji) nieznaney etykiety $e \in A_e$ na podstawie wektora cech $\mathbf{C}(r)$ klasyfikowanego obiektu r . W GMM klasyfikowanym w sposób niezależny obiektem jest piksel identyfikowany przez indeks i , a jedyną cechą obiektu jest wartość piksela f_i . Klasyfikatorem $d : A_f \rightarrow A_e$ dającym najmniejszy poziom błędu klasyfikacji jest klasyfikator bayesowski.

W przypadku K klas, klasyfikator bayesowski ma postać:

$$d_B(x_i) = \arg \max_k P(k | f_i) = \arg \max_k \pi_k P_k(f_i) \quad (4.8)$$

Maksymalne prawdopodobieństwo a posteriori przynależności do regionu decyduje o ustaleniu etykiety piksela. Reguła klasyfikacji bayesowskiej wykorzystuje oszacowane rozkłady warunkowe a priori $P_k(f_i)$, ukształtowane według wyznaczonych parametrów lokalnych i globalnych, do ustalenia prawdopodobieństw a posteriori $P(k | f_i)$ przynależności piksela do każdego z R_k . Ponieważ $P(f_i | k) = \frac{P(k|f_i)P(f_i)}{P(k)}$, otrzymujemy $P(k | f_i) = \frac{P(f_i|k)P(k)}{P(f_i)}$. Przyjmijmy stosowane wcześniej oznaczenia $P(k) = P(R_k) = \pi_k$, $P(f_i | k) = P_k(f_i)$, uwzględniając (4.3).

Dla każdego piksela o indeksie i , $i = 1, \dots, I$ przypisanego obiektowi k , $k = 1, \dots, K$ mamy predykcję

$$P(k | f_i) = \frac{\pi_k P_k(f_i)}{\sum_{j=1}^K \pi_j P_j(f_i)} \quad (4.9)$$

Aby określić regiony, potrzebne jest więc wstępne ustalenie ich liczności π_k oraz rozkładów prawdopodobieństw *a priori*. Można to ustalić poprzez wstępne pogrupowanie danych metodą bez znajomości modelu, np. k -średnich, na podstawie zbioru danych o jednakowym rozkładzie (generowanych z tego samego źródła, o ustalonym modelu) lub też innej wiedzy dostępnej *a priori*. W przypadku modelowania pojedynczego obrazu wykorzystuje się zwykle wstępną metodę grupowania.

Kolejnym etapem jest wyznaczenie modelu poprzez ustalenie jego parametrów lokalnych i globalnych, przy ustalonej koncepcji mieszaniny regionów oraz ogólnej postaci rozkładów. Metoda klasteryzacji, którą można teraz zastosować, powinna bazować na przyjętej formie modelu jako wiedzy *a priori*. Szacując parametry modelu stosowane jest ogólne kryterium maksymalnej wiarygodności, czyli $\mathcal{P}^* = \arg \max_{\mathcal{P}} P(\mathbf{f} | \text{GMM}, \mathcal{P})$, gdzie $\mathcal{P} = \mathcal{P}_s + \mathcal{P}_l$. Często stosowaną metodą wyznaczania parametrów GMM jest EM (ang. *Expectation - Maximization*).

Ze względu na ograniczające założenia GMM, konieczne jest także dodatkowe uwzględnienie kontekstu przy etykietowaniu pikseli, poprawiające efekt klasteryzacji. Jednym z rozwiązań może być metoda lokalnej modyfikacji parametru π_k – CBRL, korygująca model ze względu na rozkład etykiet w najbliższym sąsiedztwie piksela.

W takim rozwiązaniu na stochastyczny model obrazu składa się sekwencja algorytmów k -średnich – GMM – EM – CBRL.

Wyznaczanie parametrów modelu Algorytm EM służy do wyznaczenia parametrów znanych rozkładów rozwiązując problem klasteryzacji danych (łączenia w obiekty stanowiące istotę modelu). Wstępnie ustalane są początkowe wartości parametrów i zmiennych pomocniczych. Etap przewidywania sprowadza się do szacowania bayesowskich prawdopodobieństw *a posteriori*. Na podstawie ustalonych rozkładów oblicza się skorygowane wartości parametrów \mathcal{P}_l , przy założeniach ergodyczności modelu, zwiększając w procesie iteracyjnym wiarygodność modelu. Stosowanym często kryterium jest odległość K-L według (4.7).

Algorytm EM dla modelu GMM opisanego przez (4.6) wygląda następująco:

1) ustal wartości globalnych parametrów w badanych zakresach $K = K_{\min}, \dots, K_{\max}$,
 $\alpha = \alpha_{\min}, \dots, \alpha_{\max}$

- ustaw licznik iteracji $m = 0$ oraz zbiór początkowych wartości parametrów modelu $\mathcal{P}^{(0)} = \{\pi_k^{(0)}, \mu_k^{(0)}, \sigma_k^{(0)}, k = 1, \dots, K\}$ zgodnie z efektami wstępnej klasteryzacji (np. metodą k -średnich)
- etap E: dla kolejnych $i = 1, \dots, I$ oraz $k = 1, \dots, K$ oszacować przybliżenia prawdopodobieństw *a posteriori*

$$z_{i,k}^{(m)} = \frac{\pi_k^{(m)} P_k(f_i)}{\sum_{k=1}^K \pi_k^{(m)} P_k(f_i)} \quad (4.10)$$

- etap M: dla kolejnych $k = 1, \dots, K$ uaktualnić estymaty parametrów

$$\begin{cases} \pi_k^{(m+1)} = \frac{1}{I} \sum_{i=1}^I z_{i,k}^{(m)} \\ \mu_k^{(m+1)} = \frac{1}{I \pi_k^{(m+1)}} \sum_{i=1}^I z_{i,k}^{(m)} f_i \\ \sigma_k^{2(m+1)} = \frac{1}{I \pi_k^{(m+1)}} \sum_{i=1}^I z_{i,k}^{(m)} (f_i - \mu_k^{(m+1)})^2 \end{cases} \quad (4.11)$$

- jeśli $|H^{(m)}(P_F || P_{\mathcal{Z}}) - H^{(m+1)}(P_F || P_{\mathcal{Z}})| > \epsilon$, inkrementuj licznik $m = m + 1$ i idź do etapu E; w p.p. zachowaj ustalony \mathcal{P} .

2) zachowaj końcową wartość $H^{K,\alpha}(P_f || P_{\mathcal{Z}})$ i idź do punktu 1).

Modelowanie kontekstu Aby uczynić GMM modelem bardziej wiarygodnym, z wierniejszym opisem spójnych regionów, należy wykorzystać dostępną w \mathbf{f} informację kontekstową.

Jedną z możliwych form uzupełnienia modelu o opis lokalnych zależności danych jest koncepcja CBRL (ang. *Contextual Bayesian Relaxation Labeling*), zaproponowana przez [] i wykorzystana m.in. przez []. Niech \mathbf{c}_i oznacza najbliższe sąsiedztwo piksela i , będące np. blokiem o rozmiarach $m \times m$ o środku w i . \mathbf{c}_i

jest wykorzystywane do lokalnej korekcji wskaźnika π_k , czyniąc z niego estymatę normalizowanej liczności danego regionu w najbliższym sąsiedztwie piksela i :

$$\pi_k^{(i)} = p(e_i | \mathbf{e}_{\mathbf{c}_i}) = \frac{1}{m^2 - 1} \sum_{j \in \mathbf{c}_i, j \neq i} \mathcal{I}(k, e_j) \quad (4.12)$$

gdzie e_i jest etykietą piksela i , a $\mathbf{e}_{\mathbf{c}_i}$ to zbiór etykiet występujących w \mathbf{c}_i ; $\mathcal{I}(\cdot)$ to funkcja wskazująca:

$$\mathcal{I}(k, l_j) = \begin{cases} 1 & \text{jeśli } e_j = k \\ 0 & \text{w p.p.} \end{cases}$$

Algorytm CBRL wygląda więc następująco:

- 1) ustaw licznik iteracji $m = 0$ dla początkowej klasyfikacji pikseli do regionów modelu GMM-EM, dającej rozkład etykiet $\mathbf{e}^{(0)}$
- 2) uaktualnij etykiety pikseli
 - losowo uporządkuj wszystkie piksele o indeksach źródłowych $i = 1, \dots, I$
 - ustal etykietę kolejnego, zgodnie z losowym uporządkowaniem, piksela według zależności

$$e_j^{(m)} = \arg \max_k \left(\pi_k^{(i)} \right)^{(m)} P_k(f_i) \quad (4.13)$$

zgodnie z modyfikacją $\pi_k^{i(m)}$ według (4.12).

- 3) Jeśli mniej niż 1% pikseli zmieniło swoją etykietę zatrzymaj; w p.p. inkrementuj $m = m + 1$ i idź do punktu 2).

Warto podkreślić, że model GMM – CBRL uwzględnia sąsiedztwo \mathbf{c}_i na etapie precyzowania przynależności i do regionu, czyli aktualizacji e_i . Tworzona mapa regionów będąca interpretacją modelowanej realizacji pola etykiet \mathbf{e} powstaje więc z wykorzystaniem lokalnych zależności. Jest obiektywą korektą modelu obrazu w ramach rozkładów pola \mathbf{E} . Pośrednio wpływa to na wiarygodność rozkładu pola wartości \mathbf{E} , bazujące na niezależnym szacowaniu prawdopodobieństw pikseli przypisanych regionom. Ta koncepcja kontekstowego pola etykiet i niezależnego pola wartości została bardziej spójnie zrealizowana w modelu Markowa obrazu.

Model Markowa regionów obrazu

Na bazie pola losowego Markowa konstruowane są modele obrazu, wykorzystujące przede wszystkim rozkłady brzegowe do konstrukcji rozwiązań pragmatycznie uproszczonych, ale jednocześnie uwzględniających lokalne zależności danych opisywanych obiektywo.

