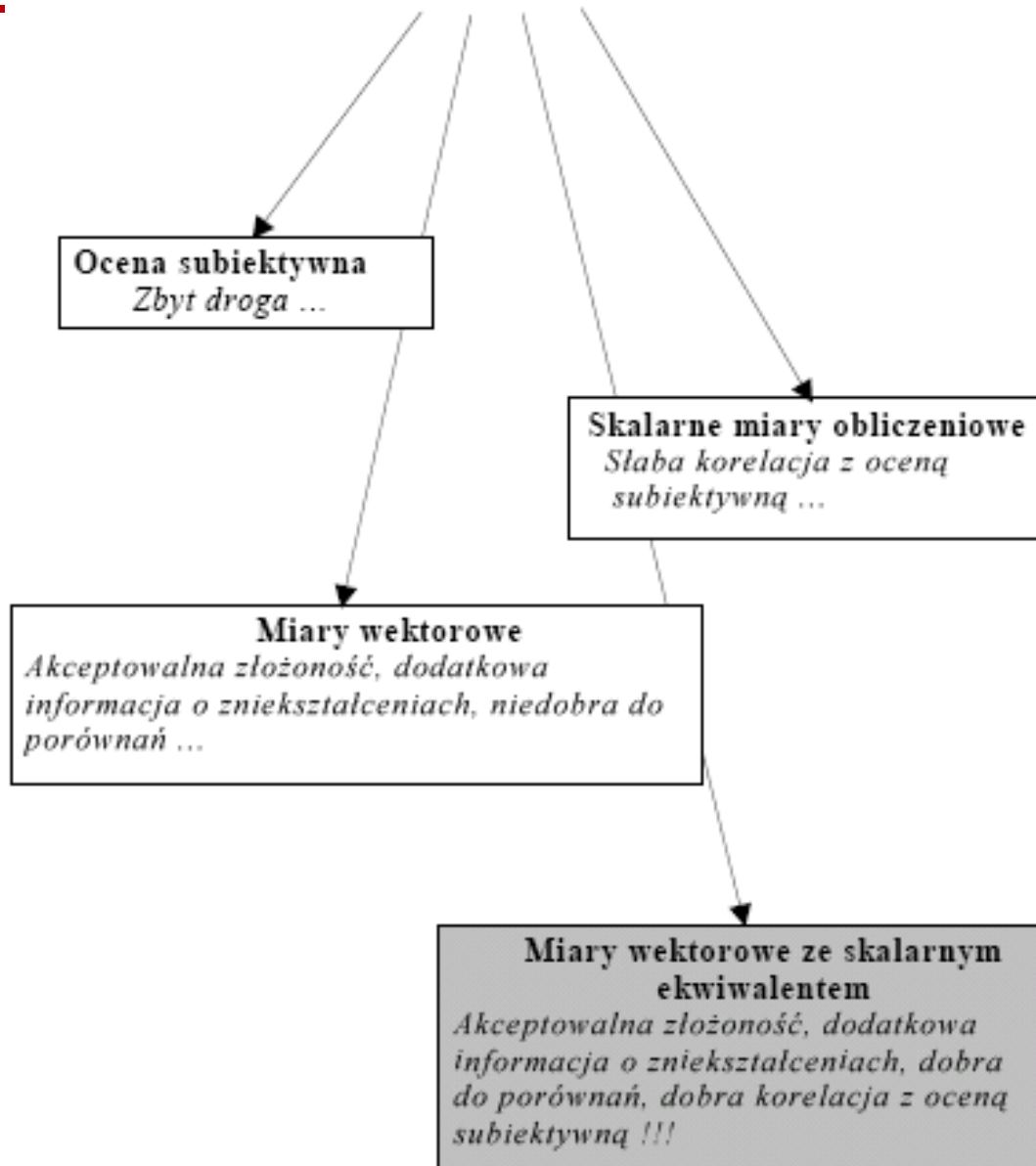


ANALIZA DANYCH OBRAZOWYCH

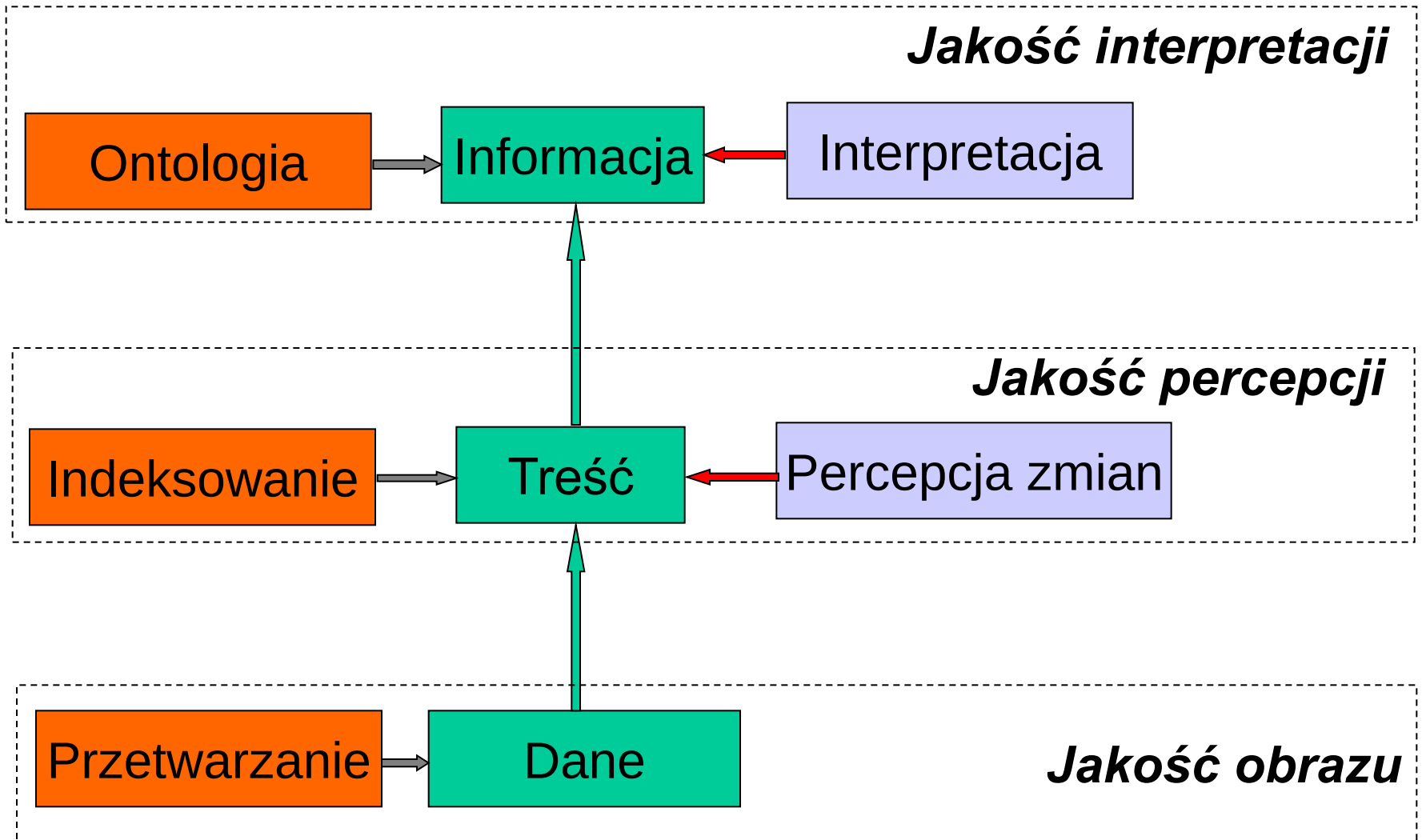
JAKOŚĆ OBRAZÓW – MIARY

Materiały przygotowane przez A.Przelaskowskiego

Jak mierzyć jakość?



Rola oceny jakości



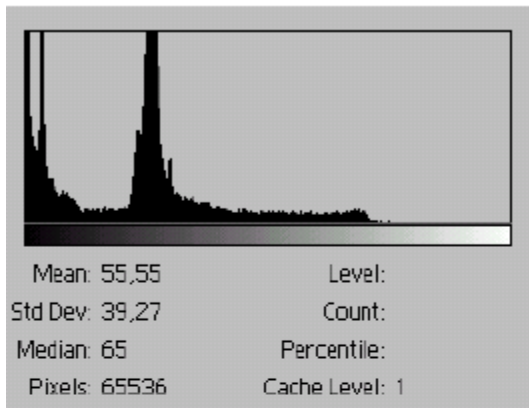
Ocena jakości obrazów - kontrast

$$K = \frac{f_{\max} - f_{\min}}{f_{\max} + f_{\min}}$$

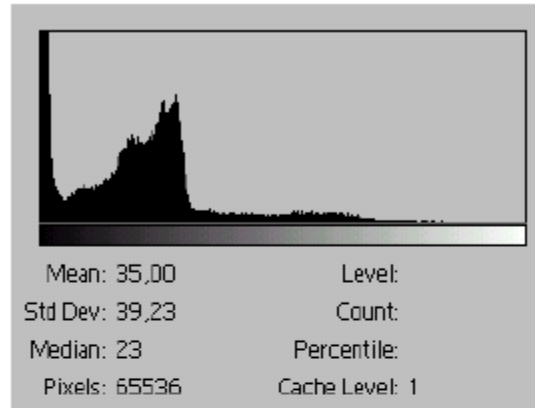
Dynamika medycznych danych obrazowych:

NM	USG	MRI
6 - 12 bitów	6 - 8 bitów	8 - 12 bitów
	CR	CT
	10 - 14 bitów	12 - 16 bitów

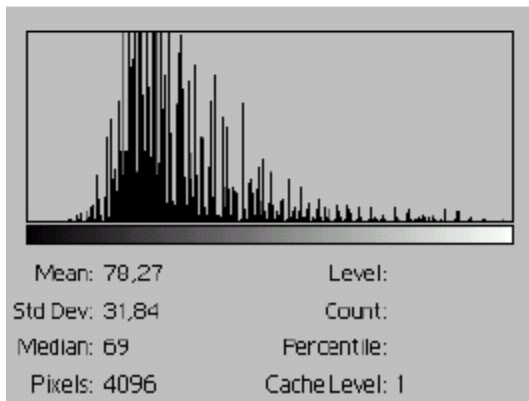
CT



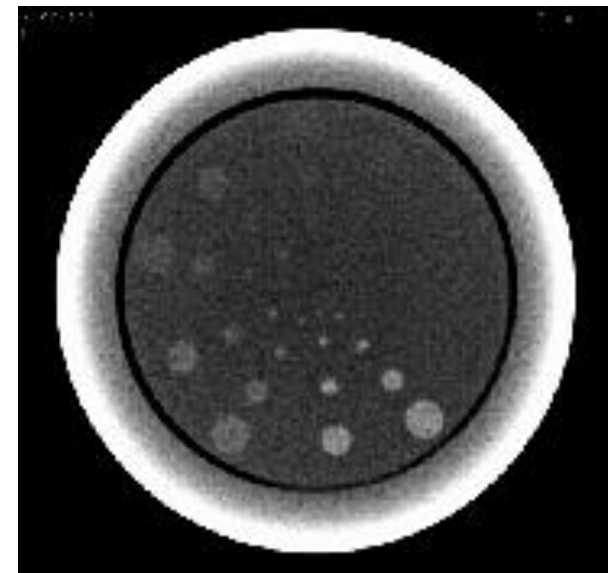
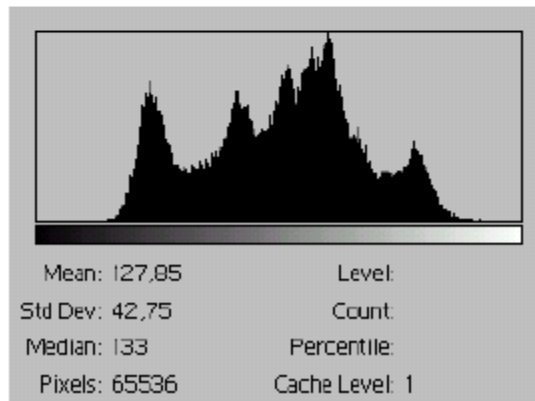
MRI



