Sygnatura: Pol J Radiol, 2005; 70(2): 47-53

Polish Journal of Radiology
www.PolRadiol.com
ORIGINAL ARTICLE

Otrzymano: 2004.05.18
Zaakceptowano: 2004.12.10

Evaluation of mammogram compression efficiency

Ocena efektywności kompresji mammogramów

Artur Przelaskowski[1], Paweł Surowski[2], Anna Kukuła[2]

[1] Instytut Radioelektroniki Politechniki Warszawskiej, Warszawa, Polska
[2] Zakład Diagnostyki Obrazowej Szpitala Wolskiego, Warszawa, Polska

Adres autora: Artur Przelaskowski, Instytut Radioelektroniki Politechniki Warszawskiej, ul. Nowowiejska 15/19, 00-665 Warszawa, Polska, emeil: arturp@ire.pw.edu.pl

Summary

Background: Lossy image coding significantly improves performance over lossless methods, but a reliable control of diagnostic accuracy regarding compressed images is necessary. The acceptable range of compression ratios must be safe with respect to as many objective criteria as possible. This study evaluates the compression efficiency of digital mammograms in both numerically lossless (reversible) and lossy (irreversible) manner. Effective compression methods and concepts were examined to increase archiving and telediagnosis performance.

Materials/Methods Lossless compression as a primary applicable tool for medical applications was verified on a set 131 mammograms. Moreover, nine radiologists participated in the evaluation of lossy compression of mammograms. Subjective rating of diagnostically important features brought a set of mean rates given for each test image. The lesion detection test resulted in binary decision data analyzed statistically. The radiologists rated and interpreted malignant and benign lesions, representative pathology symptoms, and other structures susceptible to compression distortions contained in 22 original and 62 reconstructed mammograms. Test mammograms were collected in two radiology centers for three years and then selected according to diagnostic content suitable for an evaluation of compression effects.

Results: Lossless compression efficiency of the tested coders varied, but CALIC, JPEG-LS, and SPIHT performed the best. The evaluation of lossy compression effects affecting detection ability was based on ROC-like analysis. Assuming a two-sided significance level of $p=0.05$, the null hypothesis that lower bit rate reconstructions are as useful for diagnosis as the originals was false in sensitivity tests with 0.04 bpp mammograms. However, verification of the same hypothesis with 0.1 bpp reconstructions suggested their acceptance. Moreover, the 1 bpp reconstructions were rated very similarly to the original mammograms in the diagnostic quality evaluation test, but the quality of 0.6 bpp and 0.1 bpp reconstructions was decreased.

Conclusions: The compression performance of the most effective reversible coders is rather unsatisfactory. The subjective rating with the diagnostic criteria of image quality was more sensitive to distortions caused by lossy compression compared with the pathology detection test. The observers constituted 14:1 as the accepted ratio of lossy wavelet compression for test mammograms. This is significantly higher than the mean ratio of 2:1 achieved with lossless methods.

Key words: medical image compression • subjective assessment • lossy compression • lossless compression • diagnostic quality • pathology detection • diagnostic accuracy

PDF file: http://www.polradiol.com/pub/pjr/vol_70/nr_2/5604.pdf

Wstęp

Rozwój nowoczesnych systemów radiografii cyfrowej spowodował wzrost zainteresowania efektywnymi metodami kompresji obrazów. Stale doskonalone systemy rejestracji wykorzystują detektory cyfrowe o coraz większej zdolności rozdzielczej lub też skanery pozwalające rozróżnić linie na poziomie przynajmniej 50 μm przy dużej dynamice danych. Rozmiar bitowej reprezentacji pojedynczego badania w przypadku mammografii sięga wtedy blisko 40 MB (megabajtów, przy rozmiarach obrazu 4000x5000 pikseli z dwubajtową wartością każdego z nich). Operowanie tak dużymi zbiorami danych nastręcza wiele problemów. Rosną lawinowo koszty nośników do archiwizacji badań, wydłuża się do niepraktycznych rozmiarów czas ich transmisji w sieciach lokalnych i globalnych. Szybkie przeszukiwanie baz radiologicznych systemów informacyjnych RIS, współpracujących z systemami archiwizacji i wymiany obrazów PACS, jest utrudnione. Tempo rozwoju technologicznego, dotyczącego większej pojemności nośników czy też przepustowości sieci transmisyjnych jest mniejsze od rosnących potrzeb radiologii cyfrowej. Gwałtownie zwiększa się bowiem liczba wykonywanych badań obrazowych, powstają nowe centra radiologiczne całkowicie „cyfrowe" (głównie w USA, Kanadzie i Europie Zachodniej), proces diagnozowania staje się coraz bardziej globalny (rozwój telemedycyny). Powstają struktury typu grid łączące całe kontynenty w celu tworzenia referencyjnych baz badań obrazowych, konstruowane są coraz doskonalsze narzędzia typu CAD (komputerowe wspomaganie diagnozy), systemy do telekonsultacji czy teleoperacji itd. [1,2].