Pole losowe Markowa Bardziej wiarygodnym, szczególnie przydatnym i powszechnie stosowanym stochastycznym modelem obrazu jest pole losowe Markowa – MRF (ang. Markov Random Field). Kontekstowy model źródła informacji obrazowej, uwzględniający lokalne (tj. w skończonym, najbliższym sąsiedztwie) zależności danych w przestrzeni obrazu, dostosowuje ogólną koncepcję przedstawioną w p. 1.3.1 do sąsiedztwa przestrzennego w sensie wielowymiarowym.

Określmy pole losowe etykiet \mathbf{E} jako pole Markowa. W tym celu wykorzystajmy zbiór kontekstów kolejnych pikseli $\mathbf{c}^I = \{\mathbf{c}_1, \dots, \mathbf{c}_I\}$, gdzie $\mathbf{c}_i \in A_C = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_L\}$ jest najbliższym sąsiedztwem wystąpienia piksela i w przestrzeni obrazu – $i \notin \mathbf{c}_i$ oraz $i \in \mathbf{c}_j \Leftrightarrow j \in \mathbf{c}_i$. A_C jest alfabetem możliwych kontekstów, $|A_C| = L$.

Pole losowe E na zbiorze pikseli o indeksach $i \in D$, gdzie $D = \{1, \dots, I\}$, oraz zdefiniowanym sąsiedztwie \mathbf{c}^I , jest MRF wtedy i tylko wtedy, jeśli dla każdej realizacji \mathbf{e} tego pola zachodzi $P(\mathbf{e}) > 0$ oraz $P(e_i | \mathbf{e}_{D-\{i\}}) = P(e_i | \mathbf{e}_{\mathbf{c}_i})$

Probabilistyczny opis MRF bazuje na dużej liczbie prawdopodobieństw warunkowych, zależnie od rozmiaru kontekstu i przy realnym modelowaniu obrazów pojawia się problem wiarygodnego statystycznie oszacowania tych prawdopodobieństw (rozrzedzenia kontekstu – zobacz p. ??).

Równoważnie, pole Markowa można scharakteryzować za pomocą rozkładu Gibbsa, gdzie

$$P(\mathbf{e}) = \frac{1}{Z} \exp(-U(\mathbf{e})) \quad (4.14)$$

Z jest normalizującą stałą zwaną funkcją podziału (ang. *partition function*), zaś $U(\mathbf{e})$ jest funkcją energii postaci

$$U(\mathbf{e}) = \sum_{\mathbf{c} \in \mathfrak{C}} V_{\mathbf{c}}(\mathbf{e}) \quad (4.15)$$

Funkcja ta dotyczy tzw. klik (*clique*), gdzie jako klikę rozumie się podzbiór pikseli dziedziny D , w której każdą parę różnych pikseli stanowią sąsiedzi, z wyjątkiem klik jednoelementowych. Funkcja $U(\mathbf{e})$ jest sumą potencjału $V_{\mathbf{c}}(\mathbf{e})$ klik \mathbf{c} po zbiorze wszystkich możliwych klik \mathfrak{C} . Wartość $V_{\mathbf{c}}(\mathbf{e})$ zależy od lokalnej konfiguracji klik \mathbf{c} . Przykładowa, jedna z klasycznych postaci $V_{\mathbf{c}}(\mathbf{e})$ (zobacz [121]) wygląda następująco:

$$V_{\mathbf{c}}(\mathbf{e}) = V_{\mathbf{c}}(f_{j'}, f_{j''}) = \begin{cases} -\beta & \text{jeśli } f_{j'} = f_{j''} \\ 0 & \text{w p.p.} \end{cases} \quad (4.16)$$

Model GMM – MRF

Model ukrytego pola losowego Markowa Ukryty model Markowa – HMM (ang. *Hidden Markov Model*) jest definiowany jako łańcuch Markowa o sekwencji stanów niedostępnej w bezpośredniej obserwacji. Innymi słowy, stan w jakim znajduje się źródło nie jest utożsamiany z dostarczaną przez nie obserwacją. Każda obserwacja

jest traktowana jako funkcja losowa sekwencji stanów źródła poprzedzających jej dostarczenie. Liczba stanów źródła nad określonym alfabetem $A_{SS} = \{s_1, \dots, s_t\}$ jest równa liczbie obserwacji, zaś łańcuch Markowa opisujący źródło zmienia stan zgodnie z macierzą prawdopodobieństwa przejść pomiędzy stanami o rozmiarze $t \times t$.

HMM daje więc możliwość bardziej wiarygodnego opisu złożonych zjawisk, takich jak obrazy, dając dodatkowy stopień swobody – przestrzeń stanów dopasowywaną do generowanych obserwacji źródła.

Przy rozszerzeniu tego modelu na obrazy, łańcuch (ograniczony do jednego wymiaru) zastępowany jest polem Markowa [118], dając model ukrytego pola losowego Markowa – HMRFM (ang. *Hidden Markov Random Field Model*). Obserwowanym zjawiskiem jest generowane przez źródło pole wartości \mathbf{f} , zaś pole etykiet \mathbf{e} opisuje niejawne stany źródła, w których następuje generacja informacji obrazowej. Regiony ustalone na podstawie obserwowanych wartości pikseli wyjaśniają ukrytą, tj. jedynie w pewnym stopniu domniemaną na podstawie obserwacji, formalnie poszukiwaną naturę obrazów. Istnieje więc związek przyczynowo-skutkowy w opisie źródeł informacji, opisany rozkładem $p(f_i | e_i)$, będącym funkcją zarówno f_i , jak i parametrów \mathcal{P}_{e_i} , wykorzystanych w opisie prawdopodobieństw *a priori* w (4.5).

Dla dowolnego \mathbf{e} , łączne prawdopodobieństwo pary pól losowych (\mathbf{F}, \mathbf{E}) można zapisać jako $P(\mathbf{f}, \mathbf{e}) = P(\mathbf{e})P(\mathbf{f} | \mathbf{e})$ (analogicznie do (4.1)). Zakładając tym razem warunkową niezależność pola wartości pikseli mamy:

$$P(\mathbf{f}, \mathbf{e}) = P(\mathbf{e})P(\mathbf{f} | \mathbf{e}) = P(\mathbf{e}) \prod P(f_i | e_i) \quad (4.17)$$

Zgodnie z modelem Markowa pola etykiet mamy zależność $P(e_i | \mathbf{e}_{\mathbf{c}_i})$. Tę zależność od najbliższego sąsiedztwa uwzględnijmy poprzez szacowanie prawdopodobieństwa wystąpienia (f_i, e_i) uwarunkowanego określonym kontekstem:

$$P(f_i, e_i | \mathbf{e}_{\mathbf{c}_i}) = P(f_i | e_i)P(e_i | \mathbf{e}_{\mathbf{c}_i}) \quad (4.18)$$

W przekształceniu tym wykorzystano niezależność realizacji pola wartości f_i od regionów ich przynależności.

Brzegowy rozkład prawdopodobieństwa

$$P(f_i | \mathbf{e}_{\mathbf{c}_i}) = \sum_{k \in A_e} P(f_i, k | \mathbf{e}_{\mathbf{c}_i}) = \sum_{k \in A_e} P(f_i)P(k | \mathbf{e}_{\mathbf{c}_i}) \quad (4.19)$$

Model HMRFM według (4.19) uwzględnia więc niezależne prawdopodobieństwa *a priori*, jak też lokalne zależności pola etykiet, opisane warunkowymi prawdopodobieństwami najbliższego kontekstu. Model taki jest niewątpliwie bardziej realistyczny i, jak się okazuje, znacząco bardziej przydatny w zastosowaniach [118, ?].

4.0.3 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.

Skuteczność modelu zapewnia możliwie silne podobieństwo prymitywów do kształtu dominującego w opisie treści obrazu. Możliwa adaptacja formy podstawowej do charakteru modelowanej treści wymaga jednak trafnego kryterium, szybkiego algorytmu dopasowania oraz zadawalającej stacjonarności form treściowych.

4.0.4 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 rozpoznawania wzorców, często dopasowywane do znanych *a priori* modeli, a następnie ustalane są wzajemne relacje pomiędzy obiektami obrazu.

Model obiektów są różne, od siatek prostych prymitywów geometrycznych pokrywających obraz (siatka trójkątów, prostokątów, itp.), po złożone obiekty opisane gładkim konturem i charakterystyką tekstury.

4.0.5 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.

4.0.6 Wieloskalowe modele treści obrazów

Okazuje się, geometryczny opis obiektów może być wsparty bardziej elastycznymi możliwościami geometrii analitycznej oraz uzupełniony potencjałem analizy funkcjonalnej. Kluczem jest zapewnienie skalowalności modelu obiektów, a funkcjonalny schemat analizy wieloskalowej daje podstawy konstrukcji praktycznie nieograniczonej liczby przybliżeń realnych obiektów. Wydaje się, że takie narzędzie ma wystarczający potencjał opisu bardzo silnie zróżnicowanej klasy zmian, będącej przedmiotem interpretacji informacji obrazowej. Dlatego też stale jest wykorzystywane w różnorodnych algorytmach obróbki obrazów.

Najczęściej stosowanym narzędziem są tutaj bazy falkowe, konstruowane tensorowo w przestrzeni obrazu, jak też z jądrem dwuwymiarowych – falki kierunkowe (np. *contourlets* i *curvelets*) i geometryczne (np. *wedgelets*).

