

ROZDZIAŁ 9. METODY OCENY JAKOŚCI DANYCH

Ocena jakości obrazów rekonstruowanych na podstawie reprezentacji utworzonej w procesie stratnej kompresji jest zagadnieniem wieloaspektowym, silnie zależnym od konkretnej aplikacji (klasy obrazów i sposobów ich wykorzystania), a przy tym jego waga rośnie wraz z rozwojem coraz doskonalszych technik kompresji w rozszerzającej się gamie zastosowań. Należy przy tym dodać, że dla wielu bardziej wymagających zastosowań zagadnienie to nie jest jeszcze w pełni rozwiązane. Dobrym przykładem jest ocena jakości kompresowanych stratnie obrazów medycznych. Zastosowanie wyrafinowanych metod kompresji nieodwracalnej, dających znaczne oszczędności czasowo-sprzętowe, jest tutaj praktycznie niemożliwe, gdyż spotyka się z ostrym sprzeciwem wielu lekarzy lub też w wielu krajach (np. w USA) niedopuszczalne przez prawo jest wykorzystanie w diagnostyce obrazów rekonstruowanych po stratnej kompresji. Nie ma bowiem wystarczająco pewnych metod oceny jakości tych obrazów oraz określenia wartości dopuszczalnych stopni kompresji, tj. granicznych wartości powyżej których zniekształcenia w obrazie rekonstruowanych osiągają poziom niemożliwy do zaakceptowania z punktu widzenia konkretnych zastosowań. W zastosowaniach stratnych metod do kompresji obrazów medycznych wartości dopuszczalnych stopni kompresji zależą silnie od typu systemu obrazowania, w którym powstaje konkretny obraz, oraz od rodzaju badania diagnostycznego. Warunkiem koniecznym akceptowalności jest tutaj zachowanie diagnostycznej wiarygodności rekonstruowanego obrazu.

W tym krótkim omówieniu zaprezentowano ogólną charakterystykę metod oceny jakości, a następnie na przykładzie problemu kompresji obrazów medycznych różnych modalności przedstawiono praktyczne zastosowanie tych metod i dalszą perspektywę ich rozwoju. Zdecydowana większość prezentowanych zagadnień dotyczy znacznie szerszego kręgu zastosowań. Analiza cech obrazów z punktu widzenia efektywności ich kompresji jest w wielu aspektach analogiczna do rozważań dotyczących dokładności odwzorowania przedmiotu obrazowania w tworzonym obrazie. Obraz ten jest jedynie pewnym przybliżeniem prezentowanych obiektów, obciążonym szeregiem zniekształceń powstających w czasie całego procesu konstruowania obrazu. Przedstawione metody oceny jakości obrazów rekonstruowanych po stratnej kompresji mogą być niekiedy wykorzystane do oceny jakości systemów obrazowania (obrazem odniesienia są wówczas różnego typu fantomy).

Istotną sprawą we wszelkich rozważaniach na temat sposobów oceny jakości rekonstruowanych obrazów jest jakość obrazu oryginalnego. Trzeba pamiętać o tym, że każdy system obrazowania ma swoje ograniczenia, nie wszystkie cechy prezentowanych obiektów są odzwierciedlane w rejestrowanych obrazach. Każdy system obrazowania można opisać funkcją przenoszenia (czasowo-częstotliwościową), która stanowi kompletny opis danego systemu [1]. Przykładowo funkcja ta określa częstotliwość graniczną, a obiekty składające się z częstotliwości wyższych (w analizie fourierowskiej) mogą być w ogóle nie reprezentowane w zarejestrowanym obrazie. Ponadto występuje w tych obrazach szereg dodatkowych elementów

(zniekształceń), które nie mają nic wspólnego z przedstawianymi obiektami. A skoro obraz oryginalny jest zniekształcony, to po pierwsze można stosując różne techniki przetwarzania w wielu przypadkach poprawić jego jakość, a po drugie sam proces stratnej kompresji, będący pewnego rodzaju filtracją, nie musi oznaczać automatycznie pogorszenia jakości obrazu oryginalnego. Łatwo można sobie wyobrazić sytuację, w której obraz rekonstruowany jest lepszej jakości od oryginału, bo w stratnej kompresji usunięto wysokoczęstotliwościowe pasmo szumów, znacznie pogarszające jakość obrazu oryginalnego. Formułując więc problem oceny jakości obrazów rekonstruowanych, właściwie powinniśmy porównywać ten obraz z domniemanym idealnym obrazem prezentowanych struktur (co jest oczywiście w praktyce szalenie trudne, czy wręcz niemożliwe), a przynajmniej pamiętać, że podlegający kompresji obraz oryginalny jest również zniekształcony, a pewne jego własności mogą nawet ulec poprawie podczas nieodwracalnej jego kompresji (co jest możliwe, szczególnie przy projektowaniu i przeprowadzaniu subiektywnej oceny jakości).

9.1. Charakterystyka klasy obrazów medycznych

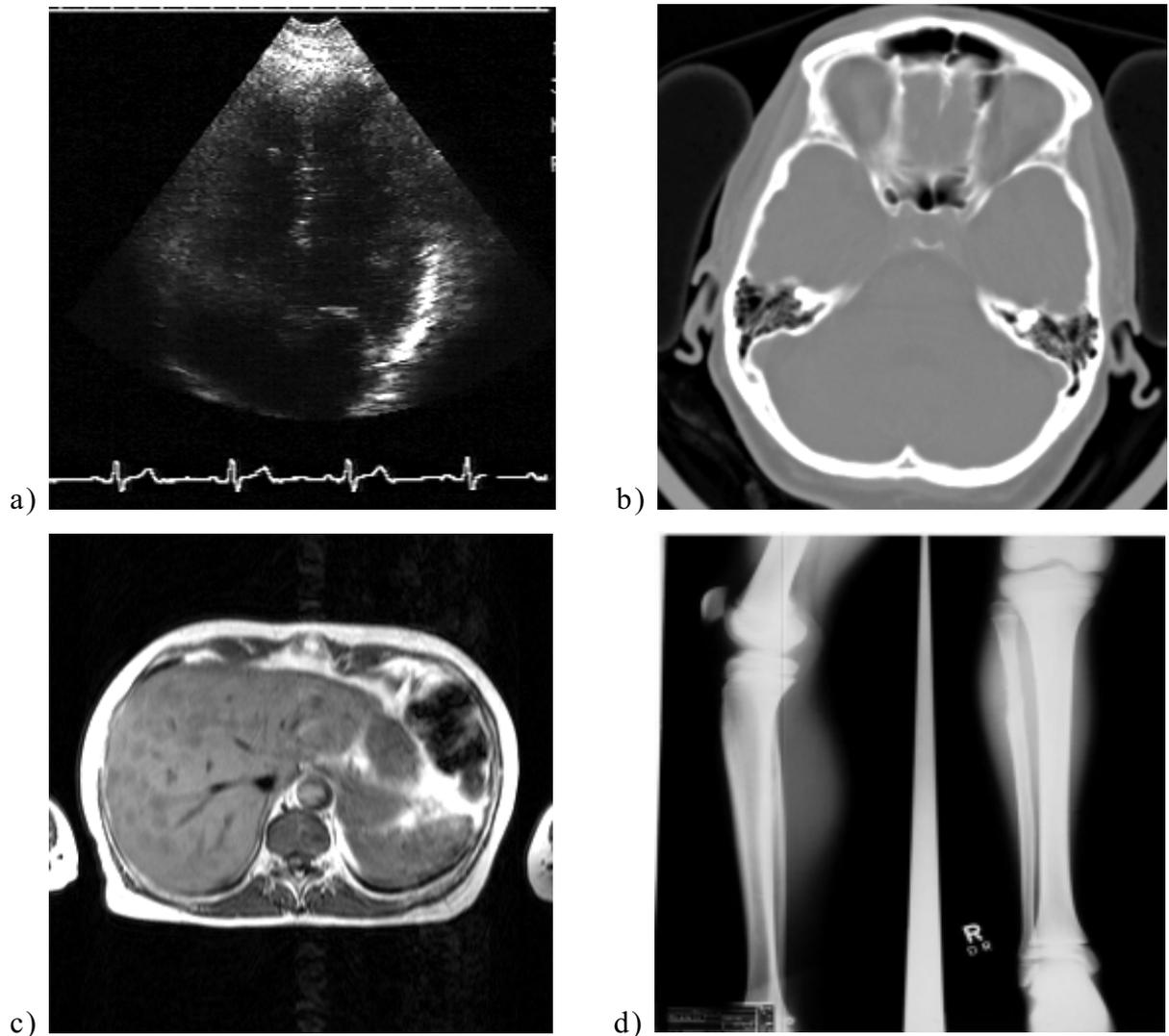
Odmienny charakter różnych technik obrazowania w medycynie wpływa na parametry tworzonych obrazów, a tym samym predestynuje te obrazy do mniej lub bardziej oszczędnej archiwizacji (czy też sprawniejszej transmisji) lub też decyduje o wyborze różnych technik prezentacji i poprawy jakości tychże obrazów.

Sposób konstruowania obrazu decyduje także o wyborze optymalnych wariantów algorytmów kompresji, dostarczając *a priori* pewnych informacji na temat przetwarzanej klasy obrazów.

Medyczne systemy obrazowania

Obrazy medyczne, tworzone wskutek różnorodnych zabiegów technologicznych, umożliwiają specjalistom obserwację różnego typu struktur i procesów fizjologicznych zachodzących we wnętrzu ciała ludzkiego. Analiza obrazów służy podjęciu właściwej diagnozy, a także planowaniu ewentualnej terapii. Obrazy medyczne są jakby oknem do ludzkiego ciała. Żadne z tych okien nie ukazuje wszystkiego. Poszczególne metody obrazowania ujawniają różne cechy charakterystyczne ciała pacjenta. Przykładowo badania MRI (ang. magnetic resonance imaging) - tomografii rezonansu magnetycznego - czy też CT (ang. computed tomography) - rentgenowskiej tomografii komputerowej - uwidaczniają głównie morfologię (budowę strukturalną) prezentowanych narządów. Z kolei badania PET (ang. positron emission tomography) - emisyjnej tomografii pozytonowej - czy SPECT (ang. single photon emission computed tomography) - jednofotonowej tomografii emisyjnej - przede wszystkim ich funkcje dynamiczne. Badania USG - ultrasonograficzne - pozwalają natomiast przy pomocy stosunkowo niedrogiej aparatury ocenić własności zarówno statyczne jak i dynamiczne struktur zbudowanych z tkanki miękkiej. Różne techniki radiografii cyfrowej CR (ang. computed radiography) pozwalają uzyskać tomograficzne obrazy różnych struktur tkanki

kostnej i miękkiej w postaci cyfrowej o jakości często przewyższającej możliwości tradycyjnego zapisu analogowego na filmach.



Rys. 9.1. Przykłady badań z różnych systemów obrazowania w medycynie: a) USG, b) CT, c) MRI, d) CR.

Przykładowe obrazy z różnych medycznych systemów obrazowania, pozwalające dostrzec podstawowe różnice co do charakteru prezentowanej informacji diagnostycznej, przedstawia rys. 9.1.

Pięć głównych elementów wpływających na proces obrazowania w medycynie to: osoby pacjenta, operatora systemu i specjalisty-obszawatora, a także konkretny system obrazowania i sam obraz. W wielu przypadkach lekarz specjalista, będący obserwatorem tworzonych obrazów, jest jednocześnie operatorem systemu. Widoczność cech anatomicznych czy funkcjonalnych zależy jest od właściwości systemu obrazowania, jakości urządzeń obrazujących, algorytmów przetwarzających i prezentujących rejestrowane dane.

Większość systemów medycznego obrazowania posiada pewną liczbę wariantowych rozwiązań, które są dobierane w zależności od rodzaju badań, cech pacjenta, wymagań specjalisty dokonującego diagnozy itp. Mogą to być

wymienne części systemu (ekrany, płytki wzmacniające, kolimatory w radiografii, przetworniki w ultrasonografii, czy cewki w tomografii rezonansu magnetycznego) lub też odpowiednio dobierane wartości różnych wielkości fizycznych związanych z procesem tworzenia obrazu (wielkość napięcia na lampie rentgenowskiej, wielkość wzmocnienia odbieranej wiązki ultradźwięków w USG, czas echa w MRI, czy też aktywność radionuklidu w badaniu scyntygraficznym). Wszystkie te elementy wpływają na jakość obrazu i decydują o dobrej widoczności prezentowanych struktur wewnętrznych. Wizualna czułość systemów medycznego obrazowania zależy więc od wielu czynników na drodze od obiektu do oka obserwatora obrazu i może być, za pomocą adaptacyjnych metod, optymalizowana w celu jak najlepszego psychowizualnego odbioru prezentowanych struktur.

Technika formowania obrazów w dużym stopniu wpływa na różnego typu nadmiarowość oryginalnej reprezentacji tworzonych obrazów. Parametry systemów obrazowania medycznego mogą być także optymalizowane pod kątem podatności tworzonych obrazów na kompresję zwiększając efektywność metod archiwizacji lub transmisji. Obowiązują tutaj następujące ogólne zasady:

- zwiększanie rozdzielczości powoduje silniejszą korelację pomiędzy wartościami pikseli, co pozwala na osiąganie wyższych stopni kompresji (oczywiście rośnie najczęściej liczba bitów potrzebna do zapisu skompresowanego obrazu, gdyż zwiększa się automatycznie liczba pikseli w obrazie, lecz przyrost względny ilości bitów jest mniejszy);
- zwiększanie liczby poziomów kwantyzacji wartości pikseli (bądź zapewnienie większej ich dynamiki w procesie formowania obrazu) zmniejsza korelację pomiędzy pikselami, co redukuje możliwości kompresji;
- występowanie jakichkolwiek szumów w obrazie, wprowadzanych na różnych etapach konstruowania obrazu, zmniejsza korelację pomiędzy pikselami, a więc ogranicza uzyskiwane stopnie kompresji.

Wartość diagnostyczna obrazów medycznych

Zasadniczym czynnikiem wpływającym na dopuszczalny poziom redukcji informacji z obrazów medycznych w procesie stratnej kompresji jest wartość diagnostyczna tych obrazów. Jest ona związana ze zdolnością obserwatora do detekcji symptomów patologicznych i zależy od trzech zasadniczych czynników:

- jakości obrazu,
- warunków obserwacji,
- charakterystyki pracy lekarza-specjalisty.

Proces kompresji nieodwracalnej wpływa bezpośrednio na jakość obrazów. Zmiany w jakości przy takich samych warunkach obserwacji mogą być nieistotne z punktu widzenia konkretnego sposobu diagnozowania, mniej czy bardziej ograniczać możliwości podejmowania poprawnych decyzji lub też w pewnych przypadkach zwiększać skuteczność detekcji patologii.

Jakość obrazu

Jakość obrazów medycznych jest bezpośrednio związana z metodą obrazowania, własnościami konkretnych urządzeń oraz wartościami zmiennych parametrów systemu dobieranymi przez operatora. Składa się na nią co najmniej pięć istotnych elementów: kontrast, rozdzielczość, stosunek sygnału użytecznego do szumów, poziom artefaktów oraz zniekształcenia przestrzenne.