USG



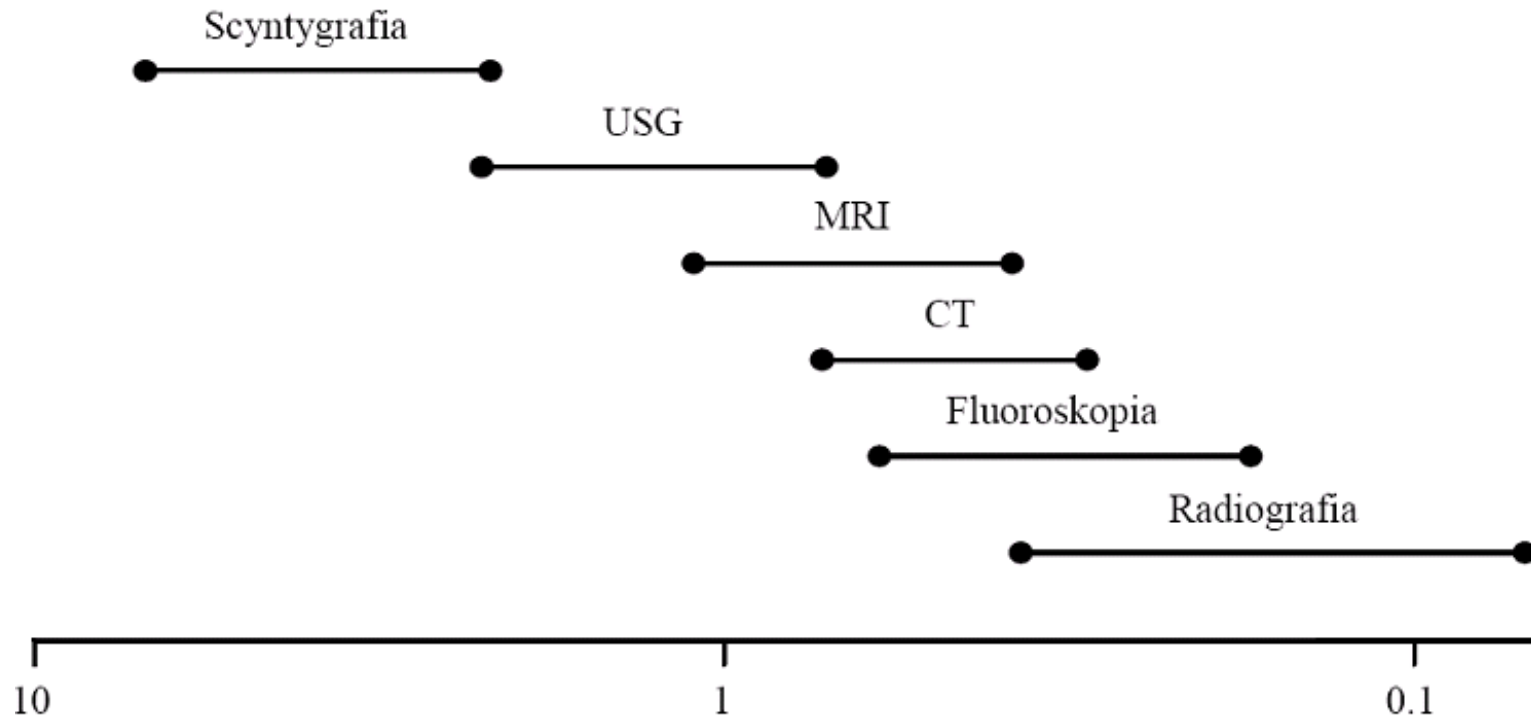
Lena



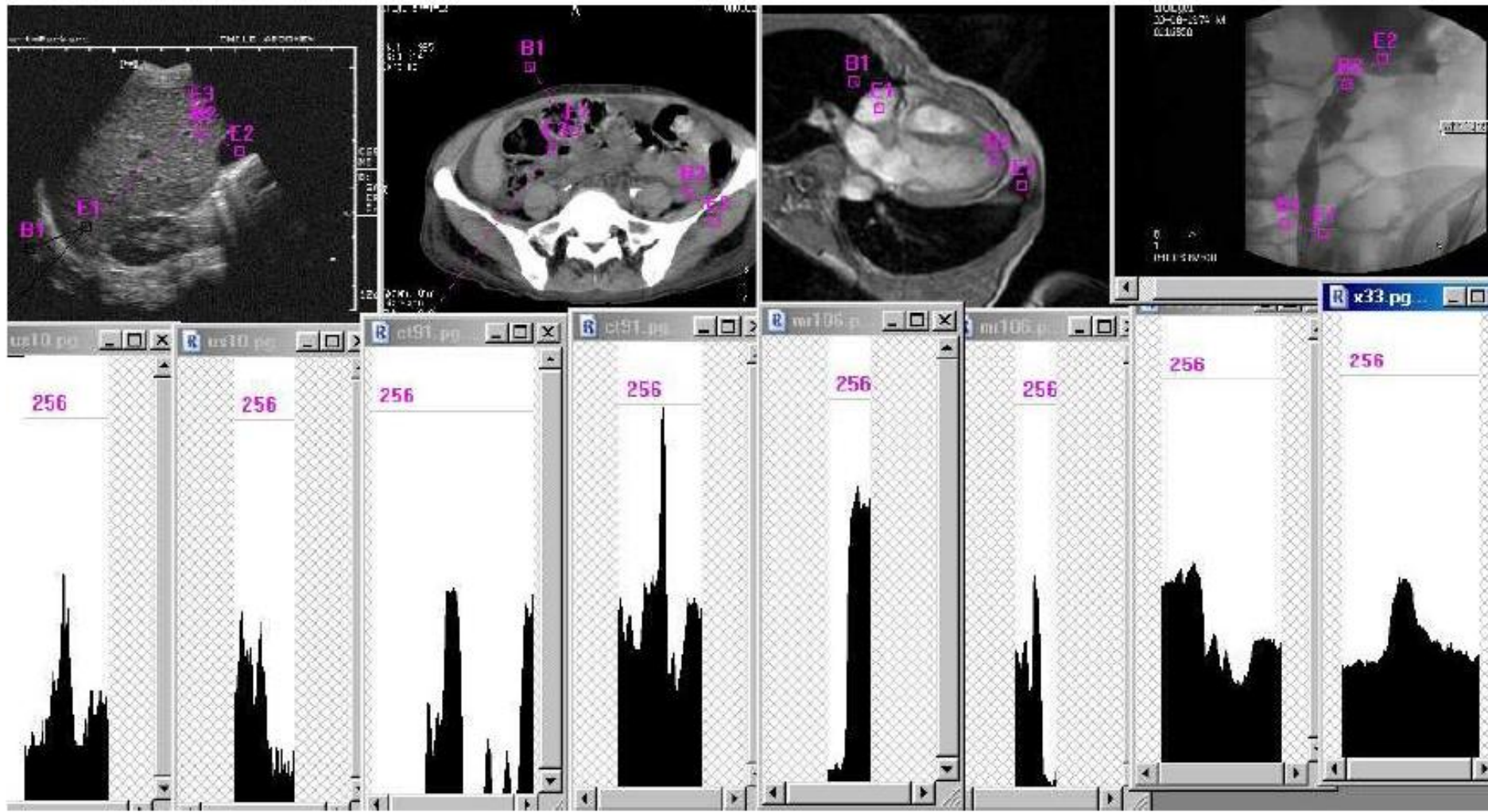
histogramy

Ocena jakości - rozdzielczość

minimalna odległość dwóch rozróżnialnych punktów



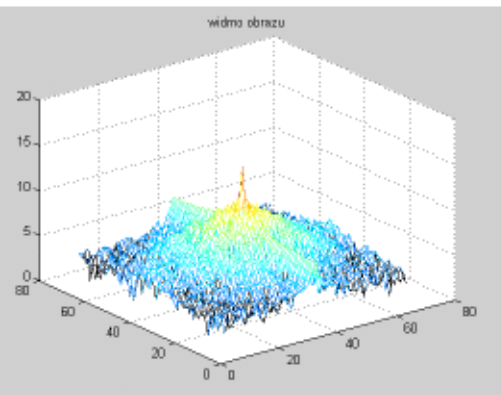
Ocena jakości - rozdzielczość



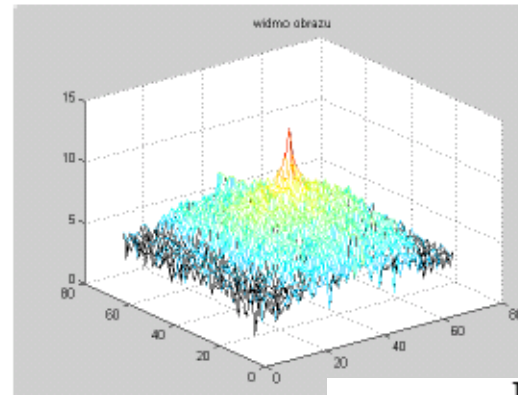
Ocena jakości - szumy

Modalność	CT	MRI	USG	Lena
Sig/szum	1.2	1.12	1.02	2.33
Cor_x	0.9734	0.8090	0.9490	0.8713
Cor_y	0.9828	0.9068	0.9043	0.9547
Autoreg_x	0.9932	0.9812	0.9912	0.9915
Autoreg_y	0.9932	0.9868	0.9788	0.9953

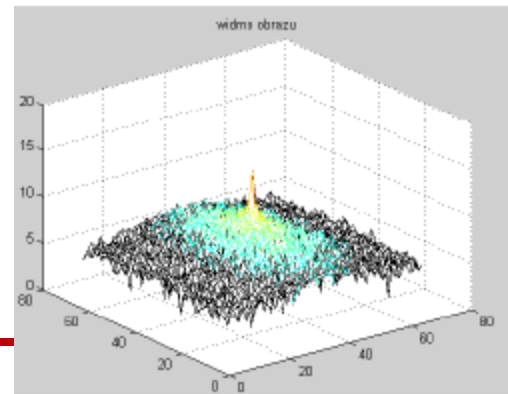
CT



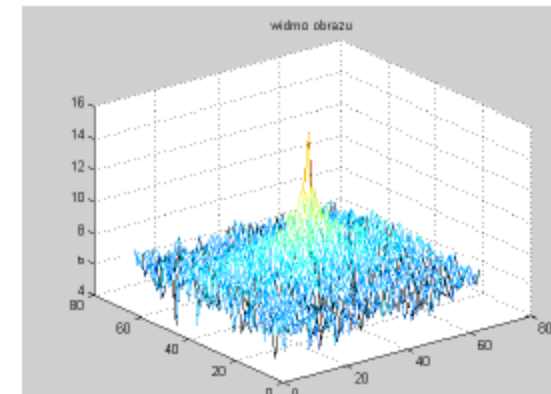
MRI



USG

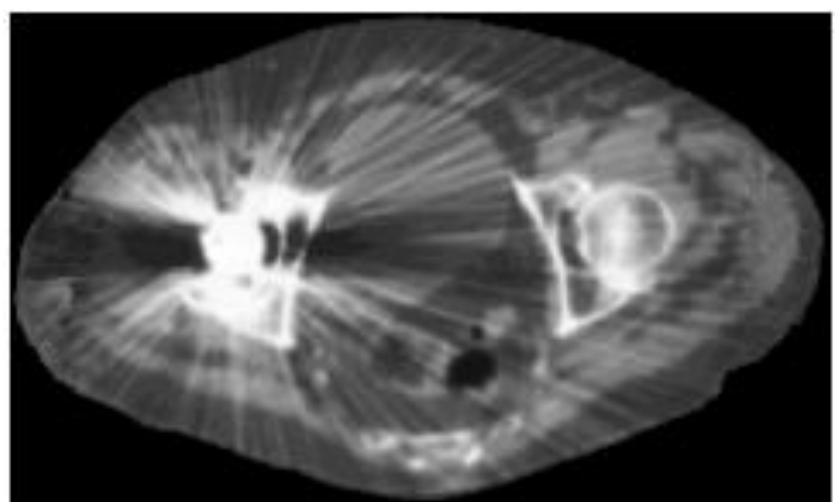
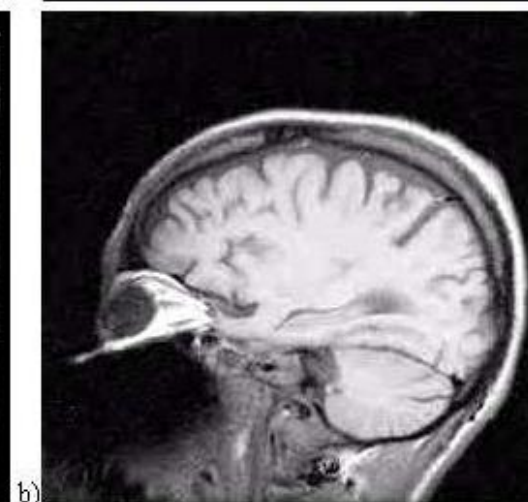
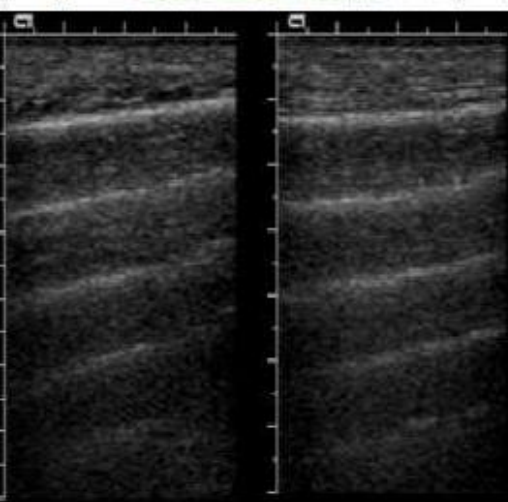
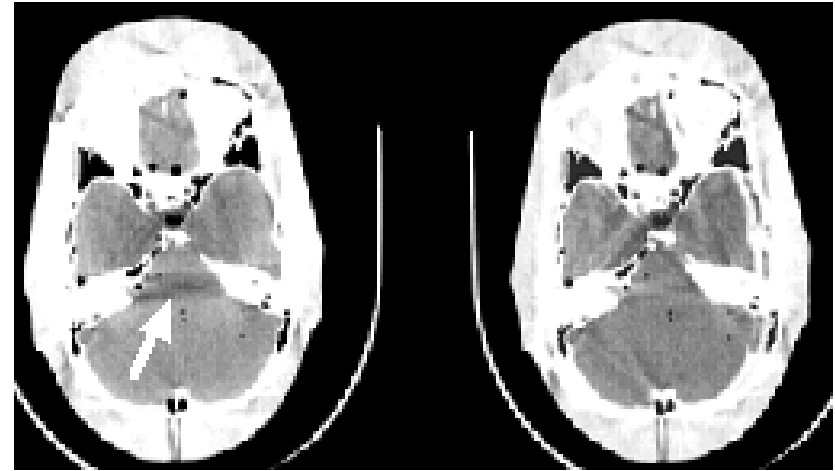
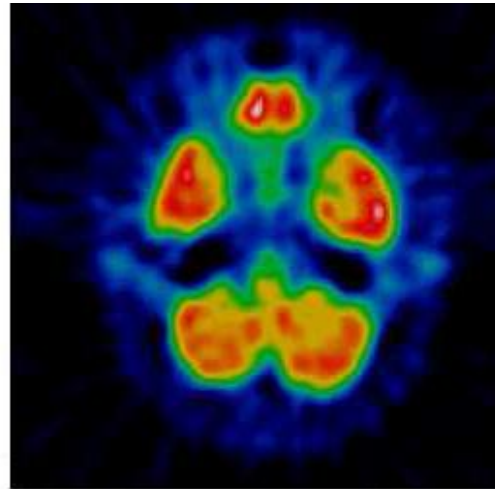
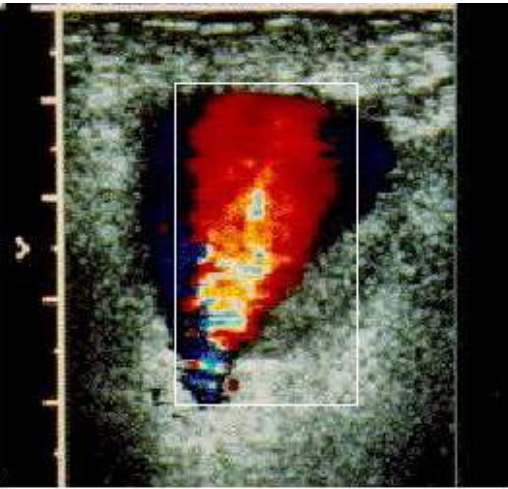