Kompresja zmniejszająca wielokrotnie rozmiary plików z danymi obrazowymi pozwala znacząco zredukować wspomniane ograniczenia i usprawnić systemy typu PACS, RIS, HIS (szpitalne systemy informacyjne), telemedyczne. Stąd niemal wszystkie medyczne systemy gromadzenia i wymiany informacji obrazowej wykorzystują różne metody efektywnej kompresji, zarówno numerycznie bezstratnej (rekonstruujące oryginał ze skompresowanej reprezentacji z dokładnością do pojedynczego bitu), jak również numerycznie stratnej, tj. odtwarzające oryginalny obraz bez dokładności bitowej, ale zachowujący w założeniu jego pełną wartość diagnostyczną.

Możliwe do uzyskania za pomocą metod bezstratnych stopnie kompresji nie przekraczają zwykle wartości 3:1. Do zastosowań medycznych wykorzystywano od dawna techniki kompresji obrazów bazujące na predykcji DPCM, HINT, hierarchicznej strukturze piramidy różnicowej czy multiplikatywnych modelach autoregresji [3]. Najbardziej skuteczne obecnie metody bazują na intuicyjnych, dobieranych eksperymentalnie schematach predykcji czy sąsiedztwa w algorytmach kodowania, co pozwala uzyskać dużą efektywność jedynie dla wybranych klas obrazów. Duża złożoność modeli informacji obrazowej powoduje, że globalna optymalizacja procesu kompresji wymaga ogromnych nakładów obliczeniowych, co prowadzi do rozwiązań niepraktycznych (koder TMW [4]). Do najbardziej efektywnych i użytecznych rozwiązań zaliczyć należy CALIC, standard JPEG-LS oraz metody falkowe SPIHT i standard JPEG2000 [5]. Podejmowane były także próby stosowania koderów obrazów czarno-białych według standardów JBIG i JBIG2 do kompresji obrazów ze skalą szarości, w tym medycznych [6].

Stratne metody kompresji stają się koniecznością, ponieważ efektywność metod bezstratnych jest ograniczona i niewystarczająca w wielu zastosowaniach, np. telemedycznych. Spowodowało to dopuszczenie przez FDA (U.S. Food and Drug Administration) kompresji stratnej w archiwizacji medycznych danych obrazowych, choć w tej kwestii nie ma pełnego konsensusu środowisk medycznych, a wiele zależy od doświadczenia i wiedzy radiologa stosującego metody stratne (więcej zobacz w [7]). Do najbardziej efektywnych narzędzi należy zaliczyć kodery falkowe, wykorzystane przez twórców standardu JPEG2000 (JPEG2000 został zaaprobowany jako metoda stratnej kompresji przez standard cyfrowych metod medycznego obrazowania DICOM).

Warunkiem koniecznym stosowania metod nieodwracalnych jest zachowanie pełnej informacji diagnostycznej zawartej w oryginałach. Stwierdzenie wiarygodności diagnostycznej obrazu rekonstruowanego przy określonym stopniu kompresji wymaga rzetelnych metod porównania wartości diagnostycznej oryginału i rekonstrukcji. Klasycznym sposobem eksperymentalnej weryfikacji diagnostycznej wiarygodności obrazów są testy detekcji patologii wykorzystujące analizę statystyczną wyników decyzji radiologów z krzywą ROC (ang. receiver operating characteristic). Wobec znanych ograniczeń tej metody stosowane są jej modyfikacje oraz testy subiektywnej oceny jakości obrazów (porównawcze, bezwzględne, ze skalą diagnostyczną) [8] i miary obliczeniowe, zarówno skalarne jak i wektorowe wykorzystujące subiektywny dobór parametrów [9].

Badania mammograficzne ze względu na trudność jednoznacznej interpretacji diagnostycznej, przy dużej użyteczności w szeroko zakrojonych badaniach przesiewowych dotyczących raka piersi (pierwszej przyczyny zgonów nowotworowych kobiet w Polsce), są obok radiografii płuc głównym obszarem zastosowań nowoczesnych systemów cyfrowej radiologii, telemedycznych i wspomagających diagnozę [10,11,12]. W pracy tej zamieszczono rezultaty i dyskusję eksperymentów, w których próbowano określić granice efektywności kompresji obrazów mammografii rentgenowskiej. Wykorzystano najbardziej efektywne narzędzia kompresji bezstratnej i stratnej. Wiarygodność diagnostyczną stratnie kompresowanych mammogramów badano za pomocą dwóch metod: testów detekcji oraz subiektywnej oceny jakości diagnostycznej obrazów. Analiza uzyskanych wyników pozwoliła określić wartości możliwych i dopuszczalnych stopni kompresji, wskazać najlepsze metody kompresji odwracalnej oraz zasugerować „bezpieczny" stopień kompresji nieodwracalnej metodami falkowymi. W testach uczestniczyło 9 radiologów z 3 różnych ośrodków warszawskich. Przedmiotem eksperymentów było ponad 300 obrazów mammograficznych: reprezentatywne, trudne diagnostycznie badania własne (ponad 200) zarejestrowane w dwóch ośrodkach w ciągu 3 lat oraz ponad 100 wybranych z DDSM [13].