Ludzkie ciało zawiera wiele struktur i obiektów, które w większości metod obrazowania są prezentowane jednocześnie. Zazwyczaj interesuje nas jednak pojedynczy obiekt w otoczeniu bezpośredniego tła i jego relacja do tegoż doraźnego otoczenia. Praktycznie, w większości metod widoczność obiektów jest określana właśnie przez tę relację, a nie przez ogólne charakterystyki wyznaczone na całym obrazie. Zadaniem każdego systemu obrazowania w medycynie jest wyrażenie w skali szarości i przestrzeni obrazu specyficznych różnic pomiędzy tkankami ciała pacjenta. Kontrastowość obrazu zależy zarówno od własności przedstawianego obiektu, jak też systemu. Decyduje o niej zróżnicowanie wielkości charakteryzującej tkanki w danej metodzie obrazowania (impedancji akustycznej w USG, liniowych współczynników pochłaniania w metodach rentgenowskich, ilość zgromadzonego izotopu o danej aktywności w MN, czy też gęstość aktywnych protonów, czas ustawiania się protonów równoległe do linii zewnętrznego pola magnetycznego po zaniku impulsu radiowego T1 oraz czas zaniku namagnesowania poprzecznego T2 w badaniach MR). Ponadto, technika przetworzenia tych różnic na wartości pikseli obrazu wpływa także znacząco na kontrast obrazu. Dobry kontrast pozwala na wyraźne rozróżnienie struktur przy zachowaniu ostrych krawędzi i prezentacji całej różnorodności obrazowanych obiektów. Uzyskuje się to poprzez odpowiednie wykorzystanie całej dynamiki danych obrazowych. Do najczęściej stosowanych globalnych miar kontrastu należy histogram całego obrazu, a do lokalnych miar kontrastu można zaliczyć lokalny histogram, wariancję rozkładu wartości funkcji jasności w pewnym obszarze, a także często stosowaną zależność:

$$K = \frac{f_{\max} - f_{\min}}{f_{\max} + f_{\min}}, \quad (9.1)$$

gdzie K - kontrast, f_{\max}, f_{\min} - maksymalna i minimalna wartość funkcji jasności w lokalnym obszarze lub też średnie wartości funkcji jasności w sąsiednich obszarach jednorodnych.

Kontrast jest najbardziej fundamentalną cechą charakterystyczną obrazu i jest związany bezpośrednio z dynamiką danych obrazowych. Format zapisu danych winien odpowiadać możliwie do uzyskania dynamice danych w konkretnym systemie obrazowania, a sposób prezentacji obrazu winien umożliwić obserwację danych z całego zakresu tej dynamiki. W praktyce dynamika medycznych danych obrazowych jest bardzo zróżnicowana, zarówno w obrębie jednej metody obrazowania, jak też pomiędzy różnymi metodami tworzenia obrazów w medycynie. Algorytmy kompresji możliwie szerokiej klasy obrazów medycznych winny być więc dostosowane do różnej dynamiki danych, zazwyczaj z zakresu 6-16 bitów. Najczęściej spotykaną dynamikę różnego typu obrazów medycznych przedstawia tabela 9.1.

Tabela 9.1. Dynamika medycznych danych obrazowych.

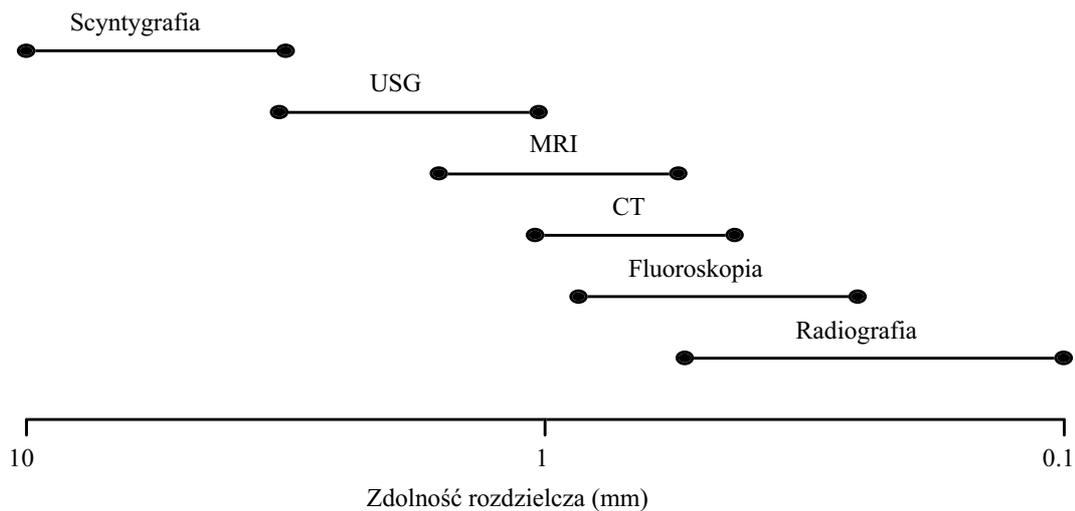
NM	USG	MRI	CR	CT
6 - 12 bitów	6 - 8 bitów	8 - 12 bitów	10 - 14 bitów	12 - 14 bitów

Wśród różnych metod obrazowania szczególne miejsce w dziedzinie kontrastu zajmuje technika CT. Obrazy otrzymywane w rentgenowskiej tomografii komputerowej mają wyższą czułość kontrastu (inaczej rozdzielczość kontrastu) niż w przypadku konwencjonalnej radiografii. Demonstruje się to w możliwości pokazywania na obrazie tych obiektów z miękkiej tkanki, które są niewidoczne w radiografii. Ze względu na dużą czułość kontrastu w obrazach CT występuje zazwyczaj duża liczba ostrych krawędzi, które mają znaczenie diagnostyczne. Wymusza to stosowanie technik kompresji wiernie odtwarzających wysokoczęstotliwościowe składowe obrazu, co jest szczególnie wymagającym warunkiem dla wszystkich metod stratnej kompresji.

Wszystkie metody obrazowania wprowadzają pewne zamazanie (rozmycie) w procesie przetwarzania cech rzeczywistych struktur wewnętrznych na postać obrazu. Podstawowym efektem rozmycia jest redukcja kontrastu i widoczności małych obiektów lub szczegółów, podczas gdy widoczność dużych obiektów praktycznie się nie zmienia. Większa rozdzielczość tworzonych w systemie obrazów eliminuje efekt rozmycia i czyni lepiej widocznymi struktury zawierające drobne elementy o dużej wadze diagnostycznej. Orientacyjne porównanie zdolności rozdzielczej różnych metod obrazowania zostało przedstawione na rys. 9.2.

Inną cechą charakterystyczną wszystkich obrazów medycznych jest szum, który pojawia się często w obrazie w postaci cętkowanej tekstury czy ziarna. Rodzaj i poziom szumu zależy od metody obrazowania. Ogólnie rzecz biorąc, wyższy poziom szumów w stosunku do sygnału użytecznego w obrazie powoduje obniżenie widoczności obiektów. Najsilniej wpływa on na obiekty o niskim kontraście, które znajdują się blisko progu widoczności. Źródłem szumów może być każdy kolejny etap tworzenia obrazu daną metodą, np. różnica energii kwantów promieniowania X, statystyka rejestrowanych kwantów, różnice w prędkościach rozchodzenia się ultradźwięków w tkankach, ugięcia na sąsiednich drobinach, szumy elektroniki itd.

Szum w obrazie staje się bardziej widoczny, jeśli zwiększa się średni poziom przenoszenia kontrastu w obrazie. Z kolei widoczność szumu może być często zredukowana przez efekt rozmycia, ponieważ szum ma zazwyczaj strukturę składającą się z drobnych elementów. Wygładzenie przypadkowej struktury szumów czyni je mniej widocznymi. Oczywiście, rozmycie może spowodować także pogorszenie widoczności użytecznych szczegółów obrazu.



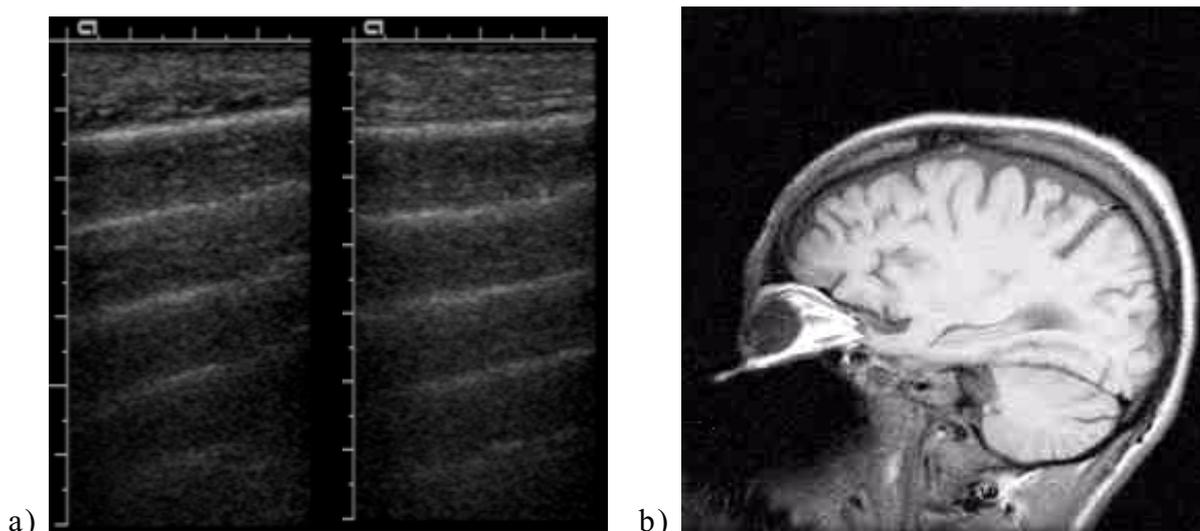
Rys. 9.2. Porównanie typowych zdolności rozdzielczych różnych metod obrazowania w medycynie.

Stratne metody kompresji powodują zazwyczaj dodatkowe zjawisko rozmycia obrazu oryginalnego, które może nawet poprawić jego jakość poprzez redukcję szumów. Warunkiem jest jednak zachowanie informacji diagnostycznej. Orientacyjne średnie wartości stosunku sygnału użytecznego do szumów w obrazach medycznych przedstawia tabela 9.2.

Tabela 9.2. Orientacyjne wartości stosunku sygnału do szumów w obrazach medycznych.

NM	USG	MRI	CR	CT
15 dB	30 dB	50 dB	60 dB	50 dB

Problemem występującym w różnych metodach obrazowania jest powstawanie pewnych cech obrazu, nazywanych artefaktami, które nie reprezentują żadnej struktury ciała, ani dodatkowego obiektu umieszczanego niekiedy celowo w polu obrazowania. W wielu przypadkach artefakty nie wpływają znacząco na widoczność obiektów i dokładność diagnozy, jednak niekiedy mogą pogorszyć widoczność w pewnych obszarach obrazu lub też być błędnie interpretowane jako cechy anatomiczne. Wiele różnych czynników związanych z daną metodą może powodować powstanie artefaktów, poczynając od sposobu przetwarzania danych i algorytmów konstruowania obrazu w systemie, a kończąc na poruszeniu pacjenta czy chociażby przesunięciu kolimatora w lampie rentgenowskiej. Do typowych artefaktów pojawiających się w obrazach można zaliczyć powstawanie cieni, przejaskrawień i rewerberacji w badaniach USG (rys. 9.3 a), nakładanie związane z kodowaniem częstotliwości, fazy i selekcją warstw, chemiczne przesunięcie i obcinanie danych w badaniach MR, ruch obiektu czy też obecność metalowych obiektów w polu obrazowania (rys 9.3 b). itd.



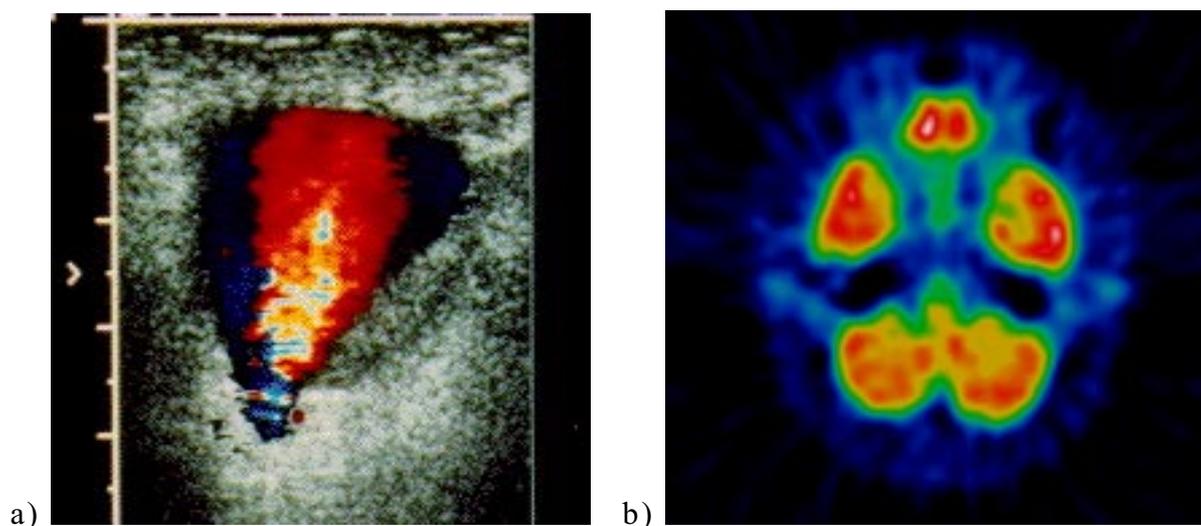
Rys. 9.3 a) artefakty w badaniach USG - przykład rewerberacji czyli wielokrotnych odbić fali ultradźwiękowej pomiędzy na granicach dwóch ośrodków o wyraźnie zróżnicowanych wartościach impedancji akustycznej; b) artefakty w badaniach MRI - przykład zaburzeń obrazu spowodowanych obecnością metalu w ustach pacjenta (metalowy ząb). Obrazy pochodzą ze źródeł internetowych.

Obraz medyczny powinien uwidocznic wewnętrzne narządy ciała, ale winien także właściwie oddać ich rozmiar, kształt oraz względne położenie. Proces obrazowania może wprowadzać zniekształcenia tych parametrów, utrudniające diagnozę. Wpływ stratnej kompresji, eliminujący wysokoczęstotliwościowe szczegóły obrazu i wprowadzający artefakty, może wprowadzać dodatkowe zniekształcenia kształtu i wzajemnego położenia struktur obrazu. Uzyskanie maksymalnej widoczności struktur w danej metodzie ograniczone jest wieloma kompromisami pomiędzy przeciwstawnymi wpływami poszczególnych elementów systemu obrazowania na jakość obrazu i warunki badania, np. pomiędzy wielkością dawki pochłoniętą przez pacjenta a czasem badania w MN. W wielu sytuacjach, jeśli optymalizujemy parametry systemu pod kątem jednego z elementów jakości obrazu, np. szumów, często wpływa to nie najlepiej na inne czynniki składające się na jakość obrazu, takie jak kontrast czy rozmycie szczegółów. Tak więc właściwości techniki obrazowania muszą być starannie dobrane do specyficznych wymagań klinicznej oceny obrazów przez specjalistów.

Warunki obserwacji

Kombinacja dwóch wskaźników stanowi o wyjątkowości każdej z metod obrazowania. Są to widoczne na obrazie cechy charakterystyczne tkanki oraz perspektywa obserwacji. Różnica wartości specyficznej dla danej metody własności tkanek jest odwzorowana w skali szarości dając pewien kontrast obrazu. Z kolei nasza zdolność do widzenia pewnych obiektów lub cech w obrazie zależy silnie od warunków, w jakich oglądamy ten obraz. Odpowiedzialny za to jest z jednej strony sposób prezentacji obrazu w danym systemie (jakość karty graficznej, monitora, możliwości zmiany palety itd.), z drugiej zaś warunki zewnętrzne, w których pracuje specjalista (oświetlenie pomieszczenia, ergonomia pracy itp.). Sposób prezentacji obrazów jest bardzo zróżnicowany i zasadniczo powinien być powiązany z jakością tych obrazów.