Zasadniczym celem stosowania metod analizy wielorozdzielczej jest możliwość silna separacja sygnału użytecznego i szumu, przy czym oba te pojęcia rozumie się zwykle w aspekcie semantycznym. Sygnałem jest struktura, zbiór cech, region lub rozkład krawędzi, które reprezentują informację (tj. wpływają na proces interpretacji obrazu przez odbiorcę – użytkownika). Szumem więc jest to wszystko, co się nie liczy, nie ma znaczenia w ocenie i interpretacji.

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.

Poszukiwanie reprezentacji obrazów, w której ukryta/subtelna informacja diagnostyczna może zostać wydobyta w procesie przybliżania, szacowania czy estymowania było i jest celem wielu badań z obszaru analizy obrazów medycznych. W wielu przypadkach poszukiwana zmiana ma charakter lokalnego zaburzenia, nie-

typowej charakterystyki tkanki rozciągającej się w niewielkim obszarze dziedziny pokrytej dominującą właściwością obrazowanego obiektu. Energia opisująca tę informację, nazwijmy ją funkcją celu, ma zwykle znikomy udział w energetycznym widmie całego sygnału, a specyfika częstotliwościowa ginie w szerokim widmie opisującym obraz. W wielu przypadkach nie sposób też wychwycić tej zmiany w probabilistycznym modelu zależności czy korelacji lokalnych, wyznaczanym na podstawie znaczącej liczby powtórzeń określonej sytuacji w danym kontekście.

Przy większych zmianach analiza cech teksturowych lub też dających się ustalić zarysów może wystarczyć do detekcji zmian podejrzanych, jednak zwykle dzieje się tak w przypadku, gdy możliwe jest ich rozpoznanie za pomocą dostrzegalnych cech wizualnych. Okazuje się jednak, że w tej innej, poszukiwanej dziedzinie zmiany te oraz ich cechy mogą być zdecydowanie bardziej uwydatnione.

Bardzo korzystną formą tej nowej dziedziny okazały się przeciwdziedziny przekształceń wieloskalowych, pozwalające analizować lokalne cechy sygnałów z różną rozdzielczością przybliżeń. Bez tych przekształceń definiują postać szczególnie dobrze opisywalnych obiektów lokalnych. Pozwalają je opisać w sposób zdecydowanie bardziej wyrazisty, poprzez koncentrację energii nowej dziedziny właśnie wokół tych obiektów. Pozostaje jedynie dobrać postać bazy do lokalnych zaburzeń, które są ukryte w dziedzinie źródłowej, by wydobyć przypisaną im poszukiwaną informację diagnostyczną.

Przedmiotem opisywanych w tym rozdziale mechanizmów reprezentacji sygnału w wielu skalach jest więc dobieranie bazy przekształceń do wzorców interpretowanych zmian celem wydobywania zmian ukrytych. A że wzorce te w przypadku wyzwania diagnostyki obrazowej z reguły nie istnieją, 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 pacjenta, systemu obrazowania, a także konkretnej formy manifestacji patologii danego przypadku.

Do zbioru wieloskalowych reprezentacji obrazów, czy ogólnie sygnałów zaliczyć 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 $2W$, z zachowaniem warunku ortogonalności bazy lub też przy słabszych warunkach, aż do ograniczenia jedynie liniową niezależnością wektorów bazy, dają niemal nieograniczoną swobodę modelowania poszukiwanych wzorców patologii. Dobór odpowiedniej bazy dokonywany jest mniej lub bardziej formalnie, w konwencji *a priori* lub *a posteriori*, z możliwością dostosowania do niekiedy bardzo wysublimowanych, złożonych opisów poszukiwanego cech sygnału.

Zasadniczym celem stosowania metod analizy wielorozdzielczej jest możliwie silna separacja sygnału użytecznego i szumu, przy czym oba te pojęcia rozumie

się zwykle w aspekcie semantycznym. Sygnałem jest struktura, zbiór cech, region lub rozkład krawędzi, które mają znacznie diagnostyczne (tj. wpływają na proces interpretacji diagnostycznej obrazu). Szumem więc jest to wszystko co się nie liczy, nie ma najmniejszego znaczenia w diagnozie.

4.1 Komputerowe przetwarzanie obrazów

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 użytkownika.

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 (obiektywnych)
 - modeli probabilistycznych
 - modeli fizycznych i pseudofizycznych (empirycznych)
- wyszukiwanie poindeksowanych danych podobnych

Zagadnienia te są ograniczone z jednej strony procesem rejestracji sygnałów (danych) naturalnych bądź specjalistycznych, z drugiej zaś strony – charakterystyką użytkownika zawartych w sygnale (w danych) informacji. Pomiedzy 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 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

- wydzielenie 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;
 - wydzielenie 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 [?];
 - 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 możliwie efektywnej prezentacji (odsluchu, 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;

4.1.1 Operacje wykonywane na obrazach

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-objekty, 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 \rightarrow **opis semantyczny**²

² (na poziomie abstrakcji użytkownika).

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, odsumiającej, 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

$$g(k, l) = \kappa \cdot f(k, l) + \beta \quad (4.20)$$

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

$$g(k, l) = F\{f(k, l)\} \quad (4.21)$$

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 też stosowana w fotografii solaryzacja

$$g(k, l) = \begin{cases} 2 \cdot f(k, l) & \text{dla } 0 \leq f(k, l) \leq 0,5 \\ 2(1 - f(k, l)) & \text{dla } 0,5 < f(k, l) \leq 1 \end{cases} \quad (4.22)$$

- 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

$$g(k, l) = \begin{cases} 0 & \text{dla } 0 \leq f(k, l) \leq d \\ \frac{1}{g-d}(f(k, l) - d) & \text{dla } d < f(k, l) < g \\ 1 & \text{dla } g \leq f(k, l) \leq 1 \end{cases} \quad (4.23)$$

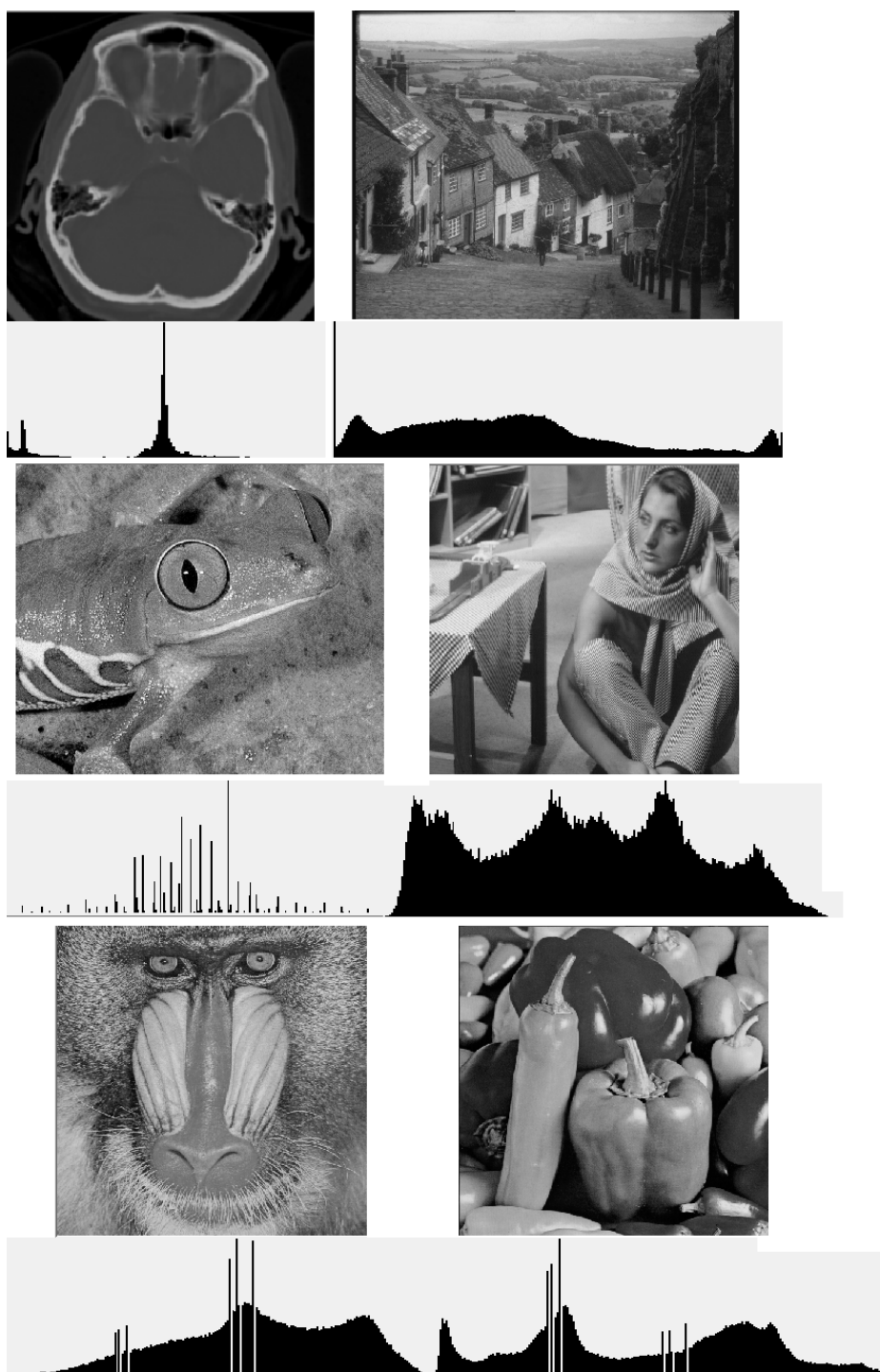
Operacje histogramowe

Histogram obrazu jest graficzną reprezentacją rozkładu wartości pikseli (inaczej wartości funkcji jasności lub poziomów jasności) w obrazie. Każda z wartości pikseli $f(k, l) \in \mathbf{f}$ należy do uporządkowanego rosnąco zbioru (alfabetu) wartości możliwych: $f(k, l) \in A_f = \{a_0, \dots, a_{M-1}\}$. Liczba wystąpień kolejnych poziomów jasności $h(m)$ dla $m = 0, \dots, M - 1$, takich że

$$h(m) = \#\{f(k, l) | f(k, l) = a_m\} \quad (4.24)$$

stanowi histogram prezentowany w określonej formie graficznej (zwykle wykresu słupkowego, niekiedy w postaci tablicy par wartości $(a_m, h(m))$).