Materiał i Metody

Kompresja bezstratna

Aby ocenić możliwą do uzyskania efektywność kompresji wykorzystano znane z wysokiej wydajności kodery CALIC, SPIHT i APT, realizacje nowych standardów JPEG-LS, JPEG2000, JBIG i JBIG2, a także pozwalający niekiedy uzyskać obiecujące rezultaty binarny koder arytmetyczny (metody scharakteryzowano w [14]). Obok oszacowania maksymalnych stopni kompresji mammogramów, istotną była także selekcja najbardziej efektywnego kodera do celów archiwizacji i telediagnozy.

Wykorzystano 131 mammogramów testowych, na które złożyło się 19 obrazów z bazy własnej (w tym 13 użytych w testach kompresji stratnej), uzupełnionych mammogramami z bazy DDSM.

Kompresja stratna

Obrazy własne wykorzystane w badaniach zostały przekonwertowane do postaci cyfrowej za pomocą skanera Umax, rozdzielczość 80 μm, dynamika 14 bpp (bitów na piksel) [15]. Dokonano wstępnej analizy badań w postaci analogowej i cyfrowej w celu selekcji ponad 100 mammogramów do oceny efektów kompresji stratnej. Wykorzystano dwa falkowe kodery obrazów: JPEG2000 oraz MBWT [16].

W testach wstępnych z udziałem 2 radiologów-ekspertów (z różnych ośrodków) ustalono wartości stopni kompresji mammogramów. W teście detekcji wybrano wartości wokół przypuszczalnej granicy akceptowalności (tylko dwie w celu minimalizacji stopnia złożoności testu). Do oceny jakości diagnostycznej wybrano kolejno graniczne wartości: niezauważalnych zmian, zniekształceń wpływających bardzo nieznacznie na jakość rekonstrukcji wybranych struktur, widocznych odkształceń, znaczących deformacji zasadniczo wpływających na percepcję i interpretację poszczególnych zmian (przynajmniej w niektórych przypadkach). Dobrano szeroki przedział wartości badanych stopni kompresji, aby opinie radiologów pokryły cały zakres przyjętej liczbowej skali ocen. Wybrano także możliwie mały, ale reprezentatywny ze względu na treść diagnostyczną i obserwowane efekty kompresji zbiór mammogramów do obu testów (podobnie powodem było zmniejszenie złożoności testów przy zachowaniu ich wiarygodności). Dla testu detekcji ustalono „złoty standard" w konwencji standardu zgodnego i osobnego, z wykorzystaniem analogowych badań oryginalnych na kliszy oraz badań dodatkowych, a także diagnoz zweryfikowanych w wyniku przebiegu procesu leczenia. Jeden z ekspertów kontrolował przebieg testów czuwając nad realizacją przyjętych założeń (głównie niezależnością pracy obserwatorów i odpowiednimi warunkami pracy).

W zasadniczych testach oceny skutków kompresji wzięło udział 7 innych radiologów-obserwatorów (z 3 różnych ośrodków). Mammogramy oceniano w warunkach pracy klinicznej (miejsce, sprzęt, oświetlenie itd.) z zachowaniem niezależności decyzji poszczególnych obserwatorów (eliminacja skojarzeń, podpowiedzi, wzajemnego wpływu ocen, możliwości rozpoznania wersji obrazu etc.). Pracowano zwykle w dwóch trybach prezentacji (wybór obserwatora): pomniejszonej wersji całego obrazu oraz przy rozmiarach rzeczywistych, z możliwością obserwacji jedynie części obrazu.

W **testach detekcji** wykorzystano 13 mammogramów, w tym 5 bez zmian patologicznych. Patologie występujące w obrazach testowych to zmiany nowotworowe (potwierdzone histopatologicznie) w postaci guzków (dobrze odgraniczonych i spikularnych), guzkowych zagęszczeń o promienistym charakterze, zaburzeń architektury z mikrozwapnieniami oraz skupiska mikrozwapnień bez zmiany guzkowej (Ryc. 1). Wybrane wartości średnich bitowych kompresowanych obrazów to 0,1 bpp oraz 0,04 bpp (badano możliwość przesunięcia granicy dopuszczalnej wartości średniej bitowej poniżej 0,15 bpp sugerowanej w [17]). Daje to liczbę 39 obrazów testowych.

Obrazy prezentowane były pojedynczo, tylko w jednej projekcji. Decyzje dotyczyły stwierdzenia obecności lub braku patologii oraz jej ewentualnej lokalizacji. W dodatkowych uwagach określano charakter patologii. Przyjęto zasadę, że decyzja prawdziwie pozytywna oznacza poprawną detekcję patologii, tj. właściwą interpretację zmiany oraz jej lokalizację (zgodnie ze „złotym standardem"). Nie było ograniczeń co do liczby zmian patologicznych w pojedynczym obrazie. Można było więc odnotować nadrozpoznania, czyli wskazania zmian patologicznych w miejscach, gdzie według „złotego standardu" one nie występowały (zwiększając liczbę decyzji fałszywie pozytywnych). Obrazy były prezentowane w trzech sesjach: najpierw obrazy potencjalnie najgorszej jakości (0,04 bpp), potem rekonstruowane ze średniej 0,1 bpp, a na końcu obrazy oryginalne. Starano się rozdzielić te sesje możliwie długą przerwą (wynosiła ona minimum pół godziny).