W aparatach USG stosowane są najczęściej niewielkie monitory o małej kontrastowości i rozdzielczości, a dodatkową informację np. prędkość przepływu z badań dopplerowskich nanosi się na podstawowy obraz (w skali szarości) jako kolorowanie tego obrazu w miejscach przepływu (barwy określonej skali pokazują prędkości przepływu) - rys. 9.4a. Lepszą jakość mają urządzenia do prezentacji stosowane w systemach rezonansu magnetycznego, podczas gdy do obserwacji badań CT używany jest zazwyczaj sprzęt wysokiej jakości, umożliwiający dokładną obserwację danych obrazowych z wykorzystaniem całej ich dynamiki (przesuwane okno). Obrazy scyntygraficzne prezentowane są często w systemach MN z wykorzystaniem interpolacji i filtracji w celu poprawy wizualnej jakości obrazów oryginalnych. Wprowadza się też zazwyczaj pseudokolorowanie (rys. 9.4b).



Rys. 9.4. Skale barwne w obrazowaniu medycznym: a) badanie USG serca - kolorowy doppler, b) pseudokolorowane PET-owych badań mózgu.

Charakterystyka pracy specjalisty

W wielu przypadkach obecność zmian patologicznych w analizowanym obrazie nie jest oczywista i, aby ją wykryć, konieczna jest wiedza, doświadczenie i uważna praca obserwatora-specjalisty. Zarówno sposób określania patologii, jak i stosowane kryteria zależą od tak wielu czynników (rodzaj badania, cechy pacjenta, dodatkowe informacje z innych badań, przebieg choroby itd.), że nie jest możliwe jednoznaczne scharakteryzowanie pracy specjalisty metodami cybernetycznymi. Stosowane są więc metody oparte na subiektywnej zdolności obserwatora do stwierdzania pewnych faktów istotnych diagnostycznie na podstawie analizowanych obrazów. Najpopularniejszą obecnie metodą, polegającą na rozpoznawaniu patologii w statystycznie istotnym zbiorze badań obrazowych przez zespół specjalistów danej dziedziny z różnych ośrodków medycznych, jest wyznaczanie krzywych ROC (*receiver operating characteristic*) [4]. W trakcie przeprowadzanych testów obok poprawnych decyzji, potwierdzających rzeczywistą obecność patologii w prezentowanym obrazie (decyzje prawdziwie pozytywne) oraz jej brak (decyzje prawdziwie negatywne), zdarzają się też wskazania błędne (fałszywe negatywne i fałszywe pozytywne), tym liczniejsze im silniejszy jest wpływ czynników pogarszających jakość obrazów (ograniczenia danej metody

obrazowania, wybór niewłaściwych parametrów systemu, słabe warunki obserwacji, niedoświadczenie czy nieuwaga obserwatora).

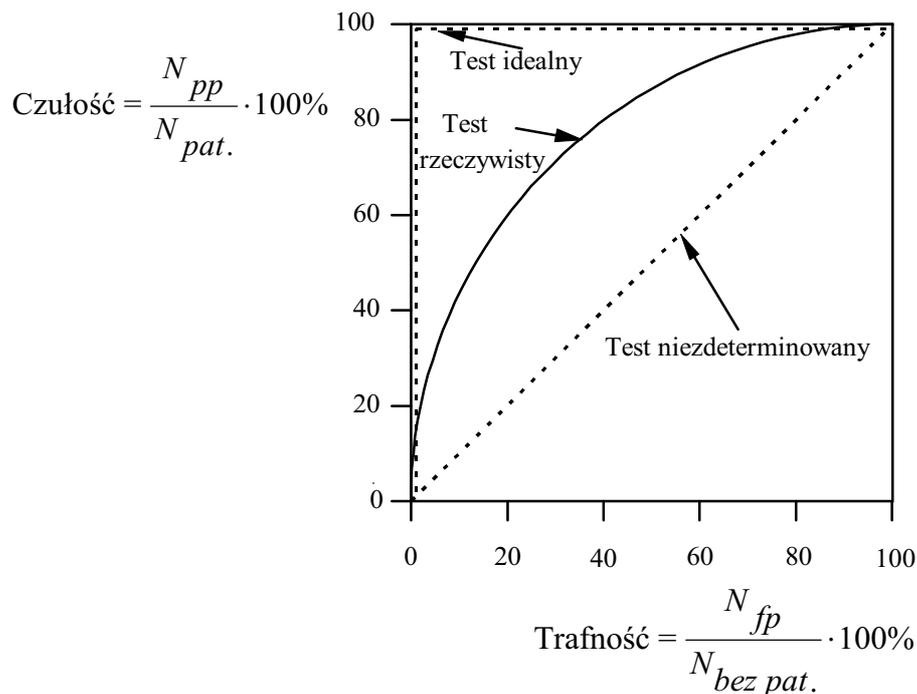
Krzywa ROC powstaje poprzez naniesienie wyników w układzie współrzędnych, w którym oś odciętych reprezentuje trafność (*specificity*), a oś rzędnych czułość (*sensitivity*). Czułość określona jest przez procentową zawartość ilości decyzji prawdziwie pozytywnych (czyli specjalista decyduje, że patologia jest obecna i rzeczywiście obraz zawiera patologię) wśród wszystkich werdyktów wydanych dla obrazów zawierających patologie - $\frac{N_{pp}}{N_{pat.}}$ [%]. Procentowy udział decyzji diagnostycznie prawdziwie negatywnych

(decyzja: brak patologii dla obrazu bez patologii) we wszystkich decyzjach dotyczących obrazów bez patologii - $\frac{N_{pn}}{N_{bez\ pat.}}$ [%] wyraża trafność decyzji, przy

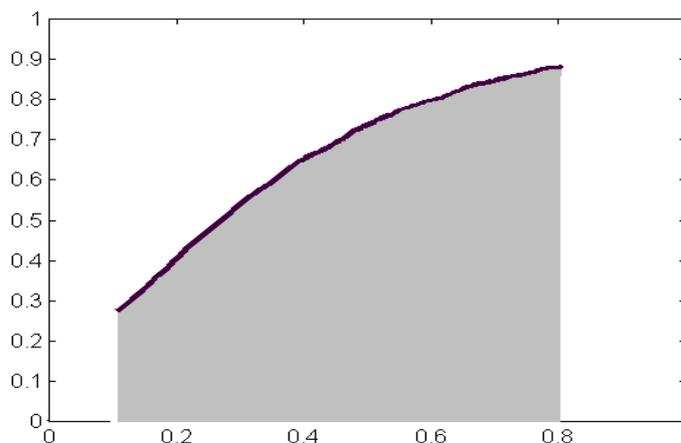
czym wówczas skala wartości trafności jest odwrócona i w początku układu współrzędnych znajduje się 100%. Częściej jako miara trafności decyzji stosowany jest procentowy stosunek ilości decyzji fałszywie pozytywnych (decyzja: jest patologia dla obrazu bez patologii) do ilości obrazów bez patologii - $\frac{N_{fp}}{N_{bez\ pat.}}$ [%], gdzie skala wartości nie jest odwrócona - 0% w

początku układu współrzędnych. Przykładowe krzywe przedstawione są na rys. 9.5. Oczywiście ideałem jest, by w czasie testu zarówno czułość jak i trafność wyniosła 100%, wówczas statystycznie (nie ma pewności co do konkretnego badania) liczba badań z patologią i bez patologii oceniona przez specjalistów pokrywa się dokładnie z wzorcem. Punkty z poszczególnych decyzji naniesione na wykres są często aproksymowane np. wielomianem czy funkcją sklejaną. Na podstawie wykreślonej krzywej ROC oblicza się różne wielkości charakterystyczne (kształt, nachylenie, pole powierzchni pod krzywą - rys. 9.6 itp.), które służą do porównań i ostatecznej oceny systemu obrazowania (wartości diagnostycznej tworzonych w nim obrazów). Wykorzystywane są też różne testy statystyczne do oceny jakości podejmowanych decyzji, a więc pośrednio wartości informacji prezentowanej przez obrazy danej klasy. Obok parametrycznych testów istotności, mających charakter jakościowy, stosowana jest czasami również parametrycznych hipotez istotności oparta na przedziałach ufności mająca charakter ilościowy. Przy pomocy tych testów porównywać można wartości pól pod krzywymi dla każdego lekarza, parametry funkcji aproksymujących lub też dokładne wartości punktów testowych

Zaletą tej metody jest względna niezależność od "dobrych intencji" obserwatora. Gdyby, przykładowo, obserwator wykazywał cechy zbytniego krytycyzmu i próbował zwiększyć liczbę wykrywanych patologii (rośnie czułość), liczba fałszywie pozytywnych (trafność) decyzji także wzrośnie.



Rys.9.5. Porównanie krzywych ROC dla przypadku idealnego, przykładowego testu praktycznego i dla przypadkowej selekcji, zupełnie niezdecydowanej użyteczną informacją.



Rys. 9.6. Obliczenie pola pod krzywą ROC jako elementu obliczeniowej oceny jakości na podstawie decyzji specjalistów.

Metody charakterystyki pracy specjalisty wykorzystywane są w technikach oceny dopuszczalnych stopni kompresji obrazów medycznych. Czynniki decydujące o wartości diagnostycznej obrazu stawiają określone wymagania odnośnie jego jakości. Wynikają one najczęściej z możliwości samej techniki obrazowania, jak też ze sposobu diagnozy w danym rodzaju badania.

W badaniach CR, CT, MRI decydujące w rozpoznawaniu patologii są często drobne fragmenty struktur, rozgałęzienia czy zgrubienia występujące w newralgicznych obszarach obrazu, zazwyczaj o ostrych krawędziach i

nieregularnych kształtach. Wymagana jest więc dobra jakość tych obrazów. W innych rodzajach badań przestrzenne relacje pomiędzy większymi strukturami oraz ich kształt stanowią zasadniczą informację diagnostyczną (głównie badania USG i NM), a więc silniejsze rozmycie czy wyższy poziom szumów nie musi powodować od razu degradacji wartości diagnostycznej tych obrazów.

Obok obserwacyjnych metod diagnozowania stosowane są także techniki obliczenia parametrów istotnych diagnostycznie bezpośrednio z obrazów, często w sposób zupełnie automatyczny, wykorzystujące różne metody przetwarzania obrazów (wyznaczania konturów, segmentacji itd.). Można więc oceniać wiarygodność diagnostyczną poprzez analizę wartości tych parametrów obliczanych z rekonstruowanych obrazów. Wyraźna zmiana tych wartości wskutek silnej redukcji informacji w zakresie małych średnich bitowych kompresji stratnej jest sygnałem, że przekroczono akceptowalną wartość stopnia kompresji [5].

Podsumowanie

Obrazy medyczne stanowią bardzo zróżnicowaną pod względem jakości klasę obrazów. Na podstawie analizy własności tych obrazów oraz czynników decydujących o wartości diagnostycznej można dokonać ogólnego podziału obrazów medycznych na dwie grupy:

- a) obrazy o 'niższej jakości' charakteryzujące się większym poziomem szumów, mniejszą rozdzielczością i łagodniejszymi kryteriami określenia ich wartości diagnostycznej (dotyczącymi głównie kształtu i przestrzennych relacji prezentowanych struktur); do grupy tej można zaliczyć przede wszystkim obrazy scyntygraficzne, USG i część obrazów MRI;
- b) obrazy o 'wyższej jakości' wyróżniające się znacznie niższym poziomem szumów, większą rozdzielczością oraz ostrymi kryteriami wartości diagnostycznej (związanej najczęściej z wiernym odtworzeniem drobnych szczegółów prezentowanych struktur oraz wielkością i kształtem charakterystycznych niewielkich obszarów); do obrazów tej grupy można zaliczyć wysokiej jakości obrazy MRI (głowy, układu nerwowego), CT oraz cyfrowe obrazy rentgenowskie CR.

9.2. Metody oceny jakości kompresowanych stratnie obrazów

W przypadku stratnych algorytmów kompresji pojęcie efektywności w znaczeniu przede wszystkim możliwie małej średniej bitowej czy dużego stopnia kompresji musi występować nierozłącznie w kontekście wnoszonych strat. Straty te rozumiane są przeważnie jako zniekształcenie danych rekonstruowanych w stosunku do zbioru danych oryginalnych, przy czym miara tych zniekształceń może być różna. Stąd też definiując efektywność stratnej metody kompresji trzeba podać obok przykładowo uzyskanego stopnia kompresji także odpowiadający mu poziom zniekształceń (np. jako błąd średniokwadratowy). Porównywanie skuteczności różnych algorytmów może się odbywać na poziomie średnich bitowych uzyskanych przy tej samej wartości zniekształcenia.

Można podejmować problem efektywności na różne sposoby. Najpierw buduje się coraz doskonalsze modele źródeł danych oraz miary zniekształceń, przy pomocy których można teoretycznie wyznaczać granicę efektywności kompresji metod stratnych - teoria stopnia zniekształceń. Potem coraz doskonalsze miary zniekształceń wykorzystuje się do oceny skuteczności technik stratnych, które w fazie konstrukcji wykorzystują często różne modele statystyczne przybliżające kompresowane zbiory danych.

Żaden układ obrazowania nie jest idealny

Istotną sprawą we wszelkich rozważaniach na temat sposobów oceny jakości rekonstruowanych obrazów jest jakość obrazu oryginalnego. Trzeba pamiętać o tym, że każdy system obrazowania ma swoje ograniczenia, nie wszystkie cechy prezentowanych obiektów są odzwierciedlane w rejestrowanych obrazach. Każdy system obrazowania można opisać funkcją przenoszenia (czasowo-częstotliwościową), która stanowi kompletny opis danego systemu. Przykładowo funkcja ta określa częstotliwość graniczną, a obiekty składające się z częstotliwości wyższych (w analizie fourierowskiej) mogą być w ogóle nie reprezentowane w zarejestrowanym obrazie. Ponadto występuje w tych obrazach szereg dodatkowych elementów (zniekształceń), które nie mają nic wspólnego z przedstawianymi obiektami. A skoro obraz oryginalny jest zniekształcony, to po pierwsze można stosując różne techniki przetwarzania w wielu przypadkach poprawić jego jakość, a po drugie sam proces stratnej kompresji, będący pewnego rodzaju filtracją, nie musi oznaczać automatycznie pogorszenia jakości obrazu oryginalnego. Łatwo można sobie wyobrazić sytuację, w której obraz rekonstruowany jest lepszej jakości od oryginału, bo w stratnej kompresji usunięto wysokoczęstotliwościowe pasmo szumów, znacznie pogarszające jakość obrazu oryginalnego. Formułując więc problem oceny jakości obrazów rekonstruowanych, właściwie powinniśmy porównywać ten obraz z domniemanym idealnym obrazem prezentowanych struktur (co jest oczywiście w praktyce szalenie trudne, czy wręcz niemożliwe), a przynajmniej pamiętać, że podlegający kompresji obraz oryginalny jest również zniekształcony, a pewne jego własności mogą nawet ulec poprawie podczas nieodwracalnej jego kompresji (co jest możliwe, szczególnie przy projektowaniu i przeprowadzaniu subiektywnej oceny jakości).

Miary jakości kompresowanych stratnie obrazów

Obraz uzyskany w wyniku przetwarzania obrazów jest "dobrej" jakości zazwyczaj wtedy, gdy według naszej percepcji wzrokowej wygląda przyjemnie (bez rzucających się w oczy zniekształceń), bądź też jest użyteczny do pewnych zastosowań. Nie istnieje niestety jedna skuteczna miara pozwalająca określić jakość odtwarzanego obrazu w każdym przypadku, stosowane są natomiast trzy zasadnicze metody określania jakości:

- **obliczeniowo obiektywne miary zniekształceń** (miary automatyczne) - wielkości skalarne bądź wektorowe wyznaczone automatycznie;
- **subiektywne miary jakości** (miary obserwacyjne) - psychowizualne testy porównawcze ze wskazywaniem obrazów o wyższej jakości lub też klasyfikacja w pewnej skali ocen (wyrażonej najczęściej liczbowo);

- **miary oparte na symulacji i analizie statystycznej** (miary symulacyjne), najbardziej złożone, dotyczą konkretnej aplikacji - możliwie wierna symulacja rzeczywistych warunków analizy obrazów i wnikliwa analiza statystyczna odpowiednio opracowanych wyników psychowizualnych testów klasyfikacyjnych.