Histogram charakteryzuje globalne skonstrastowanie 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 zbytnej 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.



Rysunek 4.1: Przykładowe histogramy wybranych obrazów naturalnych, o zróżnicowanych walorach kontrastu globalnego (pod każdym z obrazów umieszczono jego histogram w zakresie wartości 0–255).

Przykładowe histogramy naturalnych obrazów testowych, świadczące o ich globalnym kontraście przedstawiono na rys. 4.1.

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 wysyczonej 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 $[0, a_{M-1}]$ oznacza przeliczenie wartości pikseli według prostej formuły:

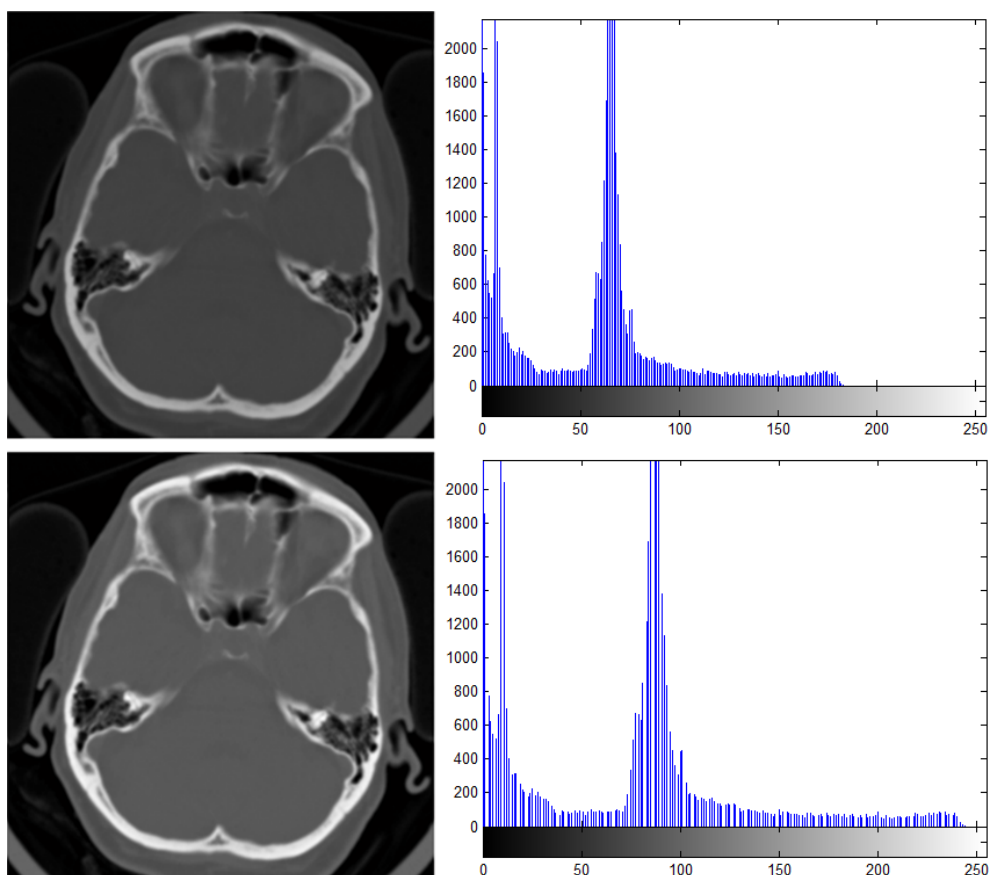
$$\bar{f}(k, l) = \frac{f(k, l) - \min_{\mathbf{f}}\{f(k, l)\}}{\max_{\mathbf{f}}\{f(k, l)\} - \min_{\mathbf{f}}\{f(k, l)\}} \cdot a_{M-1} \quad (4.25)$$

powodując pełne wykorzystanie dynamiki dopuszczonej formatem źródłowym, a opisaną alfabetem $A_{\mathbf{f}}$. Przykładowy efekt rozjaśnienia obrazu poprzez rozciąganie histogramu ukazano na rys. 4.2.

Wyrównywanie tudzież przekształcanie 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: $H_f = \sum_{m=0}^{M-1} h(m)$, służąc jako przybliżenie prawdopodobieństw wystąpienia kolejnych wartości alfabetu: $p(m) = h(m)/H_f$, 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 $A_f = \{a_0, \dots, a_{M-1}\}$ i $A_g = \{b_0, \dots, b_{N-1}\}$.



Rysunek 4.2: 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.

Precyzyjniej, skumulowany histogram estymuje skumulowane prawdopodobieństwo (inaczej dyskretną dystrybuantę) dla rozkładu prawdopodobieństw poziomów jasności postaci:

$$P(f \leq a_m) = P_f(a_m) = \sum_{k=0}^m p_f(k) \quad (4.26)$$

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

$$\sum_{l=0}^n p_g(b_l) \cong \sum_{k=0}^m p_f(a_k) \quad (4.27)$$

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

$$P_g(b_n) = P_f(a_m) \quad (4.28)$$

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 (4.28): $\int_0^{b_n} p_g(g)dg = \int_0^{a_m} p_f(f)df$, 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 $P(g \leq 1) = 1$, a $P(g \leq b_n) = P_g(b_n) = b_n$. Na podstawie (4.28) można wtedy zapisać, że znormalizowane

$$b_n = S(a_m) = P_f(a_m). \quad (4.29)$$

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 [?])

$$P_g(g) = \alpha \exp\{-\alpha(g - b_0)\} \quad (4.30)$$

wygląda następująco:

$$g = b_0 - 1/\alpha \ln[1 - P_f(f)] \quad (4.31)$$

zaś dla rozkładu Rayleigha

$$P_g(g) = \frac{g - b_0}{\alpha^2} \exp\left\{-\frac{(g - b_0)^2}{2\alpha^2}\right\} \quad (4.32)$$

mamy

$$g = b_0 + \left[2\alpha^2 \ln\left(\frac{1}{1 - P_f(f)}\right)\right]^{1/2} \quad (4.33)$$

W przypadku rzeczywistych histogramów dyskretnych stosowanie zależności analitycznych jest rzadko przydatne. Korzystając z ustaleń (4.29) można jednak zaproponować prosty algorytm wyrównywania histogramu – algorytm 4.1.

Algorytm 4.1 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

$$b_n = \left[P_f(a_m) \cdot b_{N-1} \right] \quad (4.34)$$

gdzie operator $[]$ oznacza przybliżenie do najbliższego symbolu alfabetu A_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 (4.34). Zakończ.

□

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. 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

$$f * h(t) = \int_{D_h} f(t-x)h(x) dx \quad (4.35)$$

gdzie D_h jest dziedziną określoności filtru (poza tą dziedziną odpowiedź impulsowa filtru jest zerowa).

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 próbkowanej 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 $N \ll |\Omega_f|$, postać odpowiedzi impulsowej filtru.

Operację filtracji liniowej przedstawia więc zależność

$$f * h(k) = \sum_n f(k-n)h_n \quad (4.36)$$

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

$$g(k, l) = f * h(k, l) = \sum_{(m,n) \in W_h} f(i-m, j-n)h_{m,n} \quad (4.37)$$

Filtr można zdefiniować za pomocą macierzy współczynników wyrażających dyskretną odpowiedź impulsową filtru, o rozmiarze $(2R+1) \times (2R+1)$, indeksowanej 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 wolnozmiennnej, do ekstrakcji szczegółów, gradientowe – do detekcji krawędzi i konturów;
- filtry pasmowo-przepustowe, wykorzystujące operatory laplasjanowe, maskowanie nieostrości (*unsharp masking*) czy filtracje kierunkową - do uwytklania określonych cech, podkreślania krawędzi, uwidocznienia tekstury, wyostrenia itp.

Filtry definiowane są przez ich maski, przy czym istotnych jest kilka wskazówek:

- wynik splotowych obliczeń według (4.37) 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

$$\rho = \frac{1}{\sum_{(m,n) \in W_h} h_{m,n}} \quad (4.38)$$

filtru odsumiające 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 $k = 1, \dots, K, l = 1, \dots, L$):
 - 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 odsumianie, czyli usuwanie bądź redukcja szumu maskującego użyteczną treść obrazową. Ze względu na losowy, nieobciążony charakter szumu najprostszą operacją odsumienia 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:

$$\begin{aligned}
 \text{uśredniający, równomierny: } h_{m,n} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \rho = 1/9 \\
 \text{lekkó nierównomierny: } h_{m,n} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \rho = 1/10 \\
 \text{parametryczny ogólny: } h_{m,n} &= \begin{bmatrix} 1 & p & 1 \\ p & p^2 & p \\ 1 & p & 1 \end{bmatrix}, \rho = \left(\frac{1}{p+2}\right)^2 \\
 \text{parametryczny z } p = 2: h_{m,n} &= \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \rho = 1/16 \\
 \text{gaussowski z } \sigma = 1: h_{m,n} &= \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}, \rho = 1/273
 \end{aligned} \tag{4.39}$$

Filtracje częstotliwościowe

Alternatywnym do przestrzennego sposobem filtracji są metody częstotliwościowe, co wynika przede wszystkim z bardzo skutecznej formy analizy sygnałów za pomocą reprezentacji częstotliwościowej. Wykorzystują one transformacje Fouriera sygnałów cyfrowych oraz filtrów o dobieranej postaci odpowiedzi impulsowej. W podrozdziale 1.2.2, strona 14, zdefiniowano trygonometryczne szeregi fourierowskie (1.1) 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 (1.2))

$$f(t) = \sum_{k \in \mathbb{Z}} b_k e^{j k \omega_0 t} \tag{4.40}$$

gdzie $b_i = 1/T \int_{-T/2}^{T/2} f(t) e^{-j k \omega_0 t} dt$.