Subiektywną ocenę jakości diagnostycznej przeprowadzono w celu bardziej wnikliwej oceny skutków kompresji mammogramów i wpływu wprowadzanych zniekształceń na jakość procesu interpretacji informacji diagnostycznej.

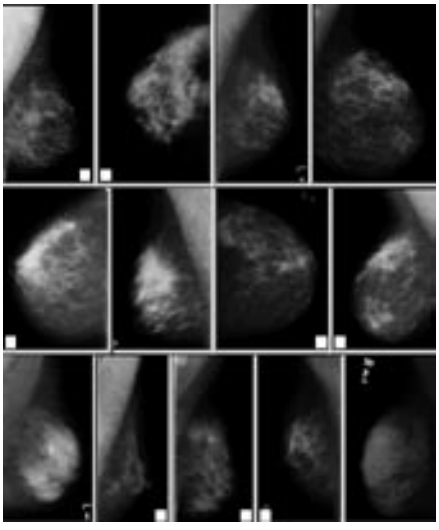


Figure 1. Mammograms used in the detection test. White squares denote images that contain a malignant lesion.

Rycina 1. Mammogramy wykorzystane w testach detekcji. Obrazy oznaczone białym kwadratem zawierają patologie.

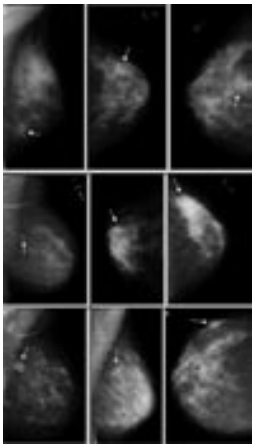


Figure 2. Mammograms used in the diagnostic quality estimation test. White arrows denote suspected regions. From left to right and top to bottom there are: 1) mass of high density with an ill-defined margin, 2) a spiculated mass; 3) a spiculated mass: original 6 in Tab. 3, which was particularly difficult to interpret, 4) a spiculated lesion, 5) a circumscribed mass of high density to differentiate between malignant and benign, 6,7) a circumscribed mass of high density, 8) a spiculated mass of high density, 9) a mass of high density with a well-defined margin.

Rycina 2. Mammogramy wykorzystane w testach oceny jakości diagnostycznej. Strzałki wskazują obszary szczególnie istotne w ocenie (sugerowane radiologom). Zawierają one, w porządku od lewej do prawej zaczynając od góry: 1) nieregularne zagęszczenie; 2) zagęszczenie o promienistym charakterze; 3, 4 i 8) guzek spikularny; 5) guzek dobrze odgraniczony o częściowo zatartych zarysach, 6, 7 i 9) guzek dobrze odgraniczony, dobrze wysycony.

Użyto 9 innych mammogramów zawierających reprezentatywne, niejednoznaczne w interpretacji przypadki patologii lub zmian łagodnych, zaburzeń regularnych struktur czy zatartych, mało widocznych szczegółów. Chodziło o przypadki „graniczne" w interpretacji, gdzie lekarz miał ocenić warunki postawienia poprawnej diagnozy. Przygotowane regiony zainteresowań ROI zawierały ważne diagnostycznie obszary, jak np. zbiegające się krawędzie i cienie mogące być interpretowane jako guzek spikularny, niskokontrastowe zmiany o nieregularnych, źle odgraniczonych zarysach w tzw. sutkach gęstych oraz obszary z podejrzeniem obecności mikrozwapnień (Ryc. 2).

Metoda oceny jakości diagnostycznej obrazów polegała na wyrażeniu, w przyjętej skali ocen, opinii dotyczącej jakości wyszczególnionych cech obrazu, które mają zasadniczy wpływ na proces detekcji i diagnozy obserwowanych zmian. Istotą oceny było śledzenie symptomów patologii, wszelkich zaburzeń normy i zmiany ich charakteru. Symptomy te to niewielkie zmiany dotyczące charakteru tekstur, zarysu krawędzi (kształt, gradient, ciągłość, relacja do wnętrza i zewnętrza struktur oraz sąsiednich krawędzi itp.), widoczności (ostrości) analizowanych szczegółów struktur. W kategoriach diagnostycznych mowa jest tutaj o poziomie wysycenia zmian, spikuli, gęstości guzków, ich kształcie, granicach, zarysie, rozmiarze oraz obecności lub braku mikrozwapnień. Uwydatnienia, odkształcenia i lokalne deformacje powodowane kompresją mogły generować dodatkowe symptomy patologii lub ukrywać rzeczywiste zmiany chorobowe.