Lena



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

Ocena jakości – artefakty, zniekształcenia geometryczne



Obliczeniowe miary porównawcze

- maksymalna różnica (ang. *Maximum Difference*), zwana też szczytowym błędem bezwzględnym PAE (ang. *Peak Absolute Error*):

$$MD = \max_{m,n} \{ |f(m,n) - \tilde{f}(m,n)| \};$$

$$f(m,n), \tilde{f}(m,n)$$

porównywane obrazy

$$0 \leq m < M, 0 \leq n < N$$

- błąd średniokwadratowy (ang. *Mean Square Error*):

$$MSE = \frac{1}{MN} \sum_{m,n} [f(m,n) - \tilde{f}(m,n)]^2;$$

- szczytowy stosunek sygnału do szumu (ang. *Peak Signal to Noise Ratio*):

$$PSNR = 10 \log \frac{MN \cdot [\max_{m,n} \{f(m,n)\}]^2}{\sum_{m,n} [f(m,n) - \tilde{f}(m,n)]^2},$$

przy czym wartość $\max_{m,n} \{f(m,n)\}$ jest zwykle ustalana na poziomie największej możliwej (a nie faktycznej) wartości funkcji jasności, np. 255 dla danych 8-bitowych;

- średnia różnica (ang. *Average Difference*):

$$AD = \frac{1}{MN} \sum_{m,n} |f(m,n) - \tilde{f}(m,n)|;$$

Obliczeniowe miary porównawcze

- jakość korelacyjna (ang. *Correlation Quality*):

$$CQ = \frac{\sum_{m,n} f(m,n) \tilde{f}(m,n)}{\sum_{m,n} f(m,n)};$$

- wierność obrazu (ang. *Image Fidelity*):

$$IF = 1 - \frac{\sum_{m,n} [f(m,n) - \tilde{f}(m,n)]^2}{\sum_{m,n} [f(m,n)]^2};$$

- miara chi-kwadrat (ang. *chi-square measure*)

$$\chi^2 = \frac{1}{MN} \sum_{m,n} \frac{[f(m,n) - \tilde{f}(m,n)]^2}{f(m,n)}.$$

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

$k(u,v)$ i $\hat{k}(u,v)$ - współczynniki DCT przed i po obróbce

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

Ocena jakości

- Miara poprawy kontrastu: $C_{||} = C_{\text{przetw}}/C_{\text{oryg}}$, gdzie $C = (o - t)/(o + t)$; o – średnia jasność obiektu, t – średnia jasność tła
- Stosunek kontrastu do szumu (speklowego):

$$CSR = \frac{|o - t|}{(\sigma_o^2 - \sigma_t^2)^2}$$

- Indeks szumu:

$$SI = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \frac{\sigma(m, n)}{\mu(m, n)}$$

gdzie lokalnie

$$\sigma(m, n) = \max_{-1 \leq a, b \leq 1} (f(m + a, n + b) - \min_{-1 \leq a, b \leq 1} (f(m + a, n + b)))$$

$$\mu(m, n) = \frac{1}{9} \sum_{a, b=-1}^1 f(m + a, n + b)$$

- MTF i inne
-

Wykresy Hosaki (algorytm), miara porównawcza – segmentacja drzewa czwórkowego

1. Kryterium jednorodności (segmentacja drzewa czwórkowego), tworzenie kilku klas bloków o różnych rozmiarach, najczęściej kwadratowych o boku 1, 2, 4, 8, 16:

Blok B jest jednorodny, jeśli $\sigma_B^2 \leq T$, gdzie

$$\sigma_B^2 = \frac{1}{MN} \sum_{(i,j) \in B} (f(i,j) - \mu_B)^2, \text{ a } T \text{ jest założoną wartością progu (często}$$

$T=100$), bloki B : $2^i \times 2^i, i=0, \dots, n$, przy czym $n=4$ (najczęściej).

Wykresy Hosaki (algorytm) – średnia bloków poszczególnych klas

2. 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 $2^i \times 2^i$.

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

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

$$d\mu_i = \mu_i - \mu$$

Wykresy Hosaki (algorytm) – odchylenie standardowe bloków poszczególnych klas

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

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

Wykresy Hosaki (algorytm) – wektory cech

4. 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:

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

$(d\mu'_0, d\mu'_1, \dots, d\mu'_n), (\sigma'_1, \dots, \sigma'_n)$ - dla obrazu rekonstruowanego

Na ich podstawie tworzymy dwa wektory różnicowe jako:

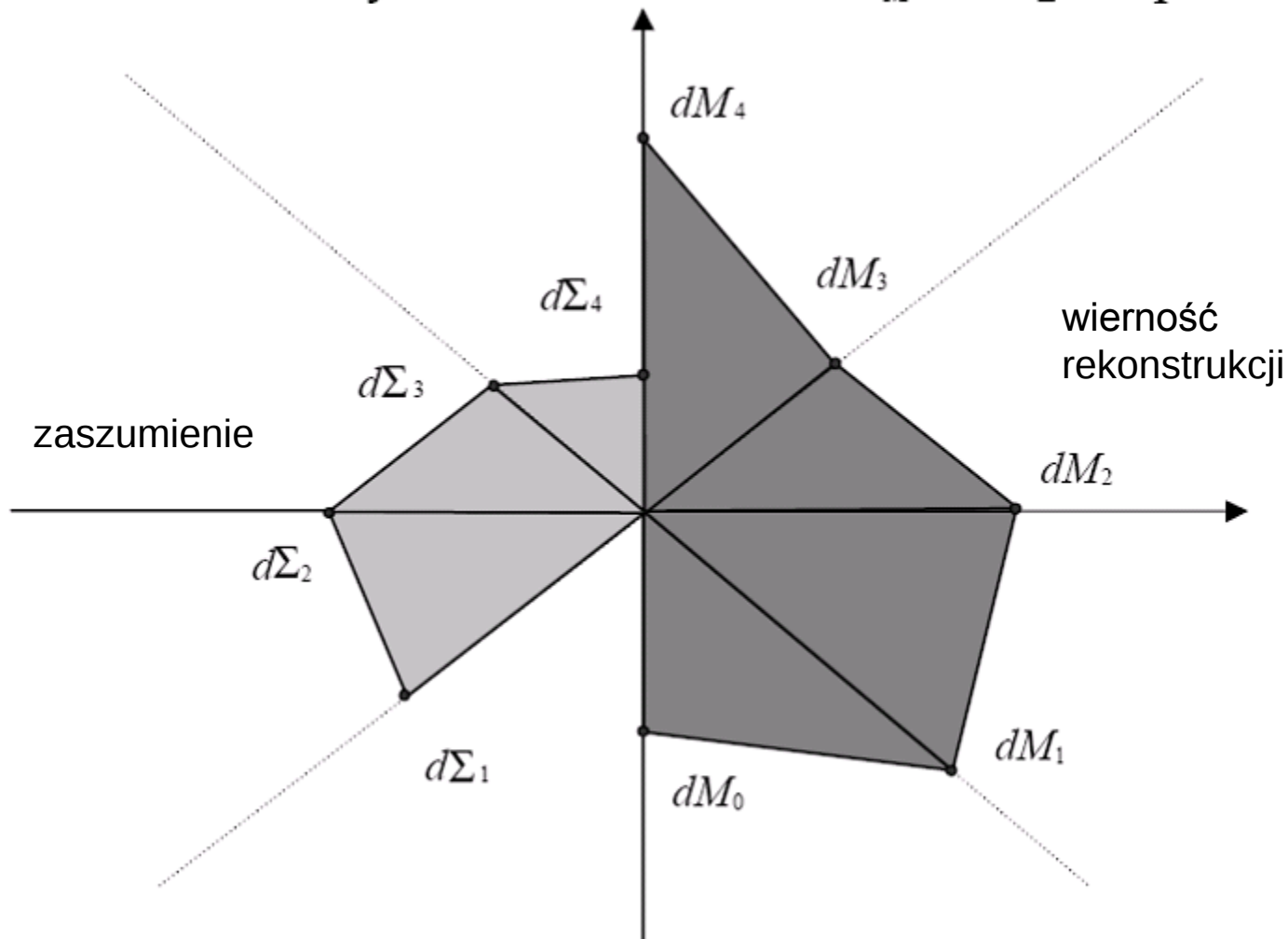
$$\vec{d}_M = (dM_0, dM_1, \dots, dM_n), \quad \vec{d}_\Sigma = (d\Sigma_1, \dots, d\Sigma_n),$$