Transformacja Fouriera jest rozszerzeniem koncepcji wykładniczych szeregów Fouriera na klasę wszystkich sygnałów $f \in L^2(\mathbb{R})$, w tym nieokresowych. Jest więc to przekształcenie $\mathcal{F} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ 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 $T \rightarrow \infty$, gdzie $\omega_0 \rightarrow 0$, co prowadzi do modelowego ucięcia częstotliwości $k\omega_0 \rightarrow \omega$. Stąd zamiast sumy w (4.40) należy zapisać całkę, co prowadzi do ciągłego zbioru współczynników $F(\omega)$, takich że

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos \omega t dt + j \int_{-\infty}^{\infty} f(t) \sin \omega t dt = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (4.41)$$

Wtedy sygnał reprezentowany jest jako

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (4.42)$$

Wyrażenie (4.41) jest **transformacją Fouriera** dająca ciągle widmo częstotliwościowe sygnału, zaś (4.42) jest odwrotnym przekształceniem Fouriera.

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 $f(t) \in L_T^2$ transformacja przyjmuje postać (analogicznie do (4.40))

$$F(u) = 1/T \int_{-T/2}^{T/2} f(t) e^{-ju \frac{2\pi}{T} t} dt \quad (4.43)$$

gdzie $F(u) \in l^2(\mathbb{Z})$, zaś odwrotna

$$f(t) = \sum_{u \in \mathbb{Z}} F(u) e^{ju \frac{2\pi}{T} t} \quad (4.44)$$

W przypadku sygnałów dyskretnych $f(k) \in l^2(\mathbb{Z})$, rozumianych analogicznie jako okresowe w sensie przejścia granicznego, transformacja jest iloczynem skalarnym sygnału z bazą funkcji sinusoidalnych

$$F(\omega) = \sum_{k \in \mathbb{Z}} f(k) e^{-jk\omega} \quad (4.45)$$

gdzie $F(\omega) \in L^2([-\pi, \pi])$, zaś odwrotna

$$f(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega) e^{jk\omega} d\omega \quad (4.46)$$

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

$$F(u) = \sum_{k=0}^{K-1} f(k) e^{-\frac{j2\pi ku}{K}}, \quad u = 0, \dots, K-1 \quad (4.47)$$

i odwrotnej

$$f(k) = \frac{1}{K} \sum_{u=0}^{K-1} F(u) e^{\frac{j2\pi ku}{K}}, \quad k = 0, \dots, K-1 \quad (4.48)$$

Zakładamy przy tym cykliczne rozszerzenie $f(k)$ z okresem K . Wyrażenia (4.47) oraz (4.48) 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 (4.47) przyjmuje ona postać

$$F(u, v) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} f(k, l) e^{-\frac{j2\pi ku}{K}} e^{-\frac{j2\pi lv}{L}} \quad (4.49)$$

gdzie $u = 0, \dots, K-1$, $v = 0, \dots, L-1$. 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* [?]) 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 (1.4) w punkcie 1.2.2). 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). 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 $\{h_n\}_{n=0}^{N-1}$, także że

$$H(u) = \sum_{n=0}^{N-1} h_n e^{-j\frac{2\pi nu}{N}} \quad (4.50)$$

gdzie $H(u)$ nazywana jest funkcją przenoszenia (inaczej transmitancją) filtru. Analogicznie do (4.49) określana jest dwuwymiarowa transmitancja filtrów stosowanych do przetwarzania obrazów.

Dyskretna postać funkcji przenoszenia 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 (4.45))

$$H(\omega) = \sum_{n \in \mathbb{Z}} h_n e^{-jn\omega} \quad (4.51)$$

a w przypadku dwuwymiarowym

$$H(\omega_x, \omega_y) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} h_{m,n} e^{-jm\omega_x} e^{-jn\omega_y} \quad (4.52)$$

Liniowej filtracji splotowej $g(k, l) = f * h(k, l)$, realizowanej w przestrzeni obrazu za pomocą filtru $h_{m,n}$ według (4.37), odpowiada (jest równoważna - zgodnie z twierdzeniem o splotcie $\mathcal{F}\{f * h\} = \mathcal{F}\{f\} * \mathcal{F}\{h\}$ [?]) filtracja realizowana w fourierowskiej dziedzinie częstotliwościowej za pomocą iloczynu transformaty Fouriera sygnału i transmitancji filtru

$$G(u, v) = F(u, v) \cdot H(u, v) \quad (4.53)$$

Stąd obraz przetworzony

$$g(k, l) = \mathcal{F}^{-1}\{F(u, v) \cdot H(u, v)\} \quad (4.54)$$

Przykładowe postacie charakterystyk częstotliwościowych filtrów stosowanych w przetwarzaniu obrazów filtrów to

- uśredniający (dolnoprzepustowy) filtr Butterwortha

$$G_B(d) = \frac{1}{1 + (d/D_0)^{2n}} \quad (4.55)$$

- górnoprzepustowy filtr Butterwortha

$$G_e(d) = \frac{1}{1 + (D_0/d)^{2n}} \quad (4.56)$$

- uśredniający filtr wykładniczy

$$H_B(d) = \exp -(d/D_0)^n \quad (4.57)$$

- górnoprzepustowy filtr wykładniczy

$$H_e(d) = \exp -(D_0/d)^n \quad (4.58)$$

Niewątpliwie komputerowe wspomaganie diagnozy medycznej, jako szczególnego przypadku odpowiedzialnego użytkownika obrazów, jest zadaniem ambitnym, fascynującym naukowo, o dużym znaczeniu praktycznym, pożytecznym społecznie, o silnym przesłaniu humanitarnym, pełnym ważkich wyzwań i dużych oczekiwań. Doskonalenie metod wymaga dużej odpowiedzialności i ludzkiej wrażliwości, a wiarygodna weryfikacja opracowanych narzędzi nabiera szczególnego znaczenia. Zagadnienie CAD obejmuje całe spektrum zagadnień istotnych także w szeregu innych zastosowaniach komputerowego wspomaganie, bazujących na informacji obrazowej, np. rozpoznawania czy identyfikowania twarzy, odczytywania zdjęć lotniczych czy satelitarnych, zastosowań wiedzenia maszynowego, monitoringu, śledzenia różnego typu obiektów, detekcji nietypowych ludzkich zachowań czy zagrożeń ruchu drogowego na podstawie rejestrowanego zapisu wideo.

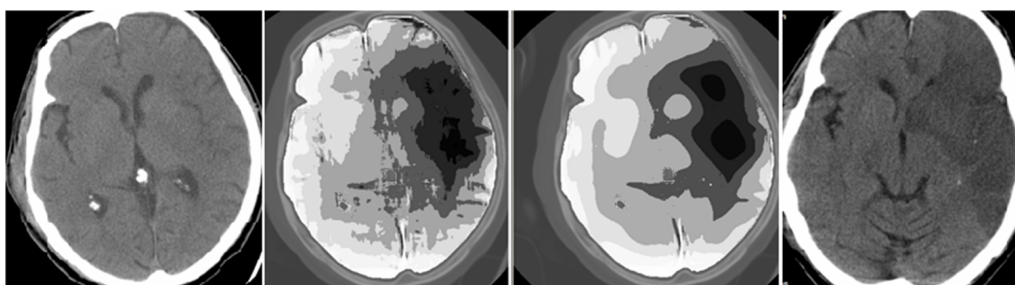
W warstwie metodologii tworzenia systemu komputerowego wspomaganie użytkownika obrazów jest bardzo dobrym, realnym przykładem najbardziej wymagających wyznań konstrukcyjnych, a przy tym różnorodnych, przydatnym w uogólnieniu problemu na szerszą kategorię komputerowego wspomaganie użytkownika obrazów, nazwanego skrótowo **CAIU** (*Computer-Aided Image Use* – propozycja własna). Wiele opisanych wyżej spostrzeżeń można więc uogólnić, redefiniując specyficzne aspekty wybranych zastosowań.