Obiektywne miary zniekształceń i częściowo miary subiektywne stosowane są zazwyczaj do porównania efektywności kompresji różnych technik, podczas gdy metody symulacji i analizy statystycznej (SiAS) oraz bardziej rozbudowane miary subiektywne służyły do określania dopuszczalnych stopni kompresji.

Obiektywne miary jakości

Wobec szeregu oczywistych wad subiektywnych miar jakości obrazu ciągle istnieje ogromne zainteresowanie rozwojem obiektywnych miar ilościowych, w formie zarówno liczbowej jak i graficznej, zbieżnych z psychowizualną oceną jakości. Błąd średniokwadratowy (MSE), najpopularniejsza z miar obliczeniowych, nie najlepiej koreluje z subiektywną oceną jakości. Niewielkie przestrzenne przesunięcie obrazu powoduje duże liczbowe zniekształcenie MSE, przy praktycznie zupełnym braku wizualnego zniekształcenia. Odwrotnie, uśredniająca cecha tego kryterium powoduje, że nawet znaczne artefakty pojawiające się w obszarach istotnych diagnostycznie o niewielkich rozmiarach, mogą mieć bardzo niewielki wpływ na wartość MSE.

Obiektywna obliczeniowo miara jakości winna mieć trzy podstawowe cechy:

- łatwość obliczeniową,
- dużą korelację z jakością perceptualną (psychowizualna ocena jakości przez obserwatorów),
- podatność w analizie (bogactwo narzędzi do analizy oraz łatwość interpretacji).

Cecha podatności w analizie daje możliwość prostego powiązania optymalizacji procesu przetwarzania obrazu z perceptualnie najlepszą jakością wynikowych obrazów. Przykładem jest popularność metod średniokwadratowych, która wiąże się głównie z bogactwem teorii oraz metod numerycznych dostępnych w analizie i syntezie systemów, które zazwyczaj są optymalizowane pod kątem minimalizacji błędu średniokwadratowego. Jednocześnie przez proste ważenie lokalnych błędów w danej dziedzinie (najlepiej częstotliwościowej) można wprowadzić model 'korekcji psychowizualnej' takiej miary. Najlepszym rozwiązaniem byłoby posiadanie miary odpowiadającej subiektywnym kryteriom oceny, która mogłaby być włączona w fazę projektowania techniki kompresji. Wówczas można by np. zaprojektować proces kwantyzacji i przydziału bitów poszczególnym współczynnikiem transformaty ze względu na ich psychowizualne znaczenie, albo zastosować - jako element wstępny kompresji - filtrację wzmacniającą istotne pasma informacji zawartej w obrazie.

Do najczęściej stosowanych obiektywnych miar jakości obrazów należą:

- 1) Średnia różnica:

$$AD = \sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)] / M \cdot N \quad (9.2)$$

2) Zawartość strukturalna (ang. structural content):

$$AC = \sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2 / \sum_{x=1}^M \sum_{y=1}^N [\hat{f}(x,y)]^2 \quad (9.3)$$

3) Znormalizowana korelacja skrośna (ang. normalized cross-correlation):

$$NK = \sum_{x=1}^M \sum_{y=1}^N f(x,y) \cdot \hat{f}(x,y) / \sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2 \quad (9.4)$$

4) Jakość korelacji (ang. correlation quality):

$$CQ = \sum_{x=1}^M \sum_{y=1}^N f(x,y) \cdot \hat{f}(x,y) / \sum_{x=1}^M \sum_{y=1}^N f(x,y) \quad (9.5)$$

5) Maksymalna różnica (ang. maximum difference), zwana też szczytowym błędem bezwzględnym (ang. peak absolute error-PAE):

$$MD = \text{Max}\{|f(x,y) - \hat{f}(x,y)|\} \quad (9.6)$$

6) Wierność obrazu (ang. image fidelity):

$$IF = 1 - \left(\sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2 / \sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2 \right) \quad (9.7)$$

7) Laplasjanowy błąd średniokwadratowy (ang. laplacian mean square error):

$$LMSE = \sum_{x=1}^{M-1} \sum_{y=2}^{N-1} [O\{f(x,y)\} - \hat{O}\{f(x,y)\}]^2 / \sum_{x=1}^{M-1} \sum_{y=2}^{N-1} [O\{f(x,y)\}]^2, \quad (9.8)$$

gdzie $O\{f(x,y)\} = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4 \cdot f(x,y)$

8) Błąd średniokwadratowy (ang. mean square error):

$$MSE = \frac{1}{M \cdot N} \sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2 \quad (9.9)$$

9) Szczytowy błąd średniokwadratowy (ang. peak mean square error):

$$PMSE = \frac{1}{M \cdot N} \sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2 / [\max\{f(x,y)\}]^2 \quad (9.10)$$

10) Znormalizowany błąd bezwzględny (ang. normalized absolute error):

$$NAE = \sum_{x=1}^M \sum_{y=1}^N |f(x,y) - \hat{f}(x,y)| / \sum_{x=1}^M \sum_{y=1}^N |f(x,y)| \quad (9.11)$$

11) Znormalizowany błąd średniokwadratowy (ang. normalized mean square error):

$$NMSE = \sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2 / \sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2 = 1 - IF \quad (9.12)$$

12) Norma L_p (Minkowskiego):

$$L_p = \left\{ \frac{1}{M \cdot N} \sum_{x=1}^M \sum_{y=1}^N |f(x,y) - \hat{f}(x,y)|^p \right\}^{1/p}, p = 1, 2, 3, \dots \quad (9.13)$$

13) Ważona odległość (ang. weighted distance):

WD: każdy element różnicowego obrazu jest normalizowany z zastosowaniem normy L_1

14) Stosunek sygnału do szumu (ang. signal to noise ratio):

$$SNR = 10 \log_{10} \left(\frac{\sum_{x=1}^M \sum_{y=1}^N [f(x,y)]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2} \right) \quad (9.14)$$

15) Szczytowy stosunek sygnału do szumu (*peak signal to noise ratio*):

$$PSNR = 10 \log_{10} \left(\frac{\sum_{x=1}^M \sum_{y=1}^N [\max\{f(x,y)\}]^2}{\sum_{x=1}^M \sum_{y=1}^N [f(x,y) - \hat{f}(x,y)]^2} \right), \quad (9.15)$$

przy czym wartość $\max\{f(x,y)\}$ jest zwykle stała (niezależna od obrazu) i wynosi np. 255 dla danych 8-mio bitowych

16) Wykresy Hosaka: graficzna miara jakości obrazów. Powierzchnia oraz kształt wykresów określają rodzaj oraz stopień degradacji jakości rekonstruowanego obrazu.

17) Histogram: inna graficzna miara jakości obrazu. Przedstawia dyskretny rozkład prawdopodobieństwa wartości pikseli w różnicowym obrazie.

$f(x,y)$ i $\hat{f}(x,y)$ oznaczają wartości pikseli odpowiednio obrazu oryginalnego i zrekonstruowanego.

Systemy HVS

Próbowano na różne sposoby (wykorzystując wiedzę o ludzkich zdolnościach psychowizualnych) stworzyć metodę oceny stopnia i rodzaju zniekształceń rekonstruowanego obrazu o własnościach zbliżonych do układu percepcji wzrokowej człowieka, tzw. systemy HVS (*human visual system*). Jako miary wykorzystuje się wówczas najczęściej różne odmiany normy Minkowskiego z liniowymi współczynnikami, odpowiadającymi własnościom HVS. Można także dokonać transformacji obrazu oraz liniowego ważenia w dziedzinie przestrzennych częstotliwości i w ten sposób uwzględnić perceptualne znaczenie zniekształceń.

Przykładem jednego z modeli HVS jest zaproponowany przez Nilla model związany z kosinusową transformatą o następująco zdefiniowanych funkcjach:

$$H(r) = \begin{cases} 0.05e^{r^{0.554}}, & \text{dla } r < 7 \\ e^{-9[|\log_{10} r - \log_{10} 9|]^{2.3}}, & \text{dla } r \geq 7 \end{cases}, \quad (9.16)$$

gdzie $r = (u^2 + v^2)^{1/2}$, a u oraz v są współrzędnymi w dziedzinie transformaty.

Wykorzystanie modelu HSV w konstruowaniu obiektywnych miar jakości może przynieść niekiedy poprawę korelacji miary z wrażeniem obserwatorów. Modyfikacja np. miary *NMSE* wykorzystująca model Nilla przedstawia się następująco:

$$NMSE = \sum_{u=1}^M \sum_{v=1}^N H\{(u^2 + v^2)^{1/2}\}^2 \cdot [k(u,v) - \hat{k}(u,v)]^2 / \sum_{u=1}^M \sum_{v=1}^N [H\{(u^2 + v^2)^{1/2}\} \cdot k(u,v)]^2 \quad (9.17)$$

gdzie $k(u,v)$ i $\hat{k}(u,v)$ - współczynniki w dziedzinie kosinusowej transformaty przed i po kwantyzacji.

Bardzo trudnym zadaniem jest więc wiarygodne porównanie jakości rekonstruowanych obrazów wyłącznie przy pomocy metody obiektywnej. Pojedyncza wartość skalarna nie może opisać szeregu różnorodnych zniekształceń. Z kolei graficzne miary jakości (histogram obrazu różnicowego, wykresy Hosaka [2], miara Eskicioglu [3] i wiele innych) pozwalają lepiej rozróżnić zarówno rodzaj zniekształceń, jak też ich wielkość i w połączeniu z miarami numerycznymi mogą dać lepszą wykładnię jakości, są jednak dużo bardziej czasochłonne i trudne do porównań.

Ponieważ ocena jakości obrazów przy pomocy miar obiektywnych wykazuje znaczący poziom korelacji z oceną psychowizualną, może być ona wykorzystana do ogólnych porównań efektywności stratnych metod kompresji obrazów medycznych, przy czym w kwestiach bardziej szczegółowych konieczne jest wspomaganie tej metody oceną subiektywną.

Wykresy Hosaka

Wśród miar graficznych wykresy Hosaka wydają się spełniać dobrze zasadnicze ich zadania w stosunku do skalarnych miar obliczeniowych, które polega na rozszerzeniu ilości dostępnej informacji o charakterze i wielkości zniekształceń przy jednoczesnym zachowaniu klarowności testów porównawczych. Miara Hosaka jest obiektywną obliczeniowo miarą porównawczą, pozwalającą zasadniczo określić jakość rekonstrukcji wartości pikseli obrazu oryginalnego, a także poziom szumu wprowadzony przez daną metodę przetwarzania obrazu. Pojęcia te należy traktować dosyć umownie (wierność rekonstrukcji i szumy), bo przybliżające je miary wykorzystują po prostu różnice wartości momentów pierwszego i drugiego rzędu rozkładów wartości pikseli podzielonych na kilka klas. W metodzie Hosaka podobnie jak w opisaney poniżej metodzie wykresów Eskicioglu wykorzystuje się do klasyfikacji prosty algorytm segmentacji drzewa czwórkowego.

Tak więc wyznaczanie wykresu Hosaka dla dwu obrazów o wartościach f' i f'' (np. oryginalnego i przetworzonego pewną metodą) opisane jest następującym algorytmem:

Algorytm 9.1. Wykresy Hosaka

1. Przyjmując kryterium jednorodności, takie że

$$\text{blok } B \text{ jest jednorodny, jeśli } \sigma_B^2 \leq T,$$

gdzie $\sigma_B^2 = \frac{1}{MN} \sum_{(i,j) \in B} (f(i,j) - \mu_B)^2$ - wariancja w tym bloku, a T jest założoną wartością progu (często $T=100$), dokonujemy podziału całego obszaru obrazu na bloki $B: 2^i \times 2^i, i=0, \dots, n$, przy czym $n=4$ (najczęściej).

Stosując metodę segmentacji drzewa czwórkowego tworzymy kilka klas kwadratowych bloków o rozmiarach boku 1, 2, 4, 8, 16 (...).

1. Obliczanie średniej w każdej klasie, a następnie różnicy tych średnich i średniej wszystkich klas: klasy C_i , gdzie $i=0, 1, \dots, n$, zawierają bloki o rozmiarach odpowiednio $2^i \times 2^i$.

$$\mu_i = \frac{1}{|C_i|} \sum_{B \in C_i} \mu_B, \quad i=0, 1, \dots, n, \quad (9.18)$$

$$\mu = \frac{1}{n+1} \sum_{i=0}^n \mu_i, \quad (9.19)$$

$$d\mu_i = \mu_i - \mu \quad (9.20)$$

2. Obliczanie średniej wartości odchylenia standardowego w każdej klasie (oczywiście z pominięciem klasy 1):

$$\sigma_i = \frac{1}{|C_i|} \sum_{B \in C_i} \sigma_B, \quad i=1, \dots, n. \quad (9.21)$$

3. Tworzenie dwóch wektorów cech: różnicowych wartości średnich i wartości odchylenia standardowego każdej z klas. Wektory te tworzymy najpierw dla obrazu oryginalnego, potem dla rekonstruowanego, a następnie tworzymy dwa różnicowe wektory cech:

$$(d\mu_0, d\mu_1, \dots, d\mu_n), (\sigma_1, \dots, \sigma_n) - \text{dla obrazu oryginalnego}$$

$$(d\mu_0', d\mu_1', \dots, d\mu_n'), (\sigma_1', \dots, \sigma_n') - \text{dla obrazu rekonstruowanego}$$

Na ich podstawie tworzymy dwa wektory różnicowe jako:

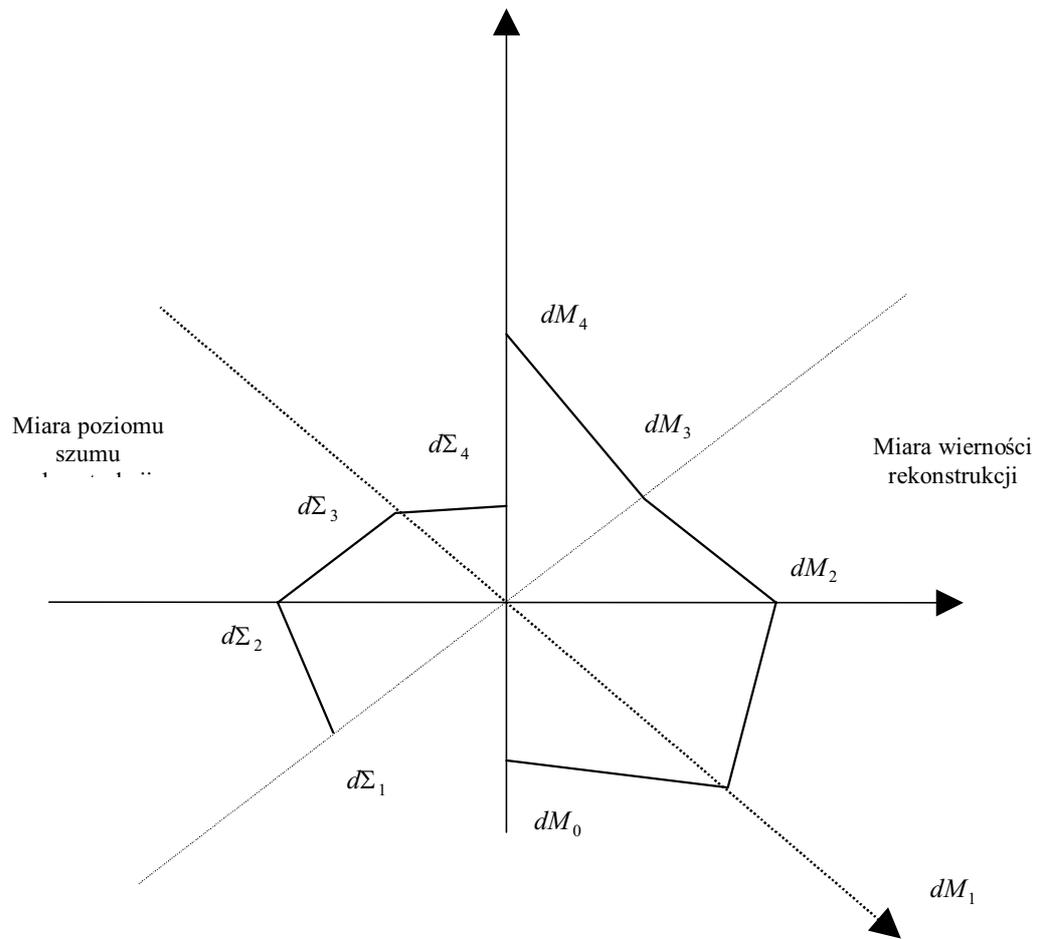
$$\vec{d}_M = (dM_0, dM_1, \dots, dM_n), \quad \vec{d}_\Sigma = (d\Sigma_1, \dots, d\Sigma_n), \quad (9.22)$$

gdzie

$$dM_i = |d\mu_i - d\mu_i'|, \quad d\Sigma_i = |\sigma_i - \sigma_i'|.$$

4. Wykreślanie różnicowych wektorów cech \vec{d}_M i \vec{d}_Σ w przestrzeni 2D.

Przykładowy wykres Hosaka został przedstawiony na rys.9.6. Prawa półpłaszczyzna wykresu zawiera wektor wartości średnich \vec{d}_M z odłożonymi na kolejnych półosiach składowymi wektora. Podobnie lewa część wykresu prezentuje wartości wektora wariancji \vec{d}_Σ .