Ocenę jakości (tj. zdolności percepcji, łatwości interpretacji) mammogramów dokonywano za pomocą następującego zestawu cech lokalnych: lokalny kontrast (względem poziomu gęstości tkanki), klarowność interpretacji (ostrość, widoczność, zdolność detekcji), kształt oraz zarysy (krawędzie, rozróżnialność konturów, relacja tekstur) wybranych struktur, w tym patologii oraz zmian łagodnych. Zastosowano skalę od 1 (słabe, niewyraźne, ledwo dostrzegalne, zniekształcone) do 3 (wyraźne, dobrze rozróżnialne, regularne, nie budzące wątpliwości), a średnia suma ocen jakości 4 cech danego obrazu (od 7 obserwatorów) była wyznacznikiem jego jakości.

Obrazy oceniano w grupach po 5 wyświetlanych razem obrazów: oryginał plus 4 rekonstrukcje po kompresji do 1 bpp, 0,6 bpp, 0,1 bpp oraz 0,04 bpp. Test przeprowadzono w kilku sesjach jednogodzinnych (w zależności od liczby obrazów testowych), bez ograniczeń czasu oceny danej grupy, z możliwością doboru optymalnych warunków prezentacji (powiększanie, korekcja jasności i kontrastu). W sumie oceniono jakość 75 obrazów (15 grup testowych). 9 obrazów było kompresowanych koderem JPEG2000, a 6 z nich dodatkowo kodowano metodą MBWT (aby zaobserwować różnice w efektach kompresji różnymi koderami nie zwiększając zbytnio stopnia złożoności testu).

Table 1. Evaluation of lossless compression efficiency: mean bit rate values of 131 compressed mammograms with sizes of 1000 x 1000 to 4500 x 4500 pixels, 12–14 bpp. Bolded results signify the lowest bit rates. JPEG2000 opt: optimized JPEG2000 coder, BKA: binary arithmetic coder.

Tabela 1. Ocena efektywności kompresji bezstratnej: wartości średniej bitowej dla zbioru 131 testowych mammogramów o rozmiarach od 1000x1000 do 4500x4500 pikseli i dynamice wartości pikseli 12–14 bitów. Pogrubiono najmniejsze wartości średniej bitowej. „JPEG2000 opt" oznacza optymalizowaną wersję tego kodera. BKA to binarny koder arytmetyczny.

Dla 131 obrazów testowych	Kodery								
	CALIC	**JPEG-LS**	**APT**	**SPIHT**	**JPEG2000**	**JPEG2000 opt**	**JBIG**	**JB2**	**BKA**
wartość średnia	**6,64**	6,68	7,14	6,69	7,86	7,04	7,09	7,31	7,05
minimum	**1,49**	1,50	1,60	1,60	1,56	1,58	1,65	1,66	1,70
maksimum	**9,28**	9,40	10,01	9,37	10,68	9,65	9,85	10,39	9,61

Table 2. Results of pathology detection: statistical analysis of data presented in Fig. 3. Bolded results are those that are significant at a level of $p=0.05$ (suggesting that the null hypothesis of the respective test is false).

Tabela 2. Wyniki testu detekcji: analiza statystyczna krzywych z Ryc. 3. Pogrubiono wyniki sugerujące odrzucenie odpowiednich hipotez zerowych (przy poziomie istotności $p=0,05$).

Wartość	Czułość			Nietrafność			PVP		
	Oryginał	0,1 bpp	0,04 bpp	Oryginał	0,1 bpp	0,04 bpp	Oryginał	0,1 bpp	0,04 bpp
Średnia	0,8095	0,7302	0,6190	0,3956	0,5055	0,4286	0,6060	0,5163	0,5015
Wariancja	0,0112	0,0323	0,0241	0,0324	0,0510	0,0175	0,0221	0,0175	0,0056
Statystyka *t*	→	0,9321	**2,4842**	→	-0,9321	-0,3616	→	1,1035	1,5382

Wyniki eksperymentów

Kompresja bezstratna

Rezultaty zamieszczono w tabeli 1. Zdecydowana większość obrazów (tj. 118) kompresowana była w stopniu poniżej 2:1 (z minimum równym 1,5:1), a jedynie dla 6 obrazów uzyskano stopień większy od 5:1 (z maksimum równym 8:1). Średnio metodami bezstratnymi udaje się redukować rozmiary cyfrowej reprezentacji mammogramów w stopniu 2:1 stosując najbardziej efektywną metodę CALIC. Niewiele mniejszą wydajność kompresji uzyskano dla standardu JPEG-LS i kodera falkowego SPIHT, przy czym dla JPEG-LS uzyskano większe zróżnicowanie wartości średniej bitowej, co może świadczyć o mniejszej uniwersalności tej metody. Optymalizacja kodera JPEG2000 poprzez dobór bardziej efektywnych postaci dekompozycji falkowej (jak w [16]) pozwoliła zmniejszyć o 10% średnią bitową w stosunku do podstawowego banku filtrów 5/3. Nie pozwoliło to jednak osiągnąć wydajności zbliżonej do CALIC. Metody kodujące kolejne mapy bitowe obrazów ze skalą szarości (JBIG, JB2, binarny koder arytmetyczny) miały o blisko 8% mniejszą średnią efektywność niż CALIC.