W poszukiwaniu efektywnych rozwiązań CAIU należy w pierwszej kolejności zwrócić uwagę na paradoks natury ogólnej: nie sposób komputerowo doścignąć ludzkiej natury, która mimo że omylna, przewyższa zdecydowanie "nieomylny"¹ komputer w inteligentnych zadaniach rozpoznawania treści scen i ich interpretacji. Idąc dalej, zasadniczym zamiarem stosowania CAIU jest humanizacja pracy, a nie odczłowieczona automatyzacja. Komputerowe wspomaganie pozwala w większym stopniu wykorzystać niezastąpiony potencjał specjalisty, m.in. poprzez:

- wzmocnienie ludzkich zdolności oceny poprzez zwiększenie czułości postrzegania, wskazywanie wszystkich cech odbiegających od normy (wzorca, standardu, tendencji), bardziej wyrazistą prezentację cech wizualnych, selekcję treści, dodatkową wizualizację rozkładu cech, które mają kluczowe znaczenie diagnostyczne itp. - zobacz przykład na rys. 4.3;

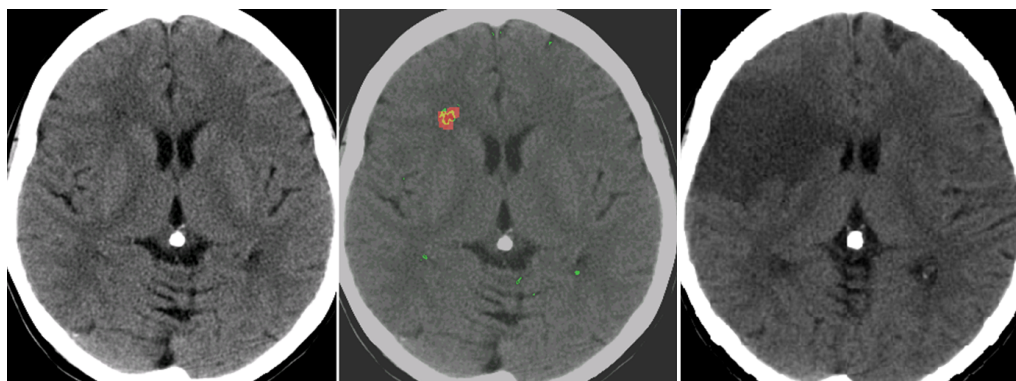
¹załóżmy, że komputer jest sprawny i nieomylnie po wielokroć realizuje zaimplementowane w nim algorytmy

- koncentrację zaangażowania specjalistów na rozwiązywaniu kluczowych problemów interpretacji obrazów poprzez oszczędne dysponowanie czasem ich pracy, teleinformatyczny automatyzm w dostarczaniu wszystkich niezbędnych w ocenie informacji oraz ułatwienie zapisu efektów interpretacji (przyjazne interfejsy użytkownika ułatwiające formalizację ocen i formułowanie wskazań, generację raportów, kontrolę protokołów ocen itp.);
- wspomoczenie procesu interpretacji obiektów i cech poprzez automatyczne wskazania obszarów zainteresowania, ekstrakcję numerycznych deskryptorów i zobiektywizowany opis przypadków w kategoriach użytkowania, z całym dostępnym ontologicznym kontekstem uporządkowanych zasobów wiedzy - zobacz przykład na rys. 4.4;
- pomoc na etapie formułowania decyzji poprzez dostarczenie zintegrowanych sugestii, bazujących na mechanizmach wnioskowania dostępnych ontologii dziedzinowych, weryfikujących wskazany zasób treści oraz wynik jej interpretacji, a także klasyfikatorach kategorii stosowanych poziomów abstrakcji, analizie referencyjnych przypadków treściowo i znaczeniowo podobnych, itp.



Rysunek 4.3: Medyczny przykład zwiększonej czułości postrzegania zmian ukrytych poprzez wizualizację kluczowych cech diagnostycznych, potwierdzonej w szeregu testów klinicznych [72]. Od lewej do prawej kolejno: a) warstwa wczesnego (do 3 godzin po wystąpieniu objawów) badania tomografii komputerowej osoby z objawami udaru, bez widocznych symptomów choroby, b-c) dwa obrazy będące efektem przetwarzania badania wczesnego, z wydobytym obszarem hipodensyjnym (czarna plama), świadczącym o występowaniu udaru niedokrwiennego, d) *follow-up* tego pacjenta (tj. badanie wykonane kilkanaście godzin po badaniu wczesnym), z widocznymi objawami udaru – opracowanie własne - metoda szerzej opisana w [71].

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



Rysunek 4.4: Przykładowe wskazania wspomagające interpretację wczesnych badań udarowych, bazujące na analizie cech teksturowych tkanki mózgowia. Od lewej: wybrana warstwa badania wczesnego, automatyczne wskazanie obszarów podejrzanych (na zielono) oraz rozpoznanych obszarów niedokrwiennych (na czerwono) oraz *follow-up* potwierdzający wystawienie udaru we wskazanej automatycznie lokalizacji.

obiektów, struktur, drobnych detali, kompozycji treści przez synergii tych znaczeń do postaci odczytywanej informacji. Dobrze opisuje ten problem koncepcja rezonansu poznawczego w diagnostyce obrazowej [80, 81] – zobacz punkt ??.

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 wspomaganie coraz częściej przyjmuje formę odwrotną: **”jak rozumieć i zobaczyć 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.

Bibliografia

- [1] J.P. Hornak (2002) Encyclopedia of imaging science and technology. John Wiley & Sons.
- [2] H.H. Barrett, K.J. Myers (2003) Foundations of image science. John Wiley & Sons.
- [3] K. Keeler, J. Westbrook (1995) 'Short encodings of planar graphs and maps. *Disc. Appl. Math.* pp. 239-252.
- [4] M. Deering (1995) Geometry compression. In SIGGRAPH'95 Conference Proceedings, pp. 13-20.
- [5] R.G. Kuehni (2002) The early development of the Munsell system. *Color Research and Application* 27(1): 20–27.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] A.A.G. Requicha (1980) Representations for rigid solids: theory methods and systems. *ACM Comput Surv* 12:437–464.

- [13] J. Gomes, L. Velho (1995) Abstraction paradigms for computer graphics. *The Visual Computer* 11:227–239.
- [14] L. Velho, P.C.P. Carvalho (2003) Mathematical optimization in graphics and vision. *Course Notes – SIGGRAPH 2003*.
- [15] P.J. Davies (1965) Interpolation and approximation. Blaisdell Publishing Company. A Division of Ginn and Company New York-Toronto-London.
- [16] J.L. Kulikowski (2010) Ontological models as tools for image content understanding. *LNCS 6374*:43–58.
- [17] R. Tadeusiewicz, J. Śmietański (2011) Pozyskiwanie obrazów medycznych oraz ich przetwarzanie, analiza, automatyczne rozpoznawanie i diagnostyczna interpretacja. Wydawnictwo Studenckiego Towarzystwa Naukowego, Kraków.
- [18] F. Galton (1883) *Inquiries into Human Faculty and its Development*. AMS Press, New York.
- [19] A. Binet (1905) New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, 12:191–244; angielskie tłumaczenie z 1916 r. – E.S. Kite w *The development of intelligence in children*. Vineland, NJ: Publications of the Training School at Vineland.
- [20] B. Pruszyński, redakcja (2000) *Diagnostyka obrazowa. Podstawy teoretyczne i metodyka badań*. Wydawnictwo Lekarskie PZWL, Warszawa.
- [21] B.J. Erickson, B. Bartholmai (2002) Computer-aided detection and diagnosis at the start of the third millennium. *J Dig Imag* 15(2):59–68.
- [22] J.M. Wilson, G. Junger (1968) *Principles and practice of screening for disease*, WHO report, Genewa.
- [23] L.T. Kohn, J.M. Corrigan, M.S. Donaldson (editors) (2000) *To err is human: building a safer health system*. National Academy Press, Washington.
- [24] B. Reason (2000) Human error: models and management, *BMJ* 320(7237):768–70.
- [25] J.T. Reason, J. Carthey, M.R. de Leval (2001) Diagnosing "vulnerable system syndrome": an essential prerequisite to effective risk management. *Quality in Health Care* 10(Suppl II):ii21–ii25.
- [26] M. Bartyzel, M. Sierackiewicz (2009) W poszukiwaniu straconego czasu - strategie poszukiwania przyczyn błędów i ich rozwiązań. *Software Developer's Journal* (174):36–38.
- [27] L.H. Garland (1959) Studies on the accuracy of diagnostic procedures. *Am J Roentgenology* 82:25–38.
- [28] D.L. Renfrew, E.A. Franken et al. (1992) Error in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology* 183:145–50.

- [29] L. Berlin (2007) Accuracy of diagnostic procedures: has it improved over the past five decades? *Am J Roentgenology* 188:1173–8.
- [30] R.L. Siegle, E.M. Baram et al. (1998) Rates of disagreement in imaging interpretation in a group of community hospitals. *Acad Radiol* 5:148–54.
- [31] J.P. Borgstede, R.S. Lewis, M. Bhargavan, J.H. Sunshine (2004) RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. *J Am Coll Radiol* 1:59–65.
- [32] L. Berlin (2007) Radiologic errors and malpractice: a blurry distinction. *Am J Roentgenology* 189:517–22.
- [33] J.R. Muhm, W.E. Miller et al. (1983) Lung cancer detected during a screening program using four-month chest radiographs. *Radiology* 148:609–15.
- [34] J.A. Harvey, L.L. Fajardo, C.A. Innis (1993) Previous mammograms in patients with impalpable breast carcinoma: retrospective vs. blinded interpretation. *Am J Roentgenology* 161:1167–72.
- [35] K.S. Berbaum (1995) Difficulty of judging retrospectively whether a diagnosis has been “missed.” *Radiology* 194:582–3.
- [36] C. Caldwell, E.R. Seamone (2007) Excusable neglect in malpractice suits against radiologists: a proposed jury instruction to recognize the human condition. *Ann Hlth Law* 16:43–77.
- [37] P. Goddard, A. Leslie et al. (2001) Error in radiology. *Brit J Radiol* 74:949–51.
- [38] H.L. Kundel (1989) Perception errors in chest radiography. *Semin Respir Med* 10:203–210.
- [39] J.R. Anderson, D. Bothell et al (2004) An integrated theory of the mind. *Psychological Review* 111(4):1036–1060.
- [40] J.H.W. Pexman, P.A. Barber, M.D. Hill et al. (2001) Use of the Alberta stroke program early CT score (ASPECTS) for assessing CT scans in patients with acute stroke. *Am J Neuroradiol* 22:1534–1542.
- [41] J.W. Oestmann, R. Greene et al. (1988) Lung lesions: correlation between viewing time and detection. *Radiology* 166:451–3.
- [42] L. Kan, I.A. Olivotto et al. (2000) Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 215:563–7.
- [43] C.T. Loy, L. Irwig (2004) Accuracy of diagnostic tests read with and without clinical information: a systematic review. *JAMA* 292:1602–8.
- [44] U.O. Aideyan, K. Berbaum, W.L. Smith (1995) Influence of prior radiologic information on the interpretation of radiographic examinations. *Acad Radiol* 2:205–8.