gdzie

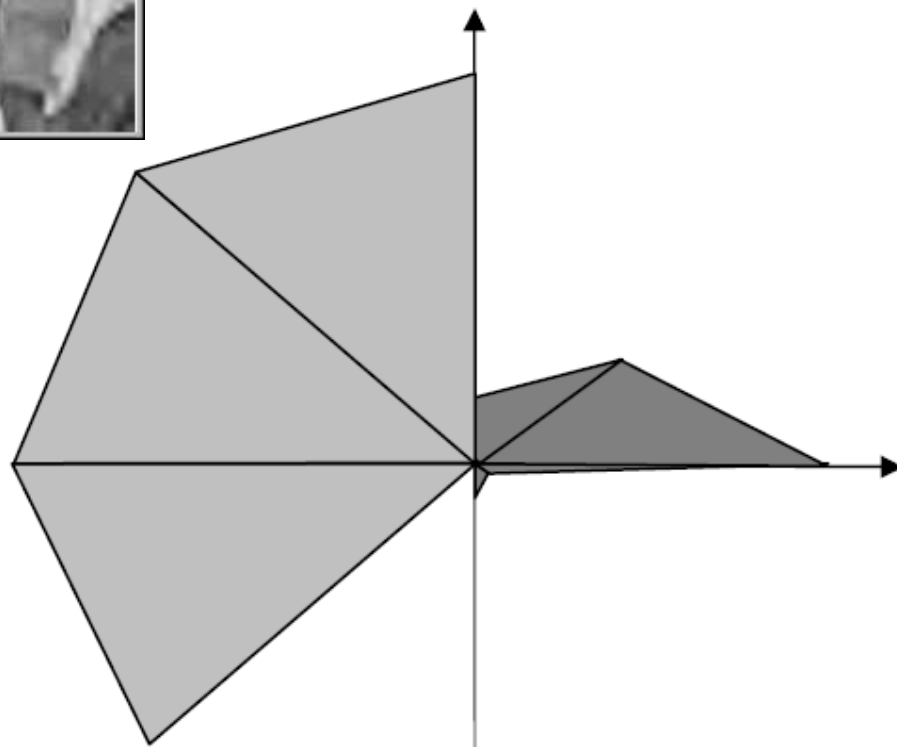
$$dM_i = |d\mu_i - d\mu'_i|, \quad d\Sigma_i = |\sigma_i - \sigma'_i|.$$

Wykresy Hosaki – forma graficzna

5. Wykreślanie różnicowych wektorów cech $\vec{d}_M$ i $\vec{d}_\Sigma$ w przestrzeni 2D.



Ocena metodami graficznymi

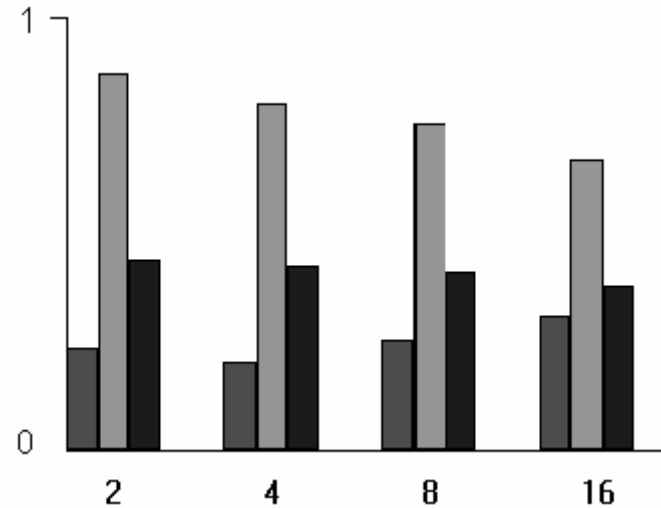
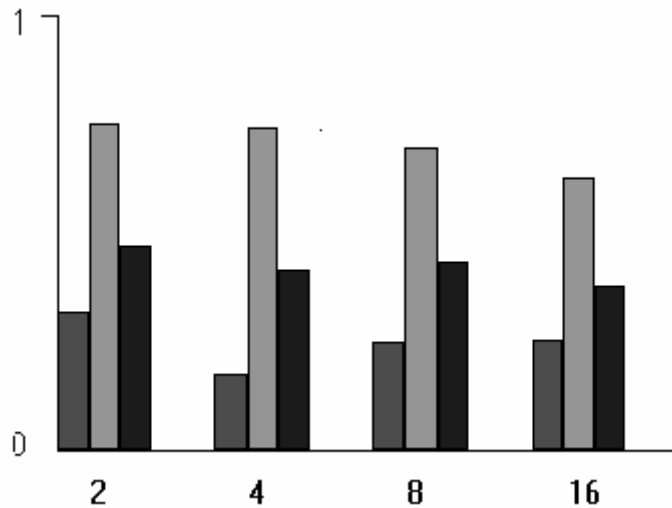
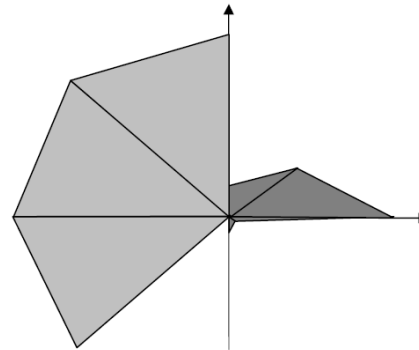


Miara Eskicioglu (algorytm), miara bezwzględna – charakterystyka obrazu

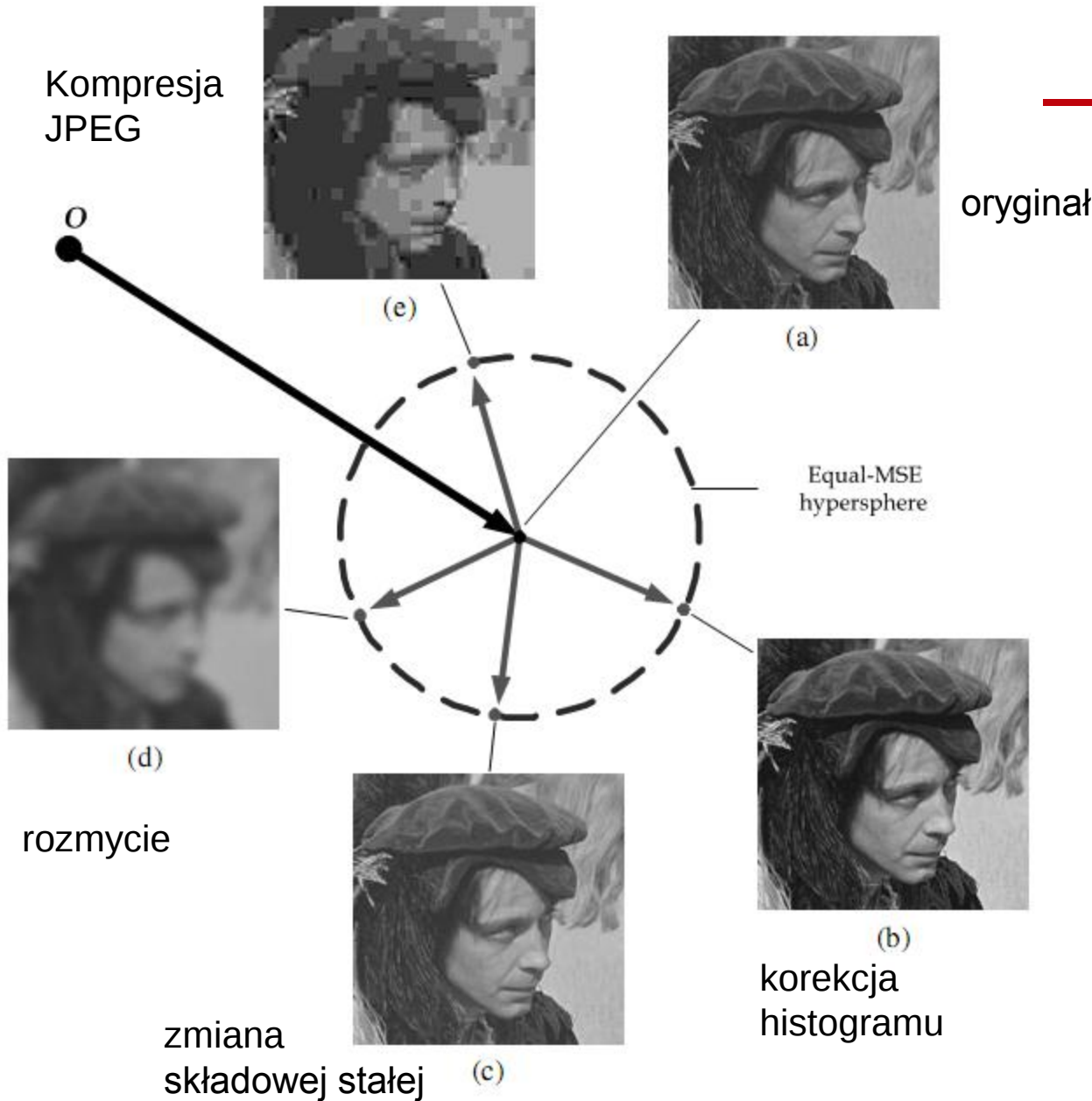
1. Segmentacja drzewa czwórkowego. Stosuje się przy tym jedynie cztery klasy: 2, 4, 8, 16.
 2. Dla każdej klasy określane są trzy cechy charakterystyczne:
 - ◆ liczba pikseli należących do bloków tej klasy/liczba wszystkich pikseli obrazu
 - ◆ liczba różnych wartości pikseli/liczba możliwych wartości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