Rys. 9.6. Przykładowy wykres Hosaka.

Miara Eskicioglu

Inną miarą graficzną jest wykreślania w postaci słupków miara Eskicioglu []. Zasadniczo mara ta jest mniej klarowna w ocenie jakości, jednak pozwala niezależnie ocenić jakość każdego obrazu, gdyż jest to miara absolutna (ang. univariate). Sposób wyznaczania miary Eskicioglu przedstawiony jest jako algorytm 9.2, a przykładowe wykresy przedstawia rys. 9.7.

Algorytm 9.2. Wyznaczanie miary Eskicioglu

1. Segmentacja drzewa czwórkowego. Stosuje się przy tym jedynie cztery klasy: 2, 4, 8, 16.
2. Dla każdej klasy określone są trzy cechy charakterystyczne:
 - ◆ liczba pikseli należących do bloków tej klasy/liczba wszystkich pikseli obrazu (256),
 - ◆ liczba różnych wartości pikseli/liczba możliwych wartości pikseli (256),
 - ◆ średnie odchylenie standardowe bloków danej klasy/przyjęte maksymalne odchylenie standardowe (8),
 - ◆ miara efektów blokowych (opcjonalnie):

$$EOBD = \{E[\Delta f(M, n)] + E[\Delta f(m, N)]\}^{1/2}, \quad (9.23)$$

gdzie:

$$\Delta f(M, n) = [f(M, n) - f(M + 1, n)]^2, \quad \Delta f(m, N) = [f(m, N) - f(m, N + 1)]^2.$$

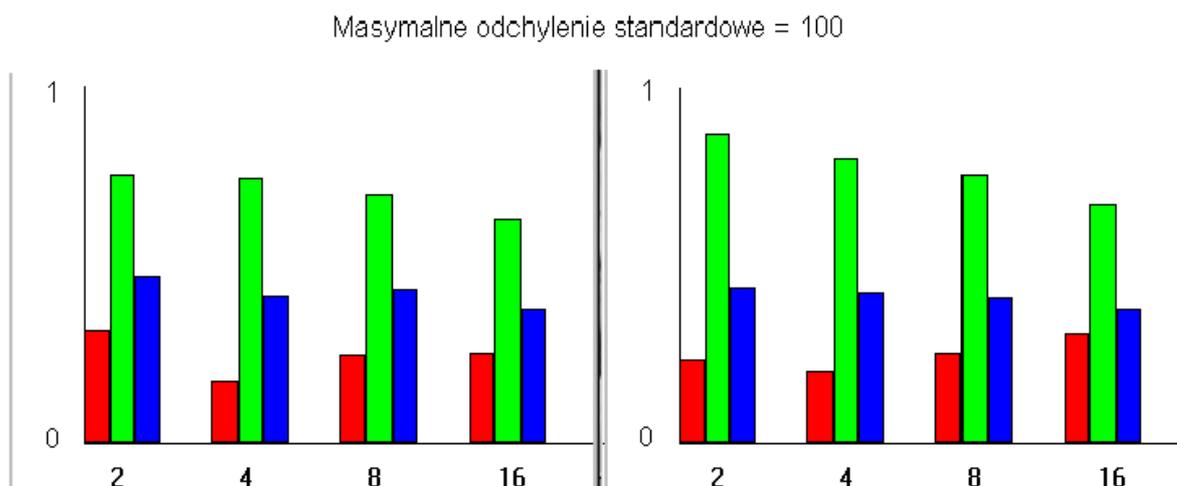
3. Wykreślanie słupków dla każdej cechy.

Przedstawiony poniżej przykład pokazuje sposób oceny jakości archiwizowanych obrazów przy pomocy miary Eskicioglu i wykresów Hosaka.

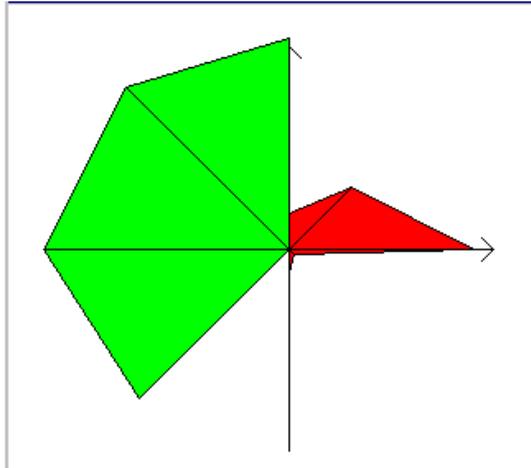
PRZYKŁAD 9.1. Obraz Lena kompresowano metodą stratną (falkową) uzyskując zamiast 8 bpp reprezentacji oryginalnej 0.1 bpp postaci skompresowanej. Na rysunku 9.7 przedstawiono obraz oryginalny i zrekonstruowany, a następnie wykresy Eskicioglu dla każdego z tych obrazów (rys.9.8) i wykres Hosaka (rys.9.9) pokazujący różnice między tymi obrazami.



Rys.9.7. Oryginalny obraz Lena (8bpp) oraz obraz rekonstruowany po stratnej kompresji falkowej (0.1bpp).



Rys. 9.8. Wykresy Eskicioglu dla obrazów z rys.9.7.



Rys.9.9. Wykres Hosaka obrazujący różnicę pomiędzy obrazem oryginalnym a rekonstruowanym z rys. 9.7.

Skala jakości obrazu (ang. Picture Quality Scale)

Jest to miara skalarna, która jednak jest budowana na szerokiej przestrzeni cech, która uwzględnia różne rodzaje zniekształceń wprowadzanych w procesie stratnej kompresji. Przestrzeń ta jest redukowana przy pomocy analizy składowych głównych, a następnie liniowa kombinacja składowych zredukowanej przestrzeni, optymalizowana metodą regresji na zgodność z oceną subiektywną, wyznacza globalną wartość zniekształcenia dla danego obrazu względem oryginału. Niektóre z cech przestrzeni pierwotnej są wyznaczone względem obrazu oryginalnego, inne zaś bezwzględnie. Są one opisane przez następujące współczynniki:

- Współczynnik F_1 (zniekształcenia losowe)

$$e_i(m,n) = i(m,n) - \hat{i}(m,n), \quad f_1(m,n) = [e_i(m,n) * w_{TV}(m,n)]^2, \quad (9.24)$$

gdzie $W_{TV}(f) = \frac{1}{1 + (f/f_c)^2}$, $f = \sqrt{u^2 + v^2}$, $f_c = 5.56$ cykla/stopień

$$F_1 = \frac{\sum_{m,n} f_1(m,n)}{\sum_{m,n} i^2(m,n)} \quad (9.25)$$

- Współczynnik F_2 (zniekształcenia losowe)

$x(m,n) = k \cdot i(m,n)^{1/2.2}$ (aproxymacja prawa Webera-Fechnera o czułości kontrastu)

$$e_w(m,n) = x(m,n) - \hat{x}(m,n), \quad s(\omega) = 1.5e^{-\sigma^2 \omega^2 / 2} - e^{-2\sigma^2 \omega^2}, \quad \text{gdzie } \sigma = 2, \omega = \frac{2\pi f}{60}, f = \sqrt{u^2 + v^2},$$

$$S_a(u,v) = s(\omega)O(\omega, \theta), \quad \text{gdzie } O(\omega, \theta) = \frac{1 + e^{\beta(\omega - \omega_0)} \cos^4 2\theta}{1 + e^{\beta(\omega - \omega_0)}}$$

$$\theta = \tan^{-1}(u/v), \beta = 8, f_0 = 11.13 \text{cykli/stopień}$$

$$f_2(m,n) = I_T(m,n)[e_w(m,n) * s_a(m,n)]^2$$

$$F_2 = \frac{\sum_{m,n} f_2(m,n)}{\sum_{m,n} \hat{i}^2(m,n)} \quad (9.26)$$

- Współczynnik F_3 (efekty blokowe)

$$f_{3h}(m,n) = I_h(m,n) \Delta_h^2(m,n), \quad \Delta_h(m,n) = e_w(m,n) - e_w(m,n+1), \quad F_{3h} = \frac{1}{N_h} \sum_{m,n} f_{3h}(m,n)$$

$$F_3 = \sqrt{F_{3h}^2 + F_{3v}^2} \quad (9.27)$$

- Współczynnik F_4 (błędy skorelowane)

$$f_4(m,n) = \sum_{(k,l) \in W} |r(m,n,k,l)|^{0.25},$$

$$r(m,n,k,l) = \frac{1}{n-1} \left[\sum e_w(i,j) e_w(i+k,j+l) - \frac{1}{n} \sum e_w(i,j) \sum e_w(i+k,j+l) \right]$$

$$F_4 = \frac{1}{MN} \sum_{m,n} f_4(m,n) \quad (9.28)$$

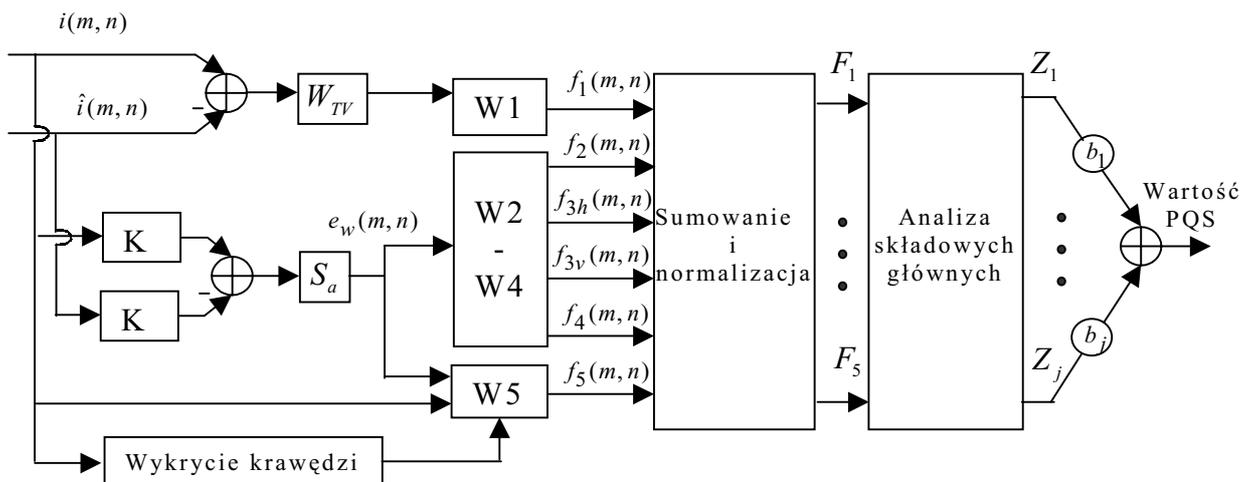
- Współczynnik F_5 (błędy strukturalne)

$$S_h(m,n) = e^{\{-0.04V_s(m,n)\}}, \quad V_h(m,n) = \frac{|i(m,n-1) - i(m,n+1)|}{2}$$

$$f_5(m,n) = I_M(m,n) |e_w(m,n)| (S_h(m,n) + S_v(m,n))$$

$$F_5 = \frac{1}{N_K} \sum_{m,n} f_5(m,n) \quad (9.29)$$

Schemat blokowy całego schematu metody PQS przedstawia rysunek 9.10.



Rys.9.10. Schemat metody oceny jakości obrazów PQS; W-F oznacza aproksymację prawa Webera-Fechnera o czułości kontrastu, W1 - W5 to algorytmy wyznaczania wartości pięciu współczynników.

Miary obserwacyjne (subiektywne)

Ponieważ ostatecznym użytkownikiem obrazów są ludzie, można skonstruować sposób oceny jakości wykorzystujący opinie obserwatorów. Każda ludzka opinia jest jednak zagrożona pewnym subiektywizmem, toteż kluczowym zadaniem przy opracowywaniu miar obserwacyjnych jest minimalizacja czynnika subiektywizmu (wynikającego z samej natury tych metod) związanego z decyzjami poszczególnych osób. Z drugiej jednak strony to specjaliści w danej dziedzinie wykorzystujący rozpatrywane zbiory danych wiedzą najlepiej, co decyduje o jakości obrazu, jakie cechy obrazu są brane pod uwagę przy jego analizie i to oni potrafią najlepiej sformułować kryteria przydatności obrazów oraz według nich przeprowadzić proces oceny jakości.

W przeprowadzanych testach wykorzystuje się przeważnie dwie grupy obserwatorów: ekspertów z danej dziedziny lub też grupę przygodnych ludzi, zupełnie przypadkową. Można też do przeprowadzenia testu zaprosić grupę specjalistów od analizy obrazów, znających ogólne czynniki, które decydują o jakości obrazu.

Subiektywna ocena jakości rekonstruowanych obrazów może być przeprowadzana na wiele sposobów. Istnieją dwa zasadnicze rodzaje miar subiektywnych:

- miary absolutne (bezwzględne): obserwatorzy stosownie do jakości obrazu umieszczają go w odpowiedniej kategorii według przyjętej skali ocen,
- miary porównawcze: obserwatorzy klasyfikują obrazy z danej grupy na podstawie wzajemnych porównań jakości poszczególnych obrazów tej grupy.

Dla miar absolutnych stosowana skala ocen winna zawierać skalę liczbową i skojarzony z każdą liczbą opis słowny, który możliwie trafnie wyrazi różne kategorie możliwych ocen obrazów danego typu (w zależności od aplikacji). W odpowiednio przygotowanych warunkach zbiorów obrazów jest prezentowany obserwatorom, którzy oceniają je w pewnej skali (rys. 9.11). Na podstawie ocen cząstkowych poszczególnych osób biorących udział w teście obliczana jest średnia ocena grupy obserwatorów według zależności:

$$R = \frac{\sum_{k=1}^K s_k n_k}{\sum_{k=1}^K n_k}, \quad (9.30)$$

gdzie K - liczba kategorii w przyjętej skali ocen, s_k - ocena związana z k -tą kategorią, n_k - liczba ocen specjalistów przypisanych danej kategorii. Przykładowo skala ocen A z rys. 9.11 ma pięć kategorii ocen z odpowiednim opisem, natomiast wartości s_k wynoszą kolejno: $s_1 = 5$, $s_2 = 4$, $s_3 = 3$, $s_4 = 2$, $s_5 = 1$.