Kompresja stratna

Na Ryc. 3 przedstawiono krzywe wyznaczone na podstawie decyzji diagnostycznych radiologów z testu detekcji (obliczano liczbę decyzji prawdziwie pozytywnych oraz fałszywie pozytywnych radiologów). Przy obserwacji oryginałów zanotowano więcej poprawnych detekcji patologii niż w przypadku diagnoz z obrazów rekonstruowanych (0,1 bpp i 0,04 bpp), przy zbliżonym poziomie pomyłek (nietrafności decyzji). Widać to na wykresie z czułością ocen poszczególnych radiologów, przy czym zdecydowanie najmniejszą czułość miały interpretacje obrazów zakodowanych do 0,04 bpp. Wartości PVP (przewidywana wartość pozytywna) są już mniej zróżnicowane. Brakuje wyraźnych tendencji pozwalających jednoznacznie stwierdzić pogorszenie wiarygodności diagnostycznej obrazów rekonstruowanych. Wnioski z wykresów potwierdzają wyniki testów statystycznych z Tab. 2.

Porównując wiarygodność oryginałów i obrazów kompresowanych ze średnią 0,04 bpp należy odrzucić hipotezę zerową o równej średniej czułości detekcji dla obu populacji (przyjęto poziom istotności $p=0,05$). Potwierdzają to komentarze obserwatorów zarejestrowane podczas testu. Ponadto wartości statystyki *t* dla rekonstrukcji z 0,1 bpp wskazują, że nie ma przesłanek do odrzucenia hipotezy w tym przypadku. Zanotowano pojedyncze przypadki, kiedy wskutek kompresji zdolność detekcji patologii uległa poprawie. Dla reprezentacji 0,1 bpp jednego z mammogramów zanotowano więcej decyzji prawdziwie pozytywnych i mniej nadrozpoznań. W innym przypadku systematycznie malała liczba nadrozpoznań przy zmniejszaniu średniej bitowej skompresowanej reprezentacji.

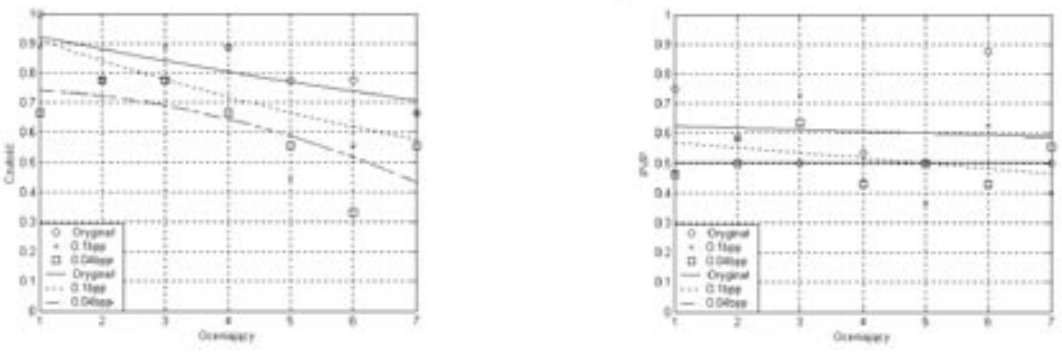


Figure 3. Results of the pathology detection test: plots of sensitivity and PVP as a function of radiologist rate. Data points were approximated by polynomials.

Rycina 3. Wyniki testu detekcji: wykresy czułości i PVP w funkcji ocen zespołu radiologów. Punkty danych aproksymowano wielomianami stopnia 2.

Table 3. Results of the subjective rating of diagnostic quality: mean rates of test mammograms (original and four compressed versions). Seven radiologists scored, on a scale of 1 to 3, four diagnostically important features of structures in mammograms. Mean values of the sum of four scores are presented in this table. Bolded results are those that are significant at a level of $p=0.05$.

Tabela 3. Wyniki testu oceny jakości diagnostycznej mammogramów: wartości średnie ocen jakości oryginału i 4 jego rekonstrukcji. Cztery diagnostycznie istotne cechy obrazów zostały ocenione przez 7 radiologów w skali od 1 do 3 (zamieszczono średnią sumy tych ocen). Pogrubiono wyniki sugerujące odrzucenie odpowiednich hipotez zerowych (przy poziomie istotności $p=0,05$).