- [45] B.J. Hillman, S.J. Hessel, R.G. Swensson, P.G. Herman (1977) Improving diagnostic accuracy: a comparison of interactive and Delphi consultations. *Invest Radiol* 12(2):112–5.
- [46] B.J. Hillman, R.G. Swensson et al. (1976) The value of consultation among radiologists. *Am J Roentgenology* 127:807–9.
- [47] E.A. Franken, K.S. Berbaum (1996) Subspecialty radiology consultation by interactive telemedicine. *J Telemed Telecare* 2:35–41.
- [48] D.B. Kopans (2000) Double reading. *Radiol Clin North Am* 38:719–24.
- [49] L.E. Duijm, J.H. Groenewoud, J.H. Hendriks, H.J. de Koning (2004) Independent double reading of screening mammograms in the Netherlands: effect of arbitration following reader disagreements. *Radiology* 231:564–70.
- [50] F.M. Hall (1997) Opinions about mammographic double reading. *Am J Roentgenol* 168:846.
- [51] M.N. Linver, S.B. Paster et al. (1992) Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases. *Radiology* 184:39–43.
- [52] M.J. Smith (1967) *Error and variation in diagnostic radiology*. C.C. Thomas (Springfield, Ill).
- [53] J. Zhang, V.L. Patel, J.P. Turley, T.R. Johnson (2005) Health informatics and medical errors. Report of Business Briefing: North American Pharmacotherapy, pp. 1–4.
- [54] J.S. Ash, P.N. Gorman, V. Seshadri, W.R. Hersh (2002) Computerized physician order entry in U.S. hospitals: results of a 2002 Survey. *J Am Med Inf Assoc* 11(2):95–9.
- [55] J. Zhang, V.L. Patel, T.R. Johnson (2002) Medical error: is the solution medical or cognitive? *J Am Med Inf Assoc* 9(6 Suppl):S75–77.
- [56] A. Przelaskowski (2010) Komputerowe wspomaganie obrazowej diagnostyki medycznej - stare wyzwania i szanse rozwoju. *Acta Bio-Optica et Informatica Medica* 16(3):245–253.
- [57] G.S. Lodwick, C.L. Haun, W.E. Smith et al. (1963) Computer diagnosis of primary bone tumor. *Radiology* 80:273–5.
- [58] H. Becker, W. Nettleton, P. Meyers, J. Sweeney, C. Jr. Nice (1964) Digital computer determination of a medical diagnostic index directly from chest X-ray images. *IEEE Trans Biomed Eng BME-11:67–72*.
- [59] P.H. Myers, C.M. Nice, H.C. Becker et al. (1964) Automated computer analysis of radiographic images. *Radiology* 83:1029–34.

- [60] R.P. Kruger, J.R. Towns, D.L. Hall et al. (1972) Automated radiographic diagnosis via feature extraction and classification of cardiac size and shapedescriptors. *IEEE Trans Biomed Eng BME-19(3)*:174–86.
- [61] R.P. Kruger, W.B. Thompson, A.F. Turner (1974) Computer diagnosis of pneumoconiosis. *IEEE Trans Syst Man Cybernetics SMC-4(1)*:44–7.
- [62] R. Connors, C. Harlow, S. Dwyer (1982) Radiographic image analysis: past and present. In *Proc. 6th. Int. Conf. Pattern Recognition, Munich, Germany*, pp. 1152–1168.
- [63] K. Doi (2007) Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comp Med Imag Graph 31*:198–211.
- [64] R. Engle (1992) Attempts to use computers as diagnostic aids in medical decision making: A thirty-year experience, In *Perspectives Biol. Med.*, 35:207–217.
- [65] G.S. Lodwick (1966) Computer-aided diagnosis in radiology. A research plan. *Invest Radiol 1(1)*:72–80.
- [66] F. Li, M. Aoyama, J. Shiraishi et al. (2004) Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT by using computer-estimated likelihood of malignancy. *AJR 183*:1209–15.
- [67] M.L. Giger, K. Doi, H. MacMahon (1988) Image feature analysis and computer-aided diagnosis in digital radiography. 3. "Automated detection of nodules in peripheral lung fields". *Med Phys 15*:158–166.
- [68] H.P. Chan, K. Doi, S. Galhotra, C.J. Vyborny, H. MacMahon, P.M. Jokich (1987) Image feature analysis and computer-aided diagnosis in digital radiography. *Med Phys 14*:538–48.
- [69] K.R. Hoffmann, K. Doi, H.P. Chan, L. Fencil, H. Fujita, A. Muraki (1986) Automated tracking of the vascular tree in DSA images using a double-square-box region-of-search algorithm. *Proc SPIE 626*:326–33.
- [70] P. Caligiuri, M.L. Giger, M. Favus, H. Jia, K. Doi, L.B. Dixon (1993) Computerized radiographic analysis of osteoporosis. *Radiology 186*:471–474.
- [71] A. Przelaskowski, K. Sklinda, B. Ciszek (2009) Modelowanie subtelných zmian chorobowych mózgowia wspomagające neurodiagnostykę. W *Neurocybernetyka teoretyczna* pod red. R. Tadeusiewicza, str. 215–268, Wydawnictwo Uniwersytetu Warszawskiego.
- [72] A. Przelaskowski, J. Walecki, K. Sklinda, G. Ostrek (2009) Improving diagnostic value of CT examinations in hyperacute ischemic stroke. *Acta Neurochirurgica Supplementum 106*:165-170.
- [73] T.C. Pan, M.N. Gurcan, S.A. Langella et al. (2007) GridCAD: grid-based computer-aided detection system. *RadioGraphics 27*:889–897.

- [74] M.N. Gurcan, T.C. Pan, A. Sharma et al. (2007) GridIMAGE: a novel use of grid computing to support interactive human and computer-assisted detection decision support. *J Dig Imag* 20(2):160–171.
- [75] K. Yuan, Z. Tian, J. Zou, Y. Bai, Q. You (2011) Brain CT image database building for computer-aided diagnosis using content-based image retrieval. *Information Processing and Management* 47(2):176–185.
- [76] T.T. Dao, F. Marin, M.C. Ho Bas Tho (2009) Ontology-based computer-aided decision system: a new architecture and application concerning the musculoskeletal system of the lower limbs. *IFMBE Proc*, 22(11):1540–1543.
- [77] B. Verma, J. Zakos (2001) A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE Tran Inf Tech Biomed* 5(1):46–54.
- [78] W.H. Land, D.W. McKee, F.R. Anderson et al. (2006) Using computational intelligence for computer-aided diagnosis of screen film mammograms. Chapter 10 in *Recent advances in breast imaging, mammography, and computer-aided diagnosis of breast cancer*. SPIE Press Monograph vol. PM155, pp. 321–381.
- [79] J.J. Fenton, S.H. Taplin, P.A. Carney et al. (2007) Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 356(14):1399–409.
- [80] R. Tadeusiewicz, M.R. Ogiela (2004) *Medical image understanding technology*. Springer-Verlag, Berlin-Heidelberg.
- [81] M.R. Ogiela, R. Tadeusiewicz (2008) *Modern computational intelligence methods for the interpretation of medical images*. Studies in Computational Intelligence 84. Springer-Verlag, Berlin-Heidelberg.
- [82] M. Grimnes, A. Aamodt (1996) A two layer case-based reasoning architecture for medical image understanding. *Lecture Notes in Computational Science* 1168:164–78.
- [83] D.M. Pisanelli, D. Zaccagnini, L. Capurso, M. Koch (2003) An ontological approach to evidence-based medicine and meta-analysis. *Stud Health Technol Inform* 5:543–8.
- [84] Q. Da, E.R. Denton, R. Zwigelaar (2006) Semantic analysis on medical images: a case study, *Proc 18th Int Con Patt Rec (ICPR'06)*, pp. 1260–3.
- [85] C.E. Kahn, D.S. Channin, D.L. Rubin (2006) An ontology for PACS integration, *J Digit Imag* 19(4):316–27.
- [86] W. Skarbek (2004) *Podstawy multimediiów*. Politechnika Warszawska.
- [87] 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.

- [88] 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.
- [89] 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.
- [90] L. Floridi (2007) In defence of the veridical nature of semantic information. Eur J Anal Philosophy 3:31–41.
- [91] 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.
- [92] 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.
- [93] 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).
- [94] R.M. Gray (1990) Entropy and information theory. Springer-Verlag, New York.
- [95] C.E. Shannon (1948) A mathematical theory of communications', Bell System Technical Journal, 27:379-423, 623-656.
- [96] 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.
- [97] W. Skarbek (1993) Metody reprezentacji obrazów cyfrowych. Akademicka Oficyna Wydawnicza PLJ, Warszawa.
- [98] M. Rabbani, P.W. Jones (1991) Digital image compression techniques. SPIE Optical Engineering Press, TT 7, Bellingham, Washington, USA.
- [99] N.S. Chang, K.S. Fu (1979) A relational database system for images. Purdue University. TR-EE 7928.
- [100] V.V. Vapnik, A.Y. Chervonenkis (1971) On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and Its Applications 16:264-280.
- [101] T. Deselaers (2002) Features for image retrieval. PhD thesis, Aachen Univeristy.
- [102] 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.
- [103] A. Puri, T. Chen (2000) Multimedia systems, standards, and networks. Marcell Dekker, Inc., Nowy Jork.