Sposób prezentacji obrazów winien być tak zaprojektowany, by zminimalizować wpływ wszelkich czynników zmniejszających obiektywność ocen (efektu uczenia, skojarzeń podobieństwa lub porządku wyświetlania

itd.). Następnie przeprowadzana analiza statystyczna polega najczęściej na wyznaczeniu wartości średniej zebranych ocen, tzw. oceny średniej oraz wariancji zbioru tychże wartości. Różnorodność rozwiązań dotyczy głównie zakresu liczbowego stosowanej skali ocen oraz opisu każdego poziomu skali (są nieraz stosowane skale bez opisu słownego). W przypadkach konkretnych aplikacji opis ten może zawierać obok cech psychowizualnej oceny jakości obrazu także charakterystykę pewnych cech obrazu, szczególnie istotnych z punktu widzenia np. diagnozy.

A:	B:	
5. Wspaniały	7. Najlepiej	
4. Dobry	6. Wyraźnie powyżej średniej	
3. Zadawalający (możliwy)	5. Nieznacznie powyżej średniej	
2. Słaby	4. Średnio	
1. Nie satysfakcjonujący (zły)	3. Nieznacznie poniżej średniej	
C:	2. Wyraźnie poniżej średniej	
1. Niezauważalne (niedostrzegalne)	1. Najgorzej	
2. Lekko zauważalne (dostrzegalne)	D:	
3. Wyraźnie zauważalne (dostrzegalne), ale tylko lekkie osłabienie	3. Znacznie lepiej	
4. Osłabienie nie przyjemne	2. Lepiej	
5. Nieco nieprzyjemne	1. Nieznacznie lepiej	
6. Wyraźnie nieprzyjemne	0. Tak samo	
7. Ekstremalnie nieprzyjemne (niewłaściwe)	-1. Nieznacznie gorzej	
E:	-2. Gorzej	
5. Nieostrzegalne	-3. Znacznie gorzej	
4. Dostrzegalne ale nie dokuczliwe	F:	
3. Nieco dokuczliwe	10, 9	Bardzo dobry
2. Dokuczliwe	8, 7	Dobry
1. Bardzo dokuczliwe	6, 5, 4	Zadawalający
	3, 2	Zły
	1, 0	Bardzo zły

Rys. 9.11. Przykładowe skale ocen jakości obrazów stosowane w psychowizualnych testach miar subiektywnych.

Metody porównawcze oceny jakości można podzielić na trzy podstawowe kategorie:

- porównywanie z oryginałem,
- porównywanie dwóch obrazów,
- porównywanie wielu obrazów.

Niektóre oceny są dokonywane poprzez porównanie przez obserwatora obrazów zniekształconych po stratnej kompresji z oryginałem i określenie stopnia podobieństwa lub niepodobieństwa tych obrazów w pewnej skali możliwych ocen. Przykładowo, radiolodzy mogą oceniać jakość rekonstruowanych obrazów w skali 0-100 poprzez porównanie ich z jednocześnie obserwowanym oryginałem, przy czym 0 oznacza zupełną nieakceptację obrazu do celów klinicznych, natomiast wartość 100 - identyczność z oryginałem.

Porównywanie jakości dwóch obrazów jednocześnie obserwowanych i podjęcie decyzji klasyfikacyjnej dla tej pary obrazów jest podstawą drugiej grupy metod porównawczych. Przykładowo rozpatrzmy grupę pięciu obrazów np. skompresowanych w tym samym stopniu pięcioma różnymi technikami, którym przypisano losowo kolejne litery A, B, C, D, E. Proces klasyfikacji rozpoczyna się od porównania jakości obrazów A i B. Załóżmy, że obserwator ustala jako właściwą kolejność BA, czyli że obraz z literką B jest wyższej jakości niż A. Następnie prezentowane mu są jednocześnie obrazy A i C i ustala on kolejność powiedzmy CA. W tym przypadku należy przeprowadzić teraz porównanie obrazów B i C. Przyjmijmy, że obserwator zdecydował BC, czyli ostateczna klasyfikacja jest następująca: BCA. Teraz analizowana jest jakość obrazu D poprzez kolejne porównanie z obrazami, zaczynając od obrazów o niższej jakości i następuje ewentualne przesuwanie obrazu D w górę klasyfikacji itd.

Innym rozwiązaniem jest schemat klasyfikacji zbioru obrazów w kolejności od obrazu o najwyższej jakości do obrazu o najniższej jakości przy jednoczesnej obserwacji całego zbioru obrazów. W przypadku, gdy różnice pomiędzy obrazami są praktycznie niezauważalne, klasyfikacja będzie dość przypadkowa, natomiast jednoznaczne decyzje wydzielające obrazy o wyraźnie zdegradowanej jakości są wskaźnikiem przekroczenia progu dopuszczalności strat.

Miary oparte na symulacji i analizie statystycznej

Najczęściej metody oceny jakości należące do tej grupy oparte są na analizie charakterystyki działania obserwatora ROC, która wywodzi się z teorii detekcji sygnału. Eksperci obserwujący odpowiednio przygotowane obrazy dokonują ich oceny, która jest zazwyczaj oparta na detekcji pewnych cech czy lokalnych własności obrazu. Wyniki ich binarnych decyzji (jest lub nie) dla wielu testowanych obrazów (dla wiarygodnej statystyki przeprowadza się zazwyczaj przynajmniej sto takich decyzji) nanoszone są w postaci punktów na charakterystyce ROC, przy czym każdy punkt wyraża prawdopodobieństwo prawdziwej i fałszywej decyzji kolejnego specjalisty. Wykorzystywane są też często wielostopniowe skale ocen. Przykładowo w skali pięciostopniowej kolejnym stopniom odpowiada następujący opis

słowny: pewna cecha jest zdecydowanie obecna, prawdopodobnie obecna, może obecna, prawdopodobnie nieobecna lub też definitywnie nieobecna. Wówczas wyrażone już w kategoriach prawdopodobieństwa poszczególne decyzje specjalistów pozwalają lepiej określić prawdopodobieństwo decyzji prawdziwej i fałszywej, podejmowanych przez danego obserwatora na podstawie obrazów rekonstruowanych.

Specjaliści uczestniczący w testach są wcześniej przygotowywani do podejmowania decyzji w przełożeniu na skalę ocen. Zapewnia się takie warunki testu, które z jednej strony jak najbardziej odpowiadają konkretnym warunkom pracy, z drugiej zaś ograniczają do minimum czynniki zakłócające obiektywną oceną (proces uczenia, skojarzenia, zmienne warunki obserwacji itd.). Wyniki decyzji poszczególnych specjalistów podlegają następnie analizie statystycznej w celu wyznaczenia sumarycznych wskaźników wyrażających ocenę jakości obrazów.

Praktyczny sposób zastosowania metod oceny jakości opartych na symulacji, gdzie podstawowym kryterium jest określenie wiarygodności diagnostycznej rekonstruowanych obrazów na zadawalającym poziomie przedstawiono w następnym rozdziale.

9.3. Ocena wiarygodności diagnostycznej obrazów medycznych

Cechami charakterystycznymi tych metod są przede wszystkim:

- duża złożoność i czasochłonność,
- wykorzystanie subiektywnych opinii lekarzy-specjalistów w danej dziedzinie, przy jednoczesnym dążeniu do maksymalnej obiektywizacji tych ocen,
- stworzenie warunków oceny jakości obrazów rekonstruowanych zbliżonych do codziennej praktyki lekarskiej.

Podstawowym problemem w ocenie przydatności stratnych technik kompresji jest duża trudność w ocenie rodzaju i ilości zniekształceń w rekonstruowanych obrazach. Przy stratnej kompresji obrazów medycznych szczególnie ważne jest określenie jakości, a przede wszystkim wiarygodności diagnostycznej rekonstruowanych obrazów, czyli wiernego odtworzenia wszystkich informacji istotnych diagnostycznie zawartych w obrazie oryginalnym. Wielu specjalistów-lekarzy wyraża się sceptycznie o możliwości zastosowania tych metod kompresji w medycznych systemach obrazowania i archiwizacji, głównie ze względu na dużą odpowiedzialność i ryzyko obniżenia jakości obrazów, a więc pogorszenia warunków diagnozy. Stąd też szczególnie istotne jest opracowanie takich miar wiarygodności diagnostycznej rekonstruowanych obrazów, które pozwolą określić wyraźne, bezpieczne granice dopuszczalnej redukcji informacji z obrazów oryginalnych w celu efektywnej ich archiwizacji i transmisji.

Metody oceny wiarygodności diagnostycznej obrazów oparte na symulacji i analizie statystycznej

Cechami charakterystycznymi tych metod są przede wszystkim: duża złożoność i czasochłonność, wykorzystanie subiektywnych opinii lekarzy-specjalistów w danej dziedzinie, przy jednoczesnym dążeniu do maksymalnej obiektywizacji tych ocen, stworzenie warunków oceny jakości obrazów rekonstruowanych zbliżonych do codziennej praktyki lekarskiej.

Krótką charakterystyka metod z ROC

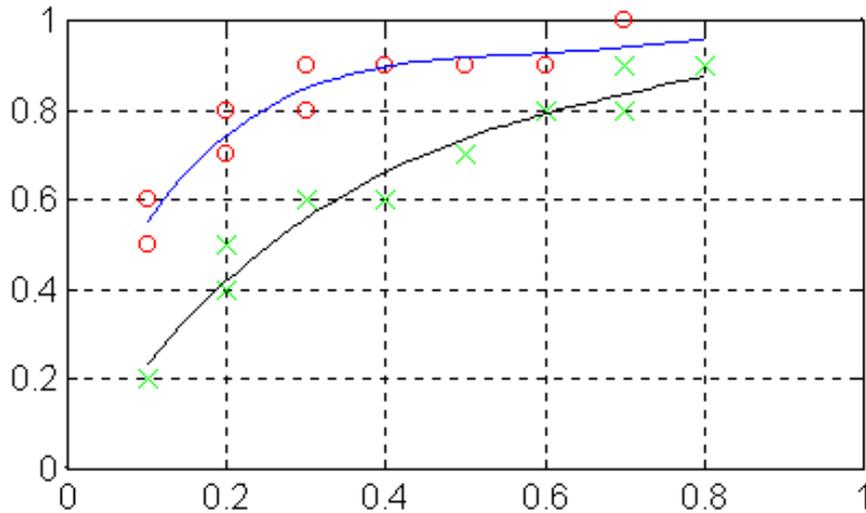
Metody oceny diagnostycznej wiarygodności obrazów medycznych wykorzystują najczęściej krzywą ROC. Eksperci dokonują diagnozy, korzystając najczęściej z pięciostopniowej skali ocen (1 - zdecydowana nieobecność pewnej patologii w obrazie, 5 - definitywna obecność patologii). Są oni wcześniej przygotowywani do podejmowania decyzji diagnostycznych w przełożeniu na skalę ocen. Ponadto, zapewnia się takie warunki testu, które z jednej strony jak najbardziej odpowiadają klinicznym warunkom pracy, z drugiej zaś ograniczają do minimum czynniki zakłócające obiektywną oceną (proces uczenia, skojarzenia, zmienne warunki obserwacji itd.). Wyniki diagnozy poszczególnych specjalistów są następnie zbierane razem, zgodnie z charakterem danych w celu wyznaczenia sumarycznych wskaźników wyrażających wiarygodność diagnostyczną.

PRZYKŁAD 9.2. Wykonano testy oceny jakości kompresowanych stratnie obrazów medycznych w warunkach jak najbardziej zbliżonych do rzeczywistych, przy czym zapewniono niezależność wydawanych ocen oraz zminimalizowano proces uczenia u lekarzy-specjalistów biorących udział w testach. Do analizy wykorzystano krzywą ROC. Przygotowano 20 obrazów radiografii cyfrowej, w tym 10 z nich zawierało patologie, które zostały poddane kompresji w stopniu 10 i 25, a po rekonstrukcji były obserwowane przez 10 specjalistów. Oczywiście w praktyce liczba wykonanych decyzji jest znacznie większa (wiarygodna statystycznie - przyjmuje się, że jeden punkt na krzywej winien być wynikiem przynajmniej stu decyzji), a uzyskane wartości prawdopodobieństw bardziej wiarygodne. W naszym przykładzie uzyskano następujące wyniki testu (tabela 9.3):

Tabela 9.3. Wyniki testu oceny wartości diagnostycznej obrazów kompresowanych w stopniu 10:1 i 25:1.

CR	lekarz	1	2	3	4	5	6	7	8	9	10
10:1	Npp	9	5	6	9	9	10	8	8	9	7
	Nfp	5	1	1	3	4	7	2	3	6	2
25:1	Npp	6	2	5	7	9	9	8	6	8	4
	Nfp	4	1	2	5	7	8	6	3	7	2

gdzie Npp - liczba decyzji prawdziwie pozytywnych, Nfp - liczba decyzji fałszywie pozytywnych, CR - stopień kompresji. Na podstawie tych wyników wykreślono krzywą ROC dla obu stopni kompresji (rys. 9.11).



Rys. 9.11. Wykres krzywej ROC dla danych z przykładu.

Poniżej przedstawione zostały przykładowe narzędzia do analizy tak wyznaczonej krzywej ROC.

Prosty test ze statystyką U

Mamy: dwie duże, niezależne próby pobrane z populacji niekoniecznie normalnych, o nieznanymi wartościami średnich m_1 i m_2 i o nieznanymi, lecz równymi wariancjach σ_1^2 i σ_2^2 . Hipoteza: $H_0: m_1 = m_2$. Statystyka:

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad (9.31)$$

której rozkład przy prawdziwości hipotezy H_0 jest asymptotycznie normalny $N(0, 1)$, n_1, n_2 - liczebności prób.

Reguła postępowania jest następująca:

- przybliżamy wartości średnich i wariancji poprzez $\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$,
- ustalamy poziom istotności α
- rozpatrujemy, którą z hipotez alternatywnych należy wziąć pod uwagę:

$$H_1: m_1 \neq m_2, \quad H_2: m_1 > m_2, \quad H_3: m_1 < m_2$$

- jeśli wybieramy H_1 , to stosujemy test dwustronny i **odrzucaamy hipotezę** H_0 na korzyść hipotezy H_1 , gdy dla obliczonej wartości

$$u_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (9.32)$$

spełniona jest nierówność $|u_0| > u_\alpha$, gdzie u_α jest wartością statystyki U wyznaczoną z tablicy rozkładu normalnego, dla której $P(|U| \geq u_\alpha) = \alpha$.

Jeśli hipotezą alternatywną względem hipotezy H_0 jest hipoteza H_2 to stosujemy test prawostronny i odrzucamy H_0 na korzyść H_2 jeśli $u_0 > u_\alpha$ itd.

Test t-Studenta

W tych zastosowaniach można użyć następującego testu: weryfikujemy hipotezę, że średnie dwóch małych prób nie różnią się istotnie, czyli $H_0: m_1 = m_2$. Przyjmujemy, że próby są niezależne, a populacje w przybliżeniu normalne o nieznanym, lecz równym wariancjach (niestety niezbyt spełnione w tym przypadku). Podstawą testu jest statystyka:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (9.33)$$

Ma ona rozkład t-Studenta o $n_1 + n_2 - 2$ stopniach swobody

Można też stosować inne testy, np. F-Snedecora dotyczący weryfikacji hipotezy o równości wariacji empirycznych rozkładów, itp.