Wersje	Obrazy testowe									średnia z M1-M9	Odchyl. st. M1-M9	Statystyka *t*
	M1	M2	M3	M4	M5	M6	M7	M8	M9			
Oryginał	9,71	10,71	9,07	9,86	10,00	9,36	9,57	10,14	10,29	9,83	1,96	↓
1 bpp	9,36	10,43	9,29	10,36	10,43	9,57	8,43	10,14	9,86	9,82	1,97	0,0367
0,6 bpp	8,43	8,43	7,07	9,64	10,14	6,64	8,86	10,71	9,29	9,02	2,24	**2,7753**
0,1 bpp	9,21	6,64	7,07	8,07	7,57	9,21	8,00	9,43	9,43	7,82	2,41	**6,5987**
0,04 bpp	5,79	4,21	4,50	5,00	5,07	4,29	4,71	5,71	5,71	4,93	1,50	**20,246**

Przeprowadzono także szereg dodatkowych analiz statystycznych rezultatów decyzji radiologów: rozszerzono skalę ocen do 4 poziomów uwzględniając oddzielnie nadrozpoznania, zastosowano bardziej złożone metody weryfikacji hipotez statystycznych (jedno i dwuwymiarowe). Opisano to w [18]. Wnioski z parametrycznych testów istotności bazujących na innych statystykach są podobne jak przy *t*.

Według średniej opinii radiologów z testu subiektywnej oceny jakości diagnostycznej ponad połowa obrazów testowych miała lepszą jakość po kompresji (Tab. 3). Tylko 4 oryginały uzyskały najwyższe oceny, natomiast w pozostałych przypadkach najlepiej oceniono 4 mammogramy rekonstruowane ze średniej 1 bpp oraz jeden z 0,6 bpp. W ocenie statystycznej jakość oryginałów i rekonstrukcji z 1 bpp jest jednakowa (niemal nie różnią się wartości średniej i odchylenia standardowego zebranych ocen, a wartości statystyki *t* nie wykazują istotnej różnicy). Potwierdza to Ryc. 4.

Nieco mniejsza średnia wartość ocen i większe odchylenie standardowe dają zbyt duże wartości statystyki dla mammogramów reprezentacji 0,6 bpp oraz 0,1 bpp. Różnica jakości diagnostycznej jest statystycznie zauważalna. Jednak w kilku przypadkach jakość diagnostyczna kompresowanych mammogramów w stosunku do oryginałów została zachowana (połowa przypadków dla 0,6 bpp i 3 dla 0,1 bpp).

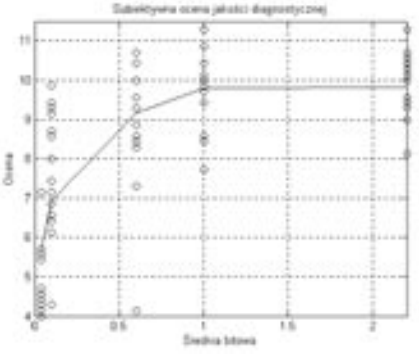


Figure 4. Mean subjective rating of diagnostic quality of the original and compressed mammograms.

Rycina 4. Uśrednione oceny jakości diagnostycznej oryginalnych i kompresowanych stratnie mammogramów.

Obrazy rekonstruowane ze średniej 0,04 bpp otrzymały najniższą ocenę w każdym przypadku, przy czym była ona zwykle prawie dwukrotnie niższa od ocen pozostałych wersji danego mammogramu. Jest to dowód wyraźnie gorszej jakości tych obrazów i degradacji cech obrazu istotnych diagnostycznie.

Dyskusja

W decyzjach radiologów patologia oznaczała *de facto* jedynie podejrzenie patologii (decyzja na podstawie tylko 1 projekcji), co dało większą czułość procedury detekcji. Większe zróżnicowanie ocen uzyskano także poprzez uwzględnienie nadrozpoznań, większą liczbę patologii w obrazie, ocenę poprawności lokalizacji zmian. Eksperci dysponujący pełną bazą badań nie tylko mammograficznych w kilku przypadkach przyznali, że „właściwie nie sposób na podstawie jednego zdjęcia stwierdzić patologii, można jednak zauważyć pewne hipotetyczne zmiany". Dość duża liczba nieprawidłowych kwalifikacji w decyzjach radiologów z testów detekcji wzięła się ze świadomego zwiększenia stopnia trudności testów w stosunku do praktyki klinicznej. Pozwoliło to przeprowadzić bardziej wiarygodne badania przy mniejszej złożoności testów, a także uwzględnić ewentualne „pozytywne" zmiany w obrazach rekonstruowanych po kompresji stratnej.

Przetworzenie danych powodowało niekiedy podkreślenie lekko zarysowanych patologii (ułatwiając ich interpretacje) lub też redukcję zniekształceń, które utrudniały wcześniej interpretację obrazu oryginalnego. Wydobycie pewnych cech poprawiających widoczność zmian zwiększa prawdopodobieństwo poprawnych decyzji lekarzy (pojedyncze przypadki w teście detekcji), jak również poprawia warunki diagnozy (test oceny jakości diagnostycznej).

Zbieżność ocen obrazów oraz opinii radiologów wyrażonych w testach i na temat testów pozwala na wysunięcie

wstępnych wniosków, że stosowanie falkowych metod kompresji w pewnych granicach wartości średnich bitowych nie zmniejsza wartości diagnostycznej obrazów. Porównanie oryginalnych obrazów cyfrowych z ich rekonstrukcjami nie wykazuje wtedy różnic istotnych diagnostycznie. Zebrane opinie radiologów sugerują bezpieczną wartość średniej bitowej dla szerokiego spektrum wykorzystanych badań mammograficznych na poziomie 1 bpp z możliwością dodatkowej redukcji rozmiaru danych skompresowanych w niektórych przypadkach do wartości 0,6 bpp, a nawet do 0,1 bpp. Jednocześnie przeprowadzenie dodatkowego testu z oceną oryginalnej wersji analogowej użytych mammogramów (na kliszy) mogłoby uczynić wnioski końcowe jeszcze bardziej wiarygodnymi.