Miara Eskicioglu (algorytm) – forma graficzna

3. Wykreślanie słupków dla każdej cechy.



Ograniczenia metod obliczeniowych



MSE=181

MODERN IMAGE
QUALITY ASSESSMENT

Zhou Wang
The University of Texas at Arlington

Alan C. Bovik
The University of Texas at Austin

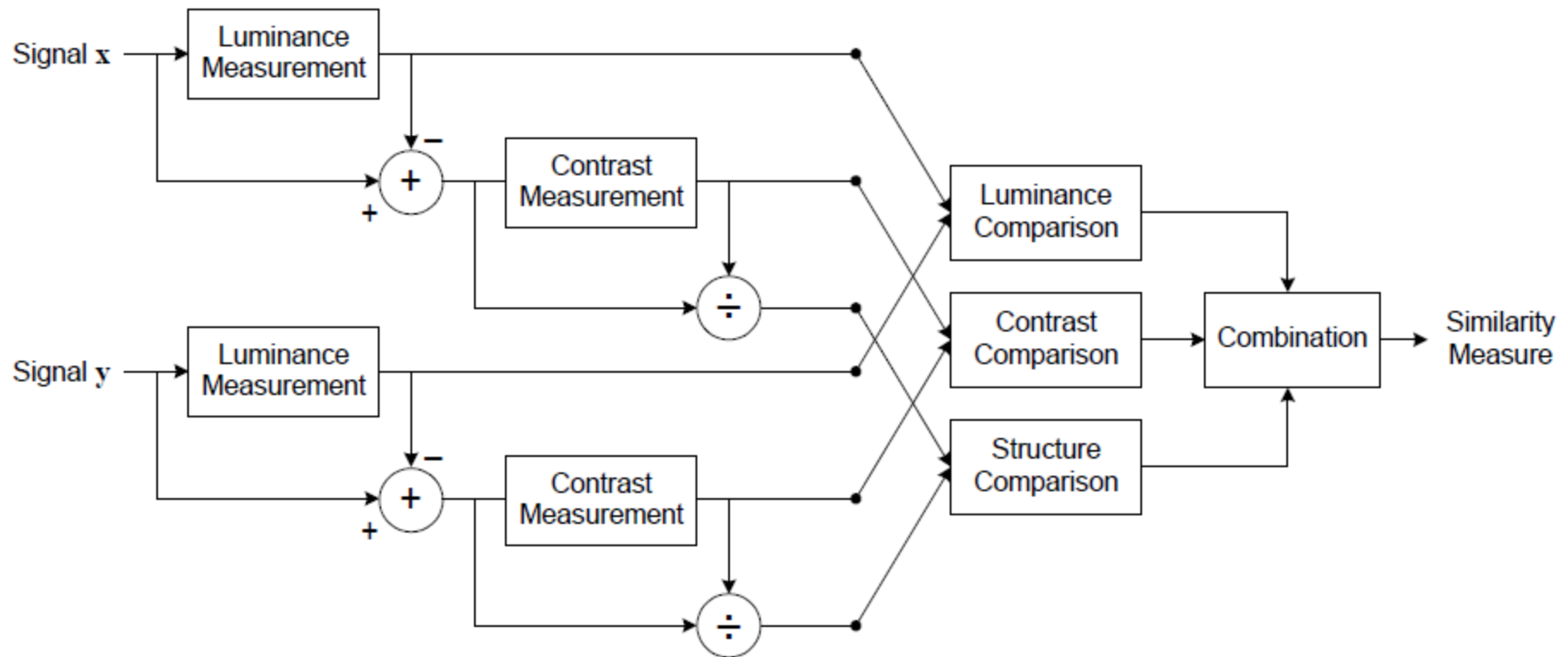
Definicja miary strukturalnej (SSIM)

- Miara podobieństwa strukturalnego w dziedzinie obrazu
- Uwzględnia też zmiany jasności i kontrastu
- Odnosi obraz badany do referencyjnego (miara porównawcza)
- Jasność: miara wartości średniej
- Kontrast: miara odchylenia standardowego
- Struktura: miara kowariancji liczonej w blokach
- Cechy: symetryczna, znormalizowana z jednoznacznym maksimum w 1 (x=y)
- Miara SSIM

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

- Realizacja: okno przesuwne 11x11, wagi przy liczeniu statystyk, warunki przesuwu okna
 - Sumaryczny wykładnik: suma - ze znormalizowanymi wagami - SSIM ze wszystkich okien
-

Konceptyjny schemat blokowy SSIM



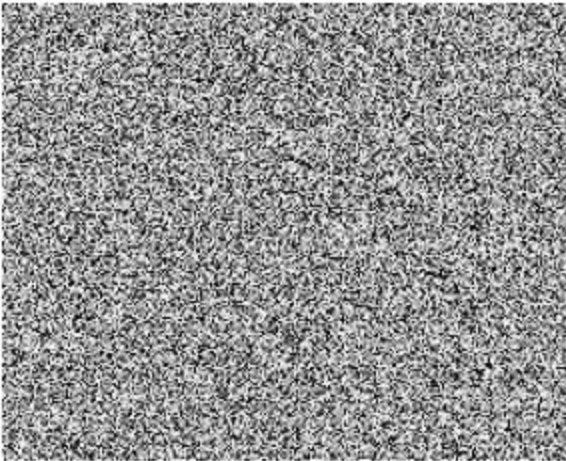
SSIM - efekty



oryginał



zasmumiony



$|zasmumiony - oryginal|$



miara SSIM

MODERN IMAGE
QUALITY ASSESSMENT

Zhou Wang
The University of Texas at Arlington

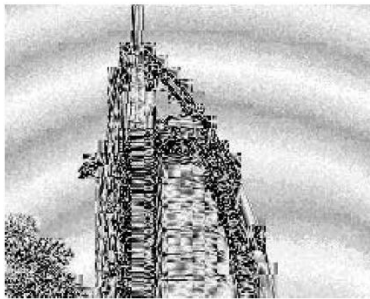
Alan C. Bovik
The University of Texas at Austin



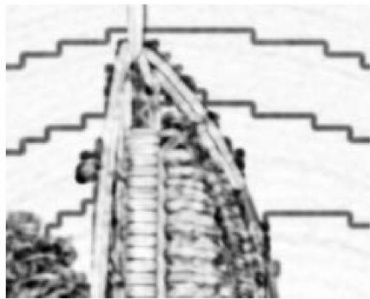
(a)



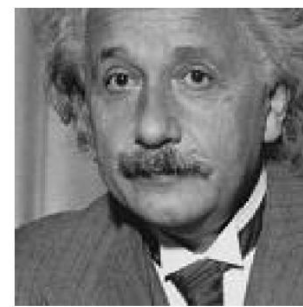
(b)



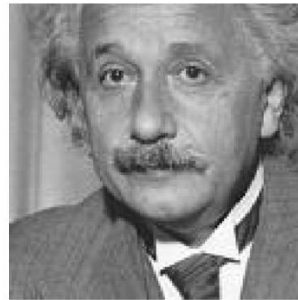
(c)



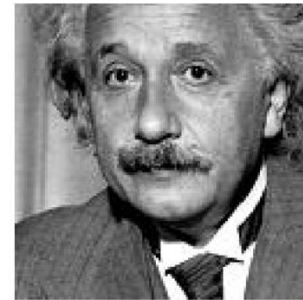
(d)



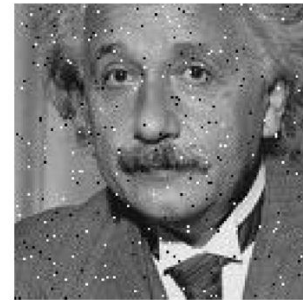
(a)



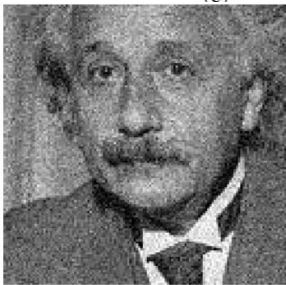
(b) MSE = 309
SSIM = 0.987
CW-SSIM = 1.000



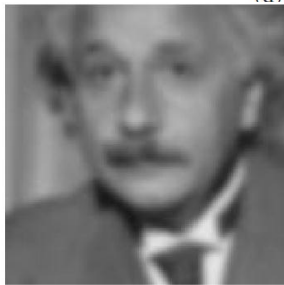
(c) MSE = 306
SSIM = 0.928
CW-SSIM = 0.938



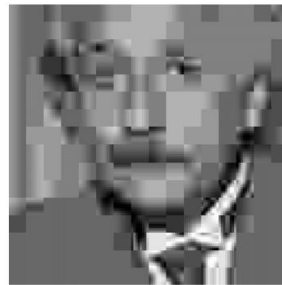
(d) MSE = 313
SSIM = 0.730
CW-SSIM = 0.811



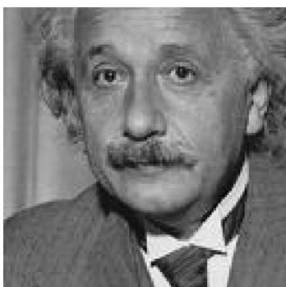
(e) MSE = 309
SSIM = 0.576
CW-SSIM = 0.814



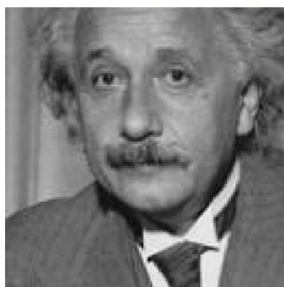
(f) MSE = 308
SSIM = 0.641
CW-SSIM = 0.603



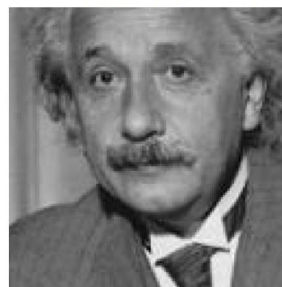
(g) MSE = 309
SSIM = 0.580
CW-SSIM = 0.633