Jakkolwiek technika ROC jest dominująca przy określaniu wartości diagnostycznej obrazów medycznych, zawiera ona szereg słabych stron związanych z jej aplikacją. Pierwsza - to konieczność zamiany normalnego trybu diagnozowania w praktyce klinicznej na wyrażenie opinii w pewnej skali ocen. Następnie, ponieważ technika ROC została stworzona przy założeniu Gaussowskiego rozkładu szumów w zbiorze analizowanych danych, jej stosowanie do oceny danych obrazowych o zazwyczaj nie-Gaussowskim charakterze nasuwa pewne wątpliwości (istnieją pewne metody redukcji błędów wynikających z tych Gaussowskich założeń). Ponadto, wiele praktycznych zadań diagnostycznych, jakie stoją przed specjalistami, nie sprowadza się do decyzji dwupoziomowej: tak lub nie. W niektórych patologich występuje kilka nieprawidłowości w różnych fragmentach obrazu, a proces decyzyjny jest dużo bardziej złożony. Jak wówczas odpowiedzieć na pytanie o nieobecność anormalności w obrazie (parametr trafności), której z nich? Dla przypadków bliskich decyzjom dwupoziomowym można jeszcze zdefiniować trafność jako średnią ważoną decyzji dla każdej patologii w zależności od konkretnego obrazu. Przy złożonych zadaniach detekcji nieprawidłowości - stosowane są rozszerzenia techniki ROC.

Modyfikacja ROC

Jedno z takich rozwiązań zawiera nieco inny sposób oceny wiarygodności diagnostycznej, eliminujący część tych ograniczeń, pozwala, w opinii autorów, lepiej ocenić wartość diagnostyczną obrazów (CT i MR) poprzez większe odzwierciedlenie w testach rzeczywistych klinicznych zadań radiologów. Specjaliści obserwują obrazy i zaznaczają obecność pewnych anormalności np. powiększonych węzłów chłonnych w obrazie CT klatki piersiowej, lub też obecność guzków w płucach, przy czym liczba anormalności jest różna w poszczególnych obrazach testowych. Warstwa decyzyjna zostaje więc rozszerzona na kilka, czy nawet kilkanaście poziomów. Analizę tak otrzymanych wyników można przeprowadzić przy

pomocy dwu parametrów: czułości i przewidywanej wartości pozytywnej (ang. predictive value positive-PVP), gdzie $PVP = \frac{N_{pp}}{N_{pp} + N_{fp}}$. Jeśli obserwator

zakreśli wszystkie anormalności w obrazie, wówczas osiąga maksymalną wartość czułości 1, a jeśli mniej - odpowiedni ułamek wyraża czułość jego decyzji. Natomiast PVP określa szansę rzeczywistej obecności anormalności w zaznaczonych miejscach. Jeżeli więc ekspert byłby zbyt agresywny w wykrywaniu anormalności, wówczas dużej wartości czułości będzie towarzyszyć mała wartość PVP, a w przypadku zbytnej ostrożności wyniki będą dokładnie odwrotne. Następnie, na wykresach przedstawiane są średnie wartości czułości i PVP (oddzielnie) dla każdego stopnia kompresji badanych obrazów, które aproksymuje się odpowiednią funkcją, np. kwadratową funkcję sklejaną z kryterium minimalizacji błędu średniokwadratowego. Porównanie czułości i PVP dla różnych stopni kompresji przeprowadzono przy pomocy testu t-Studenta z wykorzystaniem rozkładu permutacji dwuelementowych (nazywanego czasami testem Behrensa-Fishera [6]. Test ten nadaje się do danych, które nie mają gaussowskiego charakteru.

Założmy: specjalista 1 określa jakość N obrazów należących do dwóch poziomów A i B (np. obrazy kompresowane w dwu różnych stopniach). Obrazy te należą do dziewięciu grup: bez patologii, z jedną patologią, z dwoma patologiami, ..., z ośmioma patologiami. Przez N_i oznaczmy liczbę obrazów i -tej grupy, a Δ_{ij} niech reprezentuje różnicę wartości czułości (lub PVP) dla j -tego obrazu i -tej grupy oglądanego na dwóch poziomach jakości.

Niech $\bar{\Delta}_i$ będzie średnią różnicą według równania:

$$\bar{\Delta}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \Delta_{ij}. \quad (9.34)$$

Teraz zdefiniujmy:

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (\Delta_{ij} - \bar{\Delta}_i)^2, \quad (9.35)$$

a następnie t statystyka Behrensa-Fishera dana jest przez równanie:

$$t_{BF} = \frac{\sum_{i=1}^{N_i} \bar{\Delta}_i}{\sqrt{\sum_{i=1}^{N_i} \frac{S_i^2}{N_i}}} \quad (9.36)$$

Dla każdego z N obrazów można liczyć wartości testu z dwóch poziomów (A \rightarrow B i B \rightarrow A) lub nie (pełna liczba permutacji wynosi 2^N). Obliczenia t_{BF} wykonywane są dla całego rozkładu permutacji. Jeśli k jest liczbą wartości t_{BF} , które przekraczają wartość 'prawdziwą' (przypadek: ten sam obraz z poziomu A i B), wtedy wartość $\frac{(k+1)}{2^N}$ jest przyjmowana jako poziom istotności testów jednostronnych zerowej hipotezy, że jakość obrazów dla wyższego stopnia kompresji jest przynajmniej tak dobra jak obrazów niższego stopnia kompresji.

Jak widać poziom złożoności tych metod rośnie w miarę przybliżania metody oceny jakości do rzeczywistego procesu decyzyjnego lekarza. Staje się to jednak zbyt uciążliwe, by mogło być stosowane w praktyce.

Nowa metoda bez ROC:

Krótką charakterystyką metody zupełnie odchodzącą od narzędzie krzywej ROC w analizie statystycznej wyników oceny przedstawia się następująco:

- określenie złotego standardu: zgodny, osobisty, niezależny i osobny;
- obiektywna diagnoza powstaje na podstawie analizę obrazów analogowych, względem których ocenia się obrazy cyfrowe (także cyfrowy oryginał);
- nowy protokół oceny jakości obrazów, w którym lekarze wyrażają swe opinie w kategoriach jak najbardziej diagnostycznych;
- analiza statystyczna bez krzywych ROC: do zapisu wyników testu wykorzystano tablicę 2×2 postaci:

	III	dobrze	źle
dobrze		N(1,1)	N(1,2)
źle		N(2,1)	N(2,2)

gdzie I - może oznaczać obraz oryginalny analogowy a II - oryginał cyfrowy lub też I - oryginał cyfrowy, II - cyfrowy kompresowany. Decyzje **dobrze** znaczą zgodnie ze złotym standardem, a **źle** - niezgodnie.

Jeśli uzyskane tablice nie są diagonalne, to do oceny statystyki uzyskanych wyników stosuje się:

- test McNemara: porównanie $N(1,2)$ i $(N(1,2)+N(2,1))/2$ za pomocą rozkładu dwumianowego o parametrach $N(1,2)+N(2,1)$ i 0.5 lub
- test Fishera: liczymy statystykę $(N(1,2)-N(2,1))^2/(N(1,2)+N(2,1))$, która ma rozkład chi-kwadrat z jednym stopniem swobody.

Zastosowano także tablice zgodności decyzji radiologów:

	RTS	F/U	C/B	BX
RTS	12	0	5	0
F/U	0	0	0	0
C/B	3	0	12	6
BX	0	0	2	17

RTS - przypadkowy, negatywny, lub łagodny do powtórnego badania,

F/U - prawdopodobnie łagodny ale wymagający sześciomiesięcznej obserwacji,

C/B - potrzebne dodatkowe badania,

BX - biopsja.

Porównano badania analogowe i dyskretne, rozwiązania te są stosowane dla oceny jakości obrazów mammograficznych.

Perspektywa dalszych badań

Metody oceny diagnostycznej wiarygodności rekonstruowanych obrazów, a więc rzeczywistej ich jakości w punktu widzenia celu ich aplikacji, rozwijane są obecnie w kierunku możliwie pełnego wprowadzenia zrekonstruowanych obrazów w rzeczywiste warunki pracy klinicznej i wnioskowania o ich jakości na podstawie testów, które coraz silniej naśladują realną ocenę diagnostyczną prezentowanych struktur. W efekcie testy te stają się coraz bardziej złożone, pracochłonne i czasochłonne, drogie, toteż coraz trudniejsze do praktycznego wykorzystania.

Ten kierunek poszukiwań, jakkolwiek zrozumiały, zawiera w sobie jednak wiele niebezpieczeństw i może doprowadzić do rozwiązań zupełnie nieprzydatnych w praktyce, a zatem - nieistotnych. Diagnoza jest tylko środkiem służącym właściwej ocenie stanu zdrowia pacjenta i ewentualnemu podjęciu odpowiedniej metody leczenia. Może więc lepiej oceniać jakość rekonstruowanych obrazów na podstawie przebiegu procesu leczenia, albo wręcz jego wyników, czyli stanu zdrowia pacjentów po zakończeniu procesu terapeutycznego. Oczywiście, jest to myślenie mało rozsądne, bowiem same metody leczenia jak i wynik ich stosowania, jakkolwiek mogą być potwierdzeniem przydatności danej metody medycznego obrazowania i wykorzystywanych do diagnozy obrazów, zależą również od tak wielu innych, nieraz dużo bardziej istotnych czynników, tak że nikt nie jest w stanie ocenić w ten sposób jakości obrazów.

Problem określenia wiarygodności diagnostycznej jest więc niełatwy, a przecież istnieje jeszcze inny poważny problem. Metody szacujące tę wiarygodność oparte są na tzw. "złotym standardzie", który wyraża 'prawdę' diagnostyczną każdego oryginalnego obrazu (wykorzystywanego w teście) i służy jako baza do porównań wyników diagnoz dokonywanych na podstawie różnych wersji (po kompresji) tego obrazu. Złoty standard może być wyznaczany na wiele sposobów (standard zgodny, osobisty, niezależny, osobny), w zależności od tego, co tak naprawdę powinien wyrażać; czy osobiste przekonanie pojedynczego lekarza, czy też pewien kompromis pomiędzy kilku specjalistami oceniającymi później także obrazy rekonstruowane lub też zupełnie niezależnymi od późniejszych ocen. Koncepcja standardu osobnego proponuje w celu sformułowania prawdy obiektywnej o obrazie oryginalnym skorzystanie z innych badań (chirurgicznej biopsji, innych badań obrazowych), obserwacji pacjenta itd. Każde z tych sformułowań złotego standardu ma słabe strony i wprowadza pewne zafałszowanie w ocenę jakości obrazu, bądź też, mówiąc dokładniej, podnosi poziom subiektywności tej oceny, a więc jej zależności od przygotowania ekspertów, metody pracy klinicznej, rodzaju badań itd.

Niezaprzeczalny jest także fakt, iż redukcja informacji zawartej w obrazie w procesie stratnej kompresji może w pewnych przypadkach poprawić zarówno jakość tego obrazu, jak też jego wartość diagnostyczną, chociażby przez redukcję szumów w obrazach NM, osłabienie artefaktów w obrazach USG itp. Poprawa jakości obrazu w stosunku do obrazu oryginalnego nie była brana pod uwagę zarówno w obiektywnych metodach oceny jakości, jak też w

subiektywnych miarach wiarygodności diagnostycznej rekonstruowanych obrazów.

Bardzo istotne jest opracowanie doskonalszej niż istniejące, bardziej obiektywnej metody określania wiarygodności diagnostycznej rekonstruowanych obrazów w celu wyznaczenia dopuszczalnych stopni stratnej kompresji, warunkujących jej stosowanie.

Najlepszym, choć bardzo trudnym do realizacji rozwiązaniem, byłoby wyznaczenie złotego standardu, który na miarę dostępnych środków współczesnej medycyny stanowiłby **obiektywną diagnozę** rzeczywistości przedstawianej przez dany obraz, a następnie ocenianie względem tego standardu jakości obrazów zarówno oryginalnych, jak też rekonstruowanych. Takie kryterium oceny pozwoliłoby obiektywniej porównać skuteczność kompresji różnych technik, jak też w bezpieczniejszy sposób określić poziom dopuszczalnych stopni kompresji, jako np. nie wnoszący większych zniekształceń niż obraz oryginalny.

Najnowsze rozwiązania dotyczące metod oceny wiarygodności kompresowanych stratnie obrazów medycznych koncentrują się na najlepszym sposobie określenia złotego standardu, a także protokołu oceny jakości obrazów, w którym lekarze wyrażają swe opinie w kategoriach jak najbardziej diagnostycznych []. W analizie statystycznej zrezygnowano z krzywych ROC, a jako obiektywną diagnozę zastosowano analizę obrazów analogowych. Wartość diagnostyczna oryginalnych obrazów cyfrowych była oceniana względem tychże analiz, podobnie jak obrazów rekonstruowanych. Rozwiązania te są stosowane dla oceny jakości obrazów mamograficznych. Niestety, nie we wszystkich systemach obrazowania istnieje analogowy oryginał. Trzeba więc szukać innych rozwiązań.

Przykładowe miary wiarygodności

Poniżej przedstawiono dwa przykłady miar wiarygodności diagnostycznej kompresowanych obrazów o niewielkim stopniu złożoności.

PRZYKŁAD 9.3. Metoda subiektywnej oceny wiarygodności diagnostycznej obrazów oraz dopuszczalnych stopni kompresji opiera się na subiektywnej ocenie przeprowadzanej przez grupę radiologów, najlepiej specjalistów o dużym doświadczeniu z różnych ośrodków medycznych [].

Warunki obserwacji

Warunki prezentacji obrazów oryginalnych i rekonstruowanych winny odpowiadać rzeczywistym warunkom klinicznym, bądź też im dorównywać w przypadku obserwacji obrazów w innych warunkach.

Czas obserwacji i oceny wartości diagnostycznej obrazu jest nieograniczony, a decyzje poszczególnych specjalistów niezależne.

Sposób oceny wiarygodności diagnostycznej

Ocena w pierwszej fazie jest formułowana w sposób binarny, tzn. specjalista obserwując obraz i porównując go z oryginałem stwierdza jego akceptowalność (w przypadku zachowania wartości diagnostycznej) lub nieakceptowalność (w przypadku utraty lub zmniejszenia tej wartości). W

razie potrzeby można równocześnie obserwować inne obrazy rekonstruowane przy różnym poziomie stratności.

O akceptowalności decyduje poziom zachowania w obrazie informacji istotnych diagnostycznie, tzn. wygląd tych fragmentów obrazu zawierających rozgałęzienia, guzki, przewężenia, zgrubienia, wyraźne granice struktur itp., które decydują o diagnozie. Jeśli cechy obrazu, mniej lub bardziej ogólne, ulegają zniekształceniu, zagrażającemu poprawności interpretacji, wówczas powinna zapasć decyzja o nieakceptowalności.

Kryterium dopuszczalności

Zbiór ocen wszystkich specjalistów dotyczących każdego z obrazów pozwala na ustalenie jego oceny ostatecznej. Jeśli chociaż jeden ze specjalistów wydał opinię negatywną, wówczas obraz jest uznawany za nieakceptowalny, czyli jego wiarygodność diagnostyczna jest wątpliwa. Stąd też jasno wynika kryterium, określające stopień kompresji jako dopuszczalny w przypadku, gdy obraz rekonstruowany po kompresji o zadanym poziomie stratności uzyskał ostateczną ocenę pozytywną.

Charakterystyka ocen i ostateczna weryfikacja

Każdy oceniający proszony jest następnie o scharakteryzowanie powodów swych decyzji i wskazanie 'dowodów' degradacji obrazu. Pozwala to dokonać ostatecznej weryfikacji testu, porównać stosowane przez poszczególne osoby kryteria subiektywnej oceny wiarygodności diagnostycznej oraz stwierdzić przyczyny ewentualnych rozbieżności decyzji.

Przedstawiona metoda umożliwi ocenę wiarygodności diagnostycznej rekonstruowanych obrazów oraz określenie dopuszczalnych stopni kompresji przy mniejszych kosztach czasowych i organizacyjnych, głównie wskutek uproszczenia sposobu oraz kryteriów oceny.

Podstawową zaletą tej metody jest zastosowanie bardziej obiektywnego kryterium oceny wiarygodności diagnostycznej. Wykorzystano bowiem psychowizualne określenie jakości istotnych diagnostycznie fragmentów i cech obrazów (w tym też poziomu widoczności ewentualnych patologii), a nie samą tylko detekcję patologii (jednej lub wielu), która zależy jeszcze od kilku innych czynników (definicji patologii i sposobu jej określenia z obrazu, psychicznych uwarunkowań wynikających z konieczności podjęcia decyzji, dużo większego wpływu własnych doświadczeń itd.) zupełnie niezależnych od procesu kompresji.