- [104] K. Keeler, J. Westbrook (1995) 'Short encodings of planar graphs and maps. *Disc. Appl. Math.* pp. 239-252.
- [105] M. Deering (1995) Geometry compression. In *SIGGRAPH'95 Conference Proceedings*, pp. 13-20.
- [106] K. Hosaka (1986) A new picture quality evaluation method. *Proc Int Picture Coding Symposium*, Tokyo, Japan.
- [107] A.M. Eskicioglu, P.S. Fisher (1995) Image quality measures and their performance. *IEEE Tran Comm* 43(12):2959-65.
- [108] J.A. Swets (1979) ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiology* 14:109-121.
- [109] 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.
- [110] H. Li, K.J.R. Liu, S.C.B. Lo (1997) Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms. *IEEE Trans Med Imag* 16(6):785-98.
- [111] M.R. Ogiela, R. Tadeusiewicz (2008) Modern computational intelligence methods for the interpretation of medical images. *Studies in Computational Intelligence* 84. Springer-Verlag, Berlin-Heidelberg.
- [112] J.L. Kulikowski (2010) Ontological models as tools for image content understanding. *LNCS 6374:43-58*, *Proc ICCVG Part I*.
- [113] T.T. Dao, F. Marin, M.C. Ho Bas Tho (2009) Ontology-based computer-aided decision system: a new architecture and application concerning the musculoskeletal system of the lower limbs. *IFMBE Proc* 22(11):1540-1543.
- [114] T.T. Dao, F. Marin, M.C. Ho Bas Tho (2009) Clinical validated computer-aided decision system to the clubfeet deformities. *Proc International Conference of the IEEE EMBS*, pp. 6230-6233.
- [115] B. Zheng (2009) Computer-aided diagnosis in mammography using content-based image retrieval approaches: current status and future perspectives. *Algorithms* 2(2):828-49.
- [116] E. J. Candés, J. Romberg, T. Tao (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Tran Inform Theory* 52(2):489-509.
- [117] D.L. Donoho (2006) Compressed sampling. *IEEE Tran Inform Theory* 52(4):1289-1306.
- [118] 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 Tran Medical Imag* 20(1):45-57.

- [119] M. Alfó, L. Nieddu, D. Vicari (2008) A finite mixture model for image segmentation. *Stat Comput* 18:137–150.
- [120] Y. Wang, T. Adali, M.T. Freedman, S.K.Mun (1996) MR brain image analysis by distribution learning and relaxation labeling. *Proc. 15th South. Biomed. Eng. Conf.*, Dayton, OH, pp. 133–136.
- [121] M.A. Hurn, K.V. Mardiasian et al. (1996) Bayesian fused classification of medical images. *IEEE Tran Medical Imag* 15(6):850–58.
- [122] A. Przelaskowski (2010) Uproszczone wspomaganie obrazowej diagnostyki medycznej. *Przegląd telekomunikacyjny LXXXIII*(12):1748–1755.
- [123] T. Deselaers (2002) Features for image retrieval. PhD thesis, Aachen Univeristy.
- [124] 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.
- [125] J. P. Eakins, M. E. Graham (2000) Content-based image retrieval. Technical report, JISC Technology Application Program.
- [126] 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.
- [127] P.Welter, C. Hockena, T.M. Deserno, C. Grouls (2009) Workflow management of content-based image retrieval for CAD support in PACS environments based on IHE. *JCAR* 5(4):393–400.
- [128] P. Boniński (2008) Metody indeksowania obrazów medycznych na potrzeby radiologii cyfrowej, rozprawa doktorska, Politechnika Warszawska.
- [129] 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 .
- [130] 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.
- [131] 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.
- [132] R.C. Veltkamp, M. Tanase (2002) Content-based image retrieval systems: a survey. Technical Report UU-CS-2000-34 (revised October 28, 2002).
- [133] 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.

- [134] H. Eidenberger (2004) Visual Information Retrieval. PhD thesis, Technischen Universität Wien.
- [135] 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.
- [136] 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.
- [137] 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.
- [138] 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.
- [139] 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.
- [140] G. Ciocca, R. Schettini (1999) Similarity retrieval of trademark images. Proc Int Con Image Anal Proces.
- [141] R.G. Kuehni (2002) The early development of the Munsell system. Color Research and Application 27(1): 20–27.
- [142] 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.
- [143] 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.
- [144] 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.
- [145] 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.
- [146] 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.
- [147] 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.

- [148] C. Vertan, N. Boujemaa (2000) Using fuzzy histograms and distances for color image retrieval. Proc Challenge of Image Retrieval (CIR), pp. 85–89.
- [149] S. Siggelkow (2002) Feature histograms for content-based image retrieval. PhD thesis, Albert-Ludwigs-Universit at Freiburg.
- [150] L.A. Zadeh (1965) Fuzzy sets. *Information and Control* 8(3):338–353.
- [151] J. Han, K.-K. Ma (2002) Fuzzy color histogram and its use in color image retrieval. *IEEE Tran Im Proc* 11(8):944–952.
- [152] 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.
- [153] 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.
- [154] 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.
- [155] 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.
- [156] W. Ma, B. Manjunath (1996) Texture features and learning similarity. Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR'96), pp. 425–430.
- [157] 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.
- [158] 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.
- [159] 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.
- [160] 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.
- [161] 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.
- [162] R. Baeza-Yates, B. Ribeiro-Neto (1999) *Modern Information Retrieval*. Addison Wesley.

- [163] A. Przelaskowski (2002) Falkowe metody kompresji danych obrazowych. Prace naukowe, Elektronika z. 138, Oficyna Wydawnicza Politechniki Warszawskiej.
- [164] A. Przelaskowski (2005) Kompresja danych: podstawy, metody bezstratne, kodery obrazów. Wydawnictwo BTC, Warszawa.
- [165] N. Sloane, A.D. Wyner (eds.)(1993) Claude Elwood Shannon : collected papers (New York).
- [166] J. Ziv, A. Lempel (1977) A universal algorithm for sequential data compression., IEEE Trans Inf Theory 23(3):337–343.
- [167] J. Ziv, A. Lempel (1978) Compression of individual sequences via variable-rate coding. IEEE Trans Inf Theory 24(5):530–536.
- [168] J.A. Storer, T.G. Szymanski (1982) Data compression via textual substitution. J ACM 29:928–951.
- [169] T. Welch (1984) A technique for high-performance data compression. IEEE Computer 17(6):8–19.
- [170] G. Langdon (1984) An introduction to arithmetic coding. IBM J Res Dev 28(2):135–149.
- [171] 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.
- [172] I. Witten, R. Neal, J. Cleary (1987) Arithmetic coding for data compression. Comm ACM 30(6):520–540.
- [173] P.G. Howard, J.S. Vitter (1994) Arithmetic coding for data compression. Proc IEEE, 82(6):857–865.
- [174] P.G. Howard, J.S. Vitter (1993) Fast and efficient lossless image compression. Proc IEEE Data Compression Conference, pp. 351–360.
- [175] 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.
- [176] J.M. Shapiro (1993) Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans Sig Proc 41(12):3445–3462.
- [177] C.C. Cutler (1952) Differential quantization for television signals. U.S. Patent 2,605,361.
- [178] T. Endoh, T. Yamazaki (1986) Progressive coding scheme for multilevel images. Proc Picture Coding Symposium, pp. 21–22.
- [179] P. Roos, M.A. Viergever, M.C.A. van Dijke, J.H. Peters (1988) Reversible intra-frame compression of medical images, IEEE Trans Med Imag 7(4):328–336.

- [180] 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.
- [181] X. Wu (1996) Lossless compression of continuous -tone images via context selection and quantization. IEEE Trans Im Proc 5(6):656–664.
- [182] 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.
- [183] ISO/IEC 15444 International Standard (JPEG2000),Information technology - JPEG 2000 image coding system, 2000.
- [184] 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.
- [185] ISO/IEC 11544 International Standard (JBIG), Information Technology - Coded representation of picture and audio information - Progressive bi-level image compression, 1993.
- [186] ISO/IEC 14492 International Standard (JBIG2), Information Technology - Lossy/lossless coding of bi-level Images, 2001.
- [187] 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.
- [188] L. Bottou, P.G. Howard, Y. Bengio (1998) The Z-coder adaptive binary coder. Proc IEEE Data Compression Con, pp. 13–22.
- [189] Graphics Interchange Format (1990) version 89a, CompuServe Incorporated, <http://www.w3.org/Graphics/GIF/spec-gif89a.txt>
- [190] T. Boutell, A. Dilger, et al: PNG (Portable Network Graphics) specification, www.w3.org/TR/REC-png-multi.html
- [191] 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.
- [192] S.P. Lloyd (1982) Least squares quantization in PCM. IEEE Tran Inform Theory IT-28:129–137. Jest to reprodukja manuskryptu raportu technicznego Bell Laboratories z 1957 roku.
- [193] J. Max (1960) Quantizing for minimum distortion. IRE Tran Inform Theory IT-6:7–12.
- [194] 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.
- [195] A. Moffat (1990) Implementing the PPM data compression scheme. IEEE Trans Comm COM-38 11:1917-1921.

- [196] 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.
- [197] Lossless data compression software benchmarks/comparisons: <http://www.maximumcompression.com/>
- [198] SPMG/JPEG-LS V.2.2 codec, <ftp://spmge.ece.ubc.ca/pub/jpeg-ls/ver-2.2/>, Signal Processing & Multimedia Group at the University of British Columbia, USA.
- [199] <http://www.cl.cam.ac.uk/~mgk25/jbigkit/>
- [200] X. Wu, CALIC codec, ftp://ftp.csd.uwo.ca/pub/from_wu/v.arith/
- [201] 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.
- [202] http://www.msoftware.co.nz/downloads_page.php
- [203] S.G. Chang, B. Yu, M. Vetterli (2000) Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans Image Proces* 9:1532–1546.
- [204] 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.
- [205] 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.