Wnioski

Wnioski wynikające z 2 testów oceny efektów kompresji stratnej są różne. Zgodnie z analizą rezultatów testu detekcji patologii nie ma statystycznie znaczącej różnicy pomiędzy wiarygodnością oryginałów i rekonstrukcji 0,1 bpp, w przeciwieństwie do mammogramów po kompresji do 0,04 bpp. Stąd sugerowany dopuszczalny poziom redukcji średniej bitowej mammogramów w kompresji falkowej wynosi 0,1 bpp.

Konkluzje z oceny jakości diagnostycznej są bardziej restrykcyjne. Obserwacja jakości rekonstrukcji symptomów patologii wykazała konieczność odrzucenia kompresji mammogramów do wartości średniej 0,1 bpp, jak również 0,6 bpp. Nie było natomiast przesłanek pozwalających stwierdzić niewielką choćby utratę jakości przy kompresji w stopniu 14:1 (odpowiada mu średnia 1 bpp). Ustalenie takiego właśnie dopuszczalnego poziomu kompresji mammogramów testowych wydaje się bezpieczne, szczególnie wobec wątpliwości części środowisk medycznych, co do stosowania stratnej kompresji badań obrazowych.

Bezpieczna kompresja stratna (nieodwracalna) mammogramów daje wyraźnie większy stopień kompresji w stosunku do metod bezstratnych (jak 14:1 do 2:1) przy akceptowalnej złożoności obliczeniowej algorytmów kompresji. Pozwala to znacznie zredukować koszty nośników danych (problem archiwizacji) oraz usprawnić narzędzia wykorzystywane w telediagnostyce, gdzie kodery falkowe są szczególnie przydatne.

Piśmiennictwo:

1. Erickson BJ, Bartholmai B: Computer-aided detection and diagnosis at the start of the third millennium. J. Digital Imaging, 2002; 15: 5–14.
2. The Information Societies Technology project: MammoGrid – a European federated mammogram database implemented on a GRID infrastructure. EU Contract IST-2001–37614.
3. Wong S, Zaremba L, Gooden D, Huang HK: Radiologic image compression-a review. Proc. of the IEEE, 1995; 83:194–219.
4. Meyer B, Tischer P: TMW – a New Method for Lossless Image Compression. Proc PCS97 – Picture Coding Symposium, VDE-Verlag GMBH, Berlin, 1997; 533–538.
5. Przelaskowski A: Lossless encoding of medical images: hybrid modification of statistical modelling-based conception. J Electronic Imaging, 2002; 10: 966–976.
6. Ekstrand N: Lossless compression of grayscale images via context tree weighting. Proc IEEE Data Compression Conf, 1996; 132–139.
7. Przelaskowski A: Irreversible medical image compression: conditions of acceptability. Task Quarterly, 2004; 8: 303–316.
8. Cosman PC, Gray RM, Olshen RA: Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy. Proc IEEE, 1994; 82: 919–932.
9. Przelaskowski A: Vector quality measure of lossy compressed medical images. Computers in Biology and Medicine, 2004; 34: 193–207.
10. Viborny CJ, Giger ML, Nishikawa ML: Computer aided detection and diagnosis of breast cancer. Radiol Clin N Am, 2000; 38: 725–740.
11. Thursjell EL, Lernevall KA, Taube AA: Benefit of independent double reading in a population based mammography screening program. Radiology, 1994; 191:241.
12. Elmore JG, Miglioretti DL, Reisch LM et al: Screening mammograms by community radiologists: variability in false-positive rates. J Natl Cancer Inst, 2002; 94: 1373–1380.
13. *http://marathon.csee.usf.edu/Mammography/Database.html*
14. Przelaskowski A: Reversible compression of medical images: ways of performance improvement. SCAR Scientific Abstract Book, Vancouver, 2004; 44–46.
15. Kawalec T: System do wspomagania diagnostyki raka sutka. Praca dyplomowa magisterska pod kierunkiem A. Przelaskowskiego, Instytut Radioelektroniki PW, 2001.
16. Przelaskowski A: Falkowe metody kompresji danych obrazowych. Oficyna Wydawnicza PW, Warszawa 2002.
17. Perlmutter SM, Cosman PC, Gray RM et al: Image quality in lossy compressed digital mammograms. Signal Proces, 1997; 59: 189–210.
18. Przelaskowski A, Surowski P: Metody optymalizacji reprezentacji medycznych danych obrazowych do archiwizacji i transmisji telemedycznej. Sprawozdanie z grantu KBN 7 T11E 039 20, Warszawa, luty 2002.