Zastosowanie wymagających kryteriów pozwala wyznaczyć bardziej wiarygodne wartości dopuszczalnych stopni kompresji obrazów medycznych, które mogą być w większym stopniu akceptowane nawet przez największych sceptyków stosowania stratnych metod kompresji w medycynie.

Uzasadnienie swoich decyzji przez oceniających specjalistów i ich wspólna weryfikacja pozwala z jednej strony podnieść wartość ostatecznych wyników, z drugiej zaś stanowi kolejny krok ku obiektywizacji całej metody poprzez wskazanie konkretnych przyczyn utraty wiarygodności diagnostycznej kompresowanych obrazów.

Wyniki zebranych w teście ocen i opinii są następujące:

Obrazy MR

Ocenę wartości diagnostycznej różnego typu obrazów MR (w sposób opisany w p. 4.5) przeprowadzono w warunkach dających nieco lepsze możliwości obserwacji i analizy jakości obrazów niż warunki kliniczne (opinia lekarzy biorących udział w testach).

Wśród osób oceniających znalazło się trzech specjalistów-radiologów z dwu różnych ośrodków medycznych.

Wyniki obserwacji, przeprowadzonych przez dr. med. Piotra Dobkowskiego z Kliniki Diagnostyki Obrazowej na Bródnie w Warszawie, zilustrować najlepiej może cytat z jego własnego opracowania końcowego:

"Subiektywna ocena jakości obrazów MR po kompresji

Przedmiotem opracowania są obrazy różnych narządów otrzymane metodą rezonansu magnetycznego. Porównywano obrazy oryginalne z rekonstruowanymi o różnym stopniu kompresji. W ocenie kierowano się wiernością odtworzenia pierwotnego wyniku, dokładnością najmniejszych rozpoznawalnych szczegółów obrazu i obecnością (lub brakiem) artefaktów wprowadzonych przez proces kompresji.

Analizowano następujące obrazy:

1. Obraz mózgu i twarzoczaszki w przekroju strzałkowym. Obraz T1 - ważony zawiera nieprawidłową zmianę ogniskową o średnicy ok. 15 mm.
2. Obraz jamy brzusznej w przekroju poprzecznym na wysokości żył wątrobowych. Obraz wątroby z licznymi hypointensywnymi zmianami ogniskowymi w miększu.
3. Seria 5 kolejnych (sekwencja czasowa) obrazów sutków w przekroju poprzecznym, liczne niejednorodności w rozkładzie sygnału z obszarami hypo- i hyper-intensywnymi o charakterze ogniskowym. Tzw. badania dynamiczne.

Wyniki

Ad. 1. Obrazy po kompresji w stosunku 20:1 lub więcej wykazywały wyraźne artefakty w postaci wzoru "szachownicy" zakłócającego pierwotny rozkład natężenia sygnału (efekty blokowe - dop. autora). Istotne zastrzeżenia budzi również obraz skompresowany w stosunku 16:1 - zarysy poszczególnych struktur (np. zakrętów kory mózgu) są nieostre i trudne do prześledzenia, zwłaszcza w obszarach o małej kontrastowości sąsiadujących ze sobą szczegółów. Wewnętrzna struktura niewielkich fragmentów obrazu (np. zmiany ogniskowej) jest częściowo zatarta. Nie stwierdza się natomiast innych artefaktów. **Obrazy o stopniu kompresji 14 i mniejszym charakteryzowała jakość porównywalna z obrazem wyjściowym, nie skompresowanym. Nawet w dużym powiększeniu nie stwierdza się istotnych różnic w rozkładzie natężenia sygnału i kontrastowości.** Widoczne są wszystkie szczegóły możliwe do zauważenia w obrazie oryginalnym.

Ad.2. Obrazy po kompresji większej niż 22:1 są znacznie zubożone w szczegóły, zawierają artefakty w postaci wzoru "szachownicy" i charakteryzują się zniekształceniem konturów widocznych struktur (schodkowy przebieg linii krzywych). Reasumując ich jakość jest niedostateczna. Kompresja 19:1 zapewnia lepszą wierność w stosunku do oryginału, jednakże zarysy drobnych konturów (np. struktur naczyńowych) są nadal zniekształcone (w formie odcinka zamiast fragmentu krzywej). Pełną zgodność z obrazem wyjściowym zapewniają obrazy skompresowane w stosunku 16:1 lub mniejszym.

Ad.3. Kompresja w stosunku 32:1 lub większym uniemożliwia ocenę obrazu ze względu na zniekształcenie zarysów badanego narządu i mozaikową strukturę obrazu (analogicznie jak w przypadkach 1-2 występuje "szachownica", na tle której poszczególne elementy obrazu ulegają częściowemu zatarciu). Obrazy skompresowane w stosunku 25:1 i 22:1 charakteryzują się znacznie lepszą jakością, występujące nadal artefakty (tego samego rodzaju co opisane wyżej) dotyczą jedynie zarysu całego badanego narządu. Na obrazach tych poszczególne obszary hyperintensywne (np. elementy tkanki tłuszczowej) oraz hypointensywne (torbiele, zwapnienia) przedstawiają się prawidłowo - są czytelne i łatwo rozpoznawalne. Pełną wierność w stosunku do obrazu nie skompresowanego zachowują obrazy po kompresji w stosunku 19:1 - są one praktycznie nie do odróżnienia od wyjściowych.

Uwagi w p.3 dotyczą w stopniu jednakowym wszystkich elementów składowych (obrazów) porównywanych badań dynamicznych.

Podsumowanie

Na podstawie sformułowanych powyżej spostrzeżeń nasuwa się wniosek, że **kompresja w stosunku 14:1 zapewnia zachowanie dobrej jakości w przypadku każdego rodzaju badania**. Obrazy będące wynikami badań stosunkowo dużych obszarów i narządów można bez utraty czytelności skompresować w stosunku 16:1, a nawet 19:1. Wynika to z faktu, że najdrobniejsze szczegóły struktury anatomicznej nie odgrywają istotnej roli w ocenie takich obrazów. Ponadto, zwłaszcza w badaniach jamy brzusznej, obraz oryginalny zawiera mniejszą liczbę szczegółów z powodu artefaktów powstających już w procesie akwizycji obrazu (ruchy oddechowe i perystaltyczne)".

Ocena przeprowadzona przez pp. Grzegorza Rejnowskiego i Andrzeja Pytlewskiego, specjalistów radiologów z Instytutu Psychiatrii i Neurologii na ul. Sobieskiego w Warszawie, była zbliżona, chociaż nieco mniej krytyczna. Dopuszczalne według nich stopnie kompresji dla pierwszego obrazu wynoszą 16:1, a dla drugiego 17:1. Oto fragmenty ich opisu badań:

"Przeprowadzona ocena dotyczyła:

- utraty szczegółów obrazów pod wpływem procesu kompresji,
- zmian w kontraście i jasności obrazu uprzednio kompresowanego,

- strat jakości obrazów, mających znaczenie dla lekarza oceniającego obecność zmian patologicznych.

Określenie lokalizacji ogniska o podwyższonej intensywności sygnałów w badaniu mózgu było jednoznaczne i stopień kompresji obrazu nie miał wpływu na ocenę jego położenia. Także ocena jego wielkości nie ulegała zmianie. W miarę wzrostu stopnia kompresji, poszczególne struktury mózgu o tym samym stopniu szarości ulegały niewielkim zniekształceniom, które mogą powodować utrudnienie oceny diagnostycznej.

Dla niewielkich stopni kompresji w istotnych diagnostycznie fragmentach obrazu dało się zauważyć miejscową poprawę kontrastu, co sprawiało wrażenie lepszej jakości w stosunku do obrazu oryginalnego.

Obraz rozsianych, mnogich ognisk o obniżonej intensywności sygnałów, położonych w miększu wątroby, ulegał stopniowemu zatarciu w miarę rosnącej kompresji - co powodowało utratę wartości diagnostycznej."

Obrazy CT

Zrekonstruowane obrazy zostały wprowadzone do systemów rentgenowskich tomografów komputerowych i oceniane przez specjalistów w rzeczywistych warunkach klinicznych. Oceniano wyselekcjonowane wcześniej, trudne diagnostycznie (z dużą ilością drobnych szczegółów i plam) dwa obrazy klatki piersiowej oraz głowy.

W ocenie p. Andrzeja Stelmaszczyka, specjalisty rentgenologa z Centrum Onkologii na Ursynowie w Warszawie, dopuszczalne stopnie kompresji dla obrazów klatki piersiowej wynoszą odpowiednio 12 i 18, natomiast dla obrazu głowy 12. Głównymi czynnikami, degradującymi jakość obrazu przy silniejszej kompresji, są: rozmycie drobnych struktur oraz efekty blokowe.

Testy oceny wartości diagnostycznej tych samych obrazów przeprowadzano także w Zakładzie Medycyny Nuklearnej Centralnego Szpitala Klinicznego MSW. Dwóch specjalistów radiologów określiło dopuszczalne stopnie kompresji dla obrazu głowy na 12, a dla trudniejszego diagnostycznie obrazu klatki piersiowej na 18. Łagodniejsza ocena wynikała przede wszystkim z faktu obserwacji obrazów na kliszy, a także z braku w oryginalnym obrazie drobnych, a ważnych diagnostycznie struktur szczególnie wrażliwych na zniekształcenia. Według specjalisty radiologa p. Mirosława Nowickiego, w ocenie jakości rekonstruowanych obrazów głowy zasadniczą przyczyną degradacji wartości diagnostycznej jest postępujące rozmycie struktur wewnątrzczaszkowych. Specjaliści oceniali głównie zdjęcia obrazów na kliszy zgodnie z ich codzienną praktyką kliniczną.

Pani Profesor Irena Fleszarowa z II Zakładu Radiodiagnostyki AM w Warszawie, jakkolwiek zastrzegła, iż "wybrana warstwa przekroju poprzecznego głowy nie jest najlepsza do oceny działania kompresji" (oryginalny obraz głowy nie jest najlepszym reprezentantem tej klasy obrazów), to w ocenie porównawczej jakości obrazów głowy (oryginalnego i kompresowanych w różnym stopniu) nie stwierdziła istotnych różnic w przypadku obrazów kompresowanych w stopniu 8 i 12. Przy stopniu kompresji

równym 17 nastąpiło w opinii Pani Profesor "widoczne pogorszenie jakości obrazu części miękkich".

W przytoczonych opiniach lekarzy dokonano niewielkich zmian słownictwa, dostosowując tekst do używanej w pracy terminologii. Wszystkie podkreślenia pochodzą od autora.

Przedstawiona w rozdziale 3 metoda stratnej kompresji, wraz z powyższymi metodami zobiektywizowanej oceny wiarygodności diagnostycznej, może być skutecznie wykorzystywana w systemach archiwizacji medycznych danych obrazowych. Świadczą o tym przytoczone opinie specjalistów, którzy dokonali weryfikacji tej techniki pod kątem przydatności w praktyce klinicznej.

PRZYKŁAD 9.4. Skonstruowano miarę wektorową, czyli wielowymiarową miarę jakości sposób następujący. Do oceny lokalnych i globalnych zmian statystycznych służy pierwszy składnik - miara chi - kwadrat. Efekt blokowy, charakterystyczny dla standardu JPEG jest szacowany poprzez miarę artefaktów blokowych, a współczynnik wiarygodności diagnostycznej opisuje poziom wartości obrazów z punktu widzenia ich medycznej aplikacji. Miary te definiowane są w sposób następujący:

- miara chi-kwadrat (globalna i lokalna):

$$\chi_q^2 = \sum_{i=1}^{NM} \frac{(g_i - f_i)^2}{f_i} \quad (9.37)$$

- miara efektów blokowych:

$$REOBD = \{E[\Delta f(M, n)] + E[\Delta f(m, N)]\}^{1/2}, \quad (9.38)$$

gdzie

$$\Delta f(M, n) = [|f(M, n) - f(M + 1, n)| - |\hat{f}(M, n) - \hat{f}(M + 1, n)|]^2,$$

$$\Delta f(m, N) = [|f(m, N) - f(m + 1, N)| - |\hat{f}(m, N) - \hat{f}(m + 1, N)|]^2.$$

- współczynnik wiarygodności diagnostycznej:

wersja a)

$$DQC = \left| \frac{\delta_R}{\delta_M} \right|, \quad \text{for } \delta_R \leq \delta_M; \quad DQC = 1, \quad \text{for } \delta_R > \delta_M, \quad (9.39)$$

gdzie δ_R - błąd wyznaczenia wartości parametrów diagnostycznych (w porównaniu z oryginałem), δ_M - błąd metody obliczania parametrów diagnostycznych z obrazów.

wersja b)

$$DQC = \frac{N_0}{N_B}, \quad \text{for } N_0 \leq N_B; \quad DQC = 1, \quad \text{for } N_0 > N_B, \quad (9.40)$$

gdzie N_0 - liczba ocen negatywnych, N_B - graniczna liczba ocen negatywnych, która już jest nieakceptowalna.

Miara ta uwzględnia wyniki testów subiektywnej oceny wartości diagnostycznej, łącząc je z parametrami oceny obliczeniowej. W efekcie rozszerzony jest poziom informacji dostępnej do jakościowej analizy zniekształceń oraz możliwa jest bardziej wiarygodna ocena ilościowa zniekształceń.

Podsumowanie

Zakres zarysowanych zagadnień jest jedynie próbą ukazania ważkości zagadnienia oceny jakości obrazów cyfrowych. Obrazy te powstają w różnych systemach obrazowania i poddawane są wielorakim metodom przetwarzania w celu poprawy ich jakości bądź też usprawnienia warunków ich aplikacji. Na podkreślenie zasługuje fakt dużej wagi zagadnienia oceny jakości obrazów, jak również brak w pełni satysfakcjonujących rozwiązań. Stąd też powstaje obecnie wiele rozwiązań dotyczących konstrukcji coraz doskonalszych modeli systemu HVS, obliczeniowych miar jakości o cechach całkowitokrótnych (są to najczęściej miary wektorowe jak chociażby w [1]), opracowywane są nowe sposoby przeprowadzania testów psychowizualnej oceny jakości obrazów i bardziej przydatne metody statystycznej analizy wyników tych testów.

Inną stroną tego zagadnienia jest fakt, że w zdecydowanej większości prowadzonych badań z dziedziny przetwarzania obrazów jako miarę jakości rekonstruowanych obrazów stosuje się obliczeniowe miary skalarne: SNR lub MSE, przede wszystkim ze względu na ich prostotę i łatwość prowadzenia porównań. Prowadzi to oczywiście do konstruowania nowych metod przetwarzania pod kątem minimalizacji błędu średniokwadratowego, a bardziej wnikliwa analiza jakości wytwarzanych obrazów okazuje się potrzebna tylko w pewnych okolicznościach.

Istnieje bowiem pewna grupa krytycznych zastosowań, jak wspomniana w tym rozdziale kompresja obrazów medycznych czy też np. zagadnienie wprowadzenia nowych płyt obrazowych (ang. imaging plate) w cyfrowej radiografii, która z racji zastrzonych kryteriów (jakie narzuca np. ochrona zdrowia pacjenta) wymaga bardziej złożonych metod oceny jakości stosowanych metod przetwarzania danych oraz wynikowych obrazów. Ostatecznym celem nowych rozwiązań w tej dziedzinie jest nie tylko możliwie największa korelacja z psychowizualną oceną specjalisty przy jednoczesnej obiektywizacji sposobu oceny, ale wręcz wspomaganie ocen specjalisty i obiektywne wyznaczenie złotego standardu jako punktu odniesienia do wszelkich porównań i wyznaczenia progów akceptowalności nowych technik konstrukcji i przetwarzania obrazów.