(h) MSE = 871
SSIM = 0.404
CW-SSIM = 0.933



(i) MSE = 694
SSIM = 0.505
CW-SSIM = 0.925

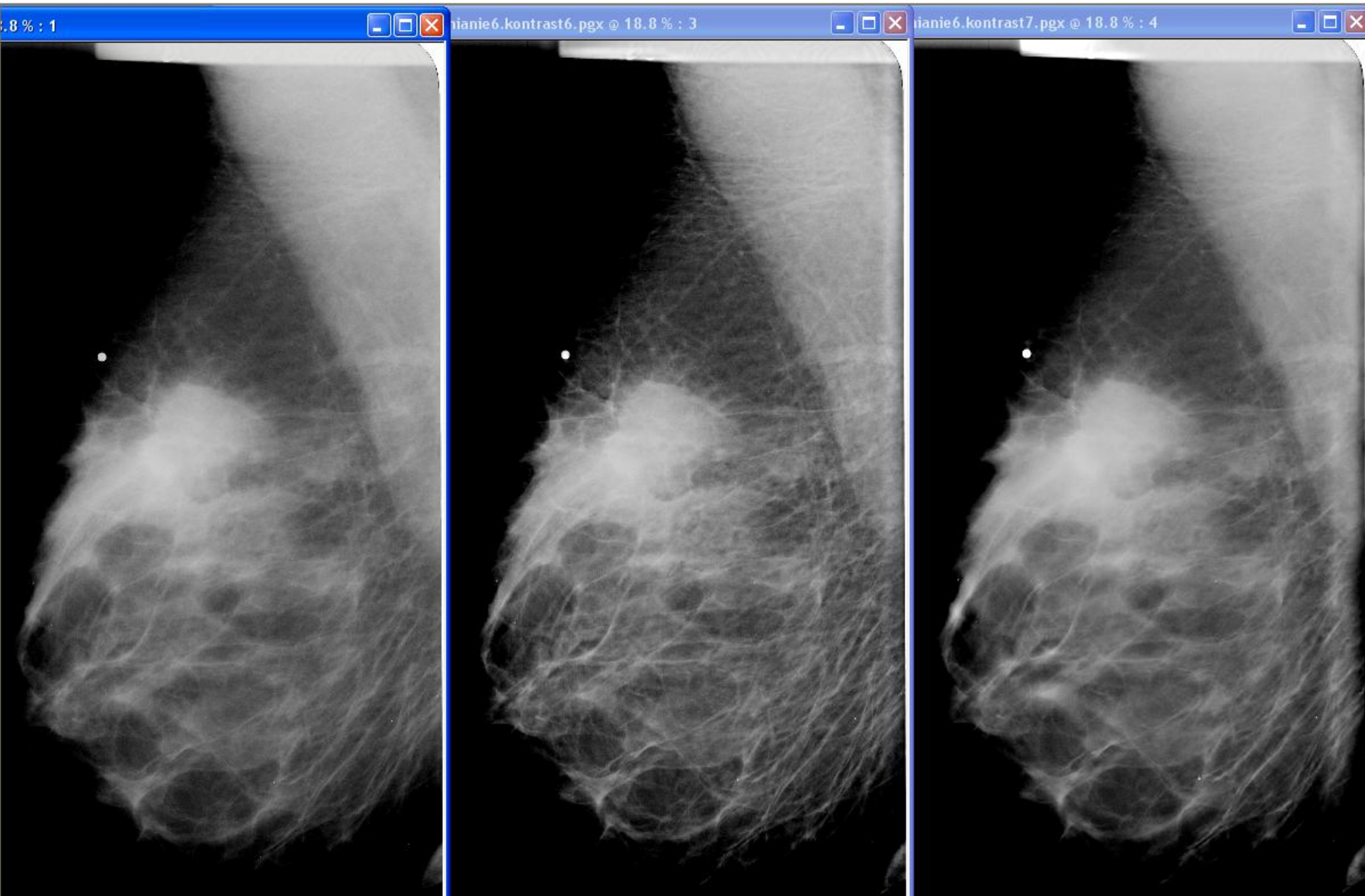


(j) MSE = 590
SSIM = 0.549
CW-SSIM = 0.917

FIGURE 3.8: (Cont.) “Einstein” image altered with different types of distortions. (a) Reference image; (b) mean luminance shift; (c) contrast stretch; (d) impulsive noise contamination; (e) Gaussian noise contamination; (f) blurring; (g) JPEG compression; (h) spatial shift (to the left); (i) spatial scaling (zooming out); and (j) rotation (counterclockwise). SSIM performs better than MSE for Images (b)–(g), but does not give consistent measurement for Images (h)–(j), which have geometric distortions. CW-SSIM overcomes this problem.

JAKOŚĆ PERCEPCJI OBRAZÓW – METODY SUBIEKTYWNE

Przykład poprawy percepcji – guz spikularny



Testy subiektywne

- Absolutne, porównawcze

- Za pomocą skali:

$$S = \sum_{k=1}^K (s_k n_k) / \sum_{k=1}^K n_k$$

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

Kategoria k	Wartość skali ocen s_k	Opis słowny charakteryzujący jakość obrazów
1	5.	Wyśmienita
2	4.	Dobra
3	3.	Średnia
4	2.	Słaba
5	1.	Zła

Kategoria k	Wartość skali ocen s_k	Opis słowny charakteryzujący jakość obrazów
1	3.	Zdecydowanie (bezwzględnie) lepsza
2	2.	Wyraźnie lepsza
3	1.	Nieznacznie lepsza
4	0.	Porównywalna z oryginałem
5	-1.	Nieznacznie gorsza
6	-2.	Wyraźnie gorsza
7	-3.	Zdecydowanie (bezwzględnie) gorsza

Poprawa percepcji mammogramu - wyniki testu subiektywnego

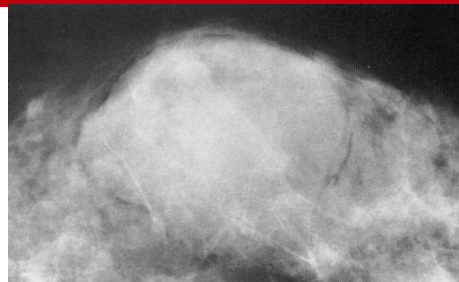
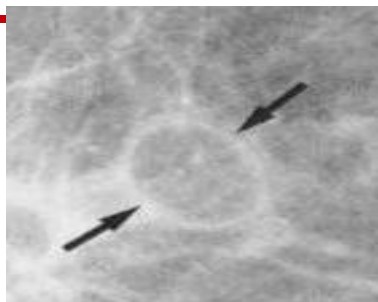
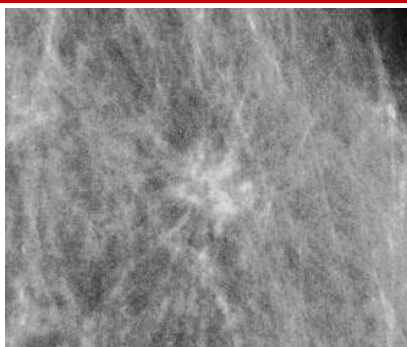
Średnia ocena lekarza		Ogólnie	Mikrozwapnienia	Guzy spikularne	Guzy dobrze odgraniczone
Zestaw parametrów	k6	+1.5	+2 (?)	+1.58	+1.25
	k7	+0.59	+0.5 (?)	+0.8	0

Średnia ocena lekarza		Subtelność (widoczność) zmiany			
		1	2	3	5
Zestaw parametrów	k6	+1	+1	+1.71	+1.5
	k7	+1	+1	+0.64	+0.33

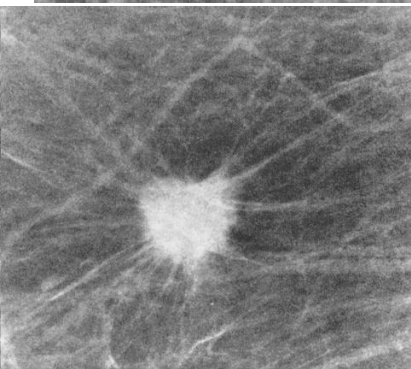
Skala ocen	Opis słowny charakteryzujący jakość obrazu
3	Zdecydowanie lepsza
2	Lepsza
1	Nieznacznie lepsza
0	Porównywalna z oryginałem
-1	Nieznacznie gorsza
-2	Gorsza
-3	Zdecydowanie gorsza

-16 obrazów z bazy DDSM o rozdzielczości 43.5 mikronów/piksel z 12-bitową głębią koloru

Testy oceny percepcji

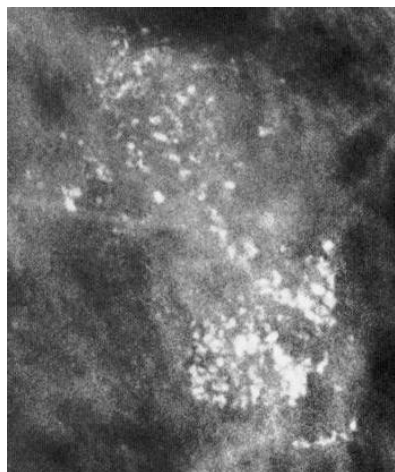


formularze testów



uj

Test OCENY	Obraz	Kontrast 1-3 (uwagi)	Ostrość 1-3 (uwagi)	Zarysy 1-3 (uwagi)	Kształt 1-3 (uwagi)
Część druga	oam				
	oa2k				
	oa0z				
	...				
	...				
	...				
	oocvbvdf				
	oogfj7				
	op4mjd8v				
Uwagi ogólne					



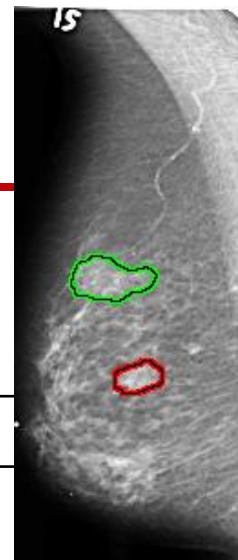
Wyniki

Lp.	Obraz	Ocena średnia					Lp.	Obraz	Ocena średnia					
		kontrast	ostrość	kształt	zarysy	suma			kontrast	ostrość	kształt	zarysy	suma	
1.	A	2,22	2,36	2,43	2,71	9,72	31.	Daj1	2,29	2,29	2,43	2,43	9,43	
2.	Aj10	2,14	2,00	2,14	2,29	8,57	32.	Daj04	1,57	1,14	1,43	1,57	5,71	
3.	Am10	2,43	2,43	2,71	2,57	10,14	33.	E	2,36	2,50	2,64	2,50	10,00	
4.	Aj6	2,00	2,00	2,14	2,14	8,29	34.	Ej10	2,71	2,57	2,86	2,71	10,86	
5.	Am6	2,29	2,00	2,14	2,14	8,57	35.	Em10	2,43	2,43	2,43	2,71	10,00	
6.	Aj1	2,29	2,14	2,43	2,29	9,14	36.	Ej6	2,43	2,43	2,43	2,29	9,57	
7.	Am1	2,29	2,00	2,43	2,57	9,29	37.	Em6	2,57	2,57	2,71	2,86	10,71	
8.	Aj04	1,29	1,00	1,00	1,14	4,43	38.	Ej1	1,71	1,29	1,71	1,86	6,57	
9.	Am04	1,71	1,29	1,86	2,29	7,14	39.	Em1	2,14	1,86	2,29	2,29	8,57	
10.	B	2,65	2,79	2,64	2,64	10,72	40.	Ej04	1,29	1,00	1,14	1,29	4,71	
11.	Bj10	2,57	2,43	2,71	2,71	10,43	41.	Em04	1,57	1,00	1,29	1,57	5,43	
12.	Bm10	2,57	2,86	2,43	2,57	10,43	42.	F	2,43	2,29	2,36	2,29	9,36	
13.	Bj6	2,29	2,29	2,43	2,57	9,57	43.	Fj10	2,57	2,57	2,29	2,29	9,71	
14.	Bm6	1,71	1,57	2,00	2,00	7,29	44.	Fm10	2,29	2,43	2,43	2,29	9,43	
15.	Bj1	1,86	1,29	1,71	2,00	6,86	45.	Fj6	2,29	2,43	1,86	1,86	8,43	
16.	Bm1	1,86	1,57	1,57	1,43	6,43	46.	Fm6	2,57	2,00	2,57	2,29	10,00	
17.	Bj04	1,14	1,00	1,00	1,00	4,14	47.	Fj1	1,57	1,57	2,00	2,00	7,14	
18.	Bm04	1,29	1,00	1,00	1,00	4,29	48.	Fm1	1,57	1,29	1,57	1,71	6,14	
19.	C	2,22	2,43	2,58	2,65	9,86	49.	Fj04	1,00	1,00	1,00	1,00	4,00	
20.	Cj10	2,57	2,86	3,00	2,86	11,29	50.	Fm04	1,14	1,00	1,14	1,29	4,57	
21.	Cm10	2,43	2,43	2,29	2,29	9,43	51.	G	2,43	2,43	2,29	2,43	9,57	
22.	Cj6	2,57	2,71	2,57	2,57	10,43	52.	Gj10	2,14	2,00	2,14	2,14	8,43	
23.	Cm6	2,14	2,29	2,14	2,29	8,86	53.	Gj6	2,14	2,29	2,29	2,14	8,86	
24.	Cj1	1,86	1,57	2,00	2,00	7,43	54.	Gj1	2,00	2,00	2,00	2,00	8,00	
25.	Cm1	2,29	1,71	2,29	2,43	8,71	55.	Gj04	1,14	1,29	1,14	1,14	4,71	
26.	Cj04	1,57	1,00	1,43	1,57	5,57	56.	H	2,29	2,43	Miary skalarne		Korelacje	
27.	Cm04	1,43	1,00	1,00	1,00	4,43	57.	Hj10	2,43	2,57	PQS: Φ_1		0.7815	
28.	Da	2,43	2,57	2,57	2,71	10,29	58.	Hj6	2,71	2,43	PQS: Φ_2		0.6115	
29.	Daj10	2,71	2,43	2,29	2,43	9,86	59.	Hj1	2,29	2,00	PQS: Φ_3		0.8112	
30.	Daj6	2,14	2,29	2,43	2,43	9,29	60.	Hj04	1,43	1,29	PQS: Φ_4		0.6374	
									PQS: Φ_5		0.7537		<i>PQS</i> (optymalizowany)	0.8459
									<i>PSNR</i>		0.5825		<i>OMW</i>	0.9028

Skala radiologiczna

Kategoria k	Wartość skali ocen s_k	Opis słowny charakteryzujący wiarygodność diagnostyczną obrazów
1	0.	Brak symptomów patologicznych
2	1.	
3	2.	Nieznacznie zarysowana zmiana, przypuszczalnie patologiczna
4	3.	
5	4.	Rozróżnialne cechy patologiczne struktur
6	5.	
7	6.	
8	7.	Wyraźne cechy o charakterze patologicznym
9	8.	
10	9.	Niewątpliwa zmiana patologiczna w obrazie
11	10.	

Skala ocen stosowana w mammografii przez radiologów

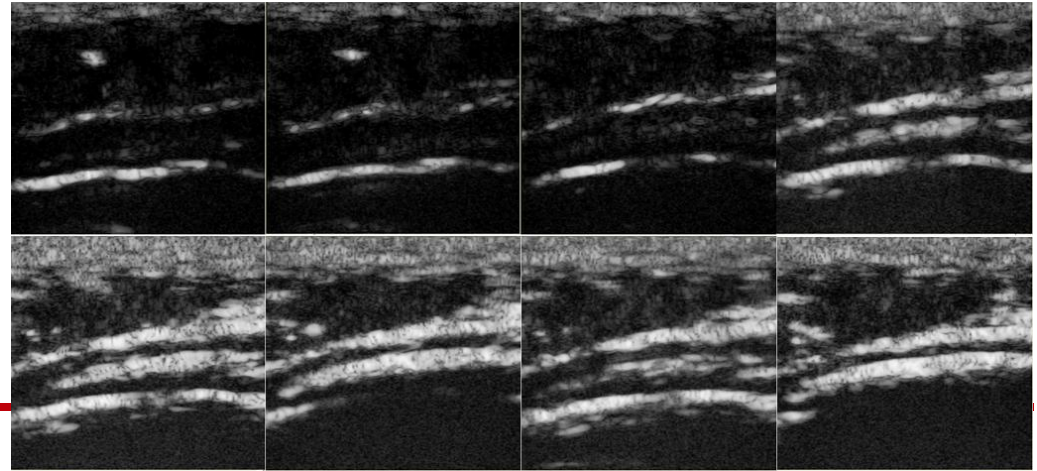
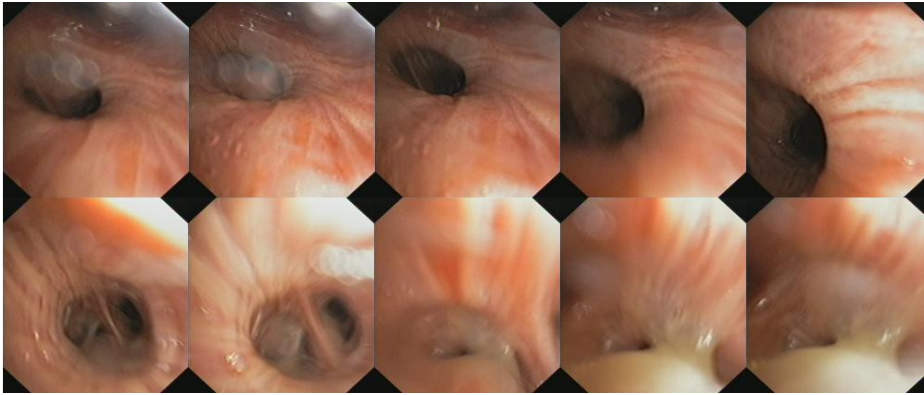


ACR-BIRADS

Kategoria	Ocena	Opis
0	Ocena niekompletna	Dodatkowe badanie obrazowe
1	Negatywna (zmiana zdecydowanie łagodna)	Bez komentarza (0 cech zmian złośliwych i wszystkie cechy zmiany łagodnej)
2	Zmiana łagodna	Zmiana zdecydowanie łagodna (0 cech zmian złośliwych)
3	Zmiana prawdopodobnie łagodna	Bardzo wysokie prawdopodobieństwo zmiany łagodnej; krótkoczasowa kontynuacja w celu uzyskania pewności diagnozy (1-2 cechy zmian złośliwych)
4	Podejrzenie patologii	Zmiana nietypowa, złośliwa ze znaczącym prawdopodobieństwem; należy rozważyć biopsję (3-4 cechy zmian złośliwych)
5	Silne przekonanie o patologii	Wysokie prawdopodobieństw rakowej zmiany złośliwej; należy podjąć odpowiednie działania (5 cech zmian złośliwych)

Ocen jakości medycznego wideo

- ❑ Video Quality Experts Group (VQEG)
- ❑ Ocena subiektywna (było wcześniej)
- ❑ Testy ITU-T ze skalami ocen (DSCQS – Double Stimulus Continuous Quality Scale według VQEG)
- ❑ Testy diagnostyczne (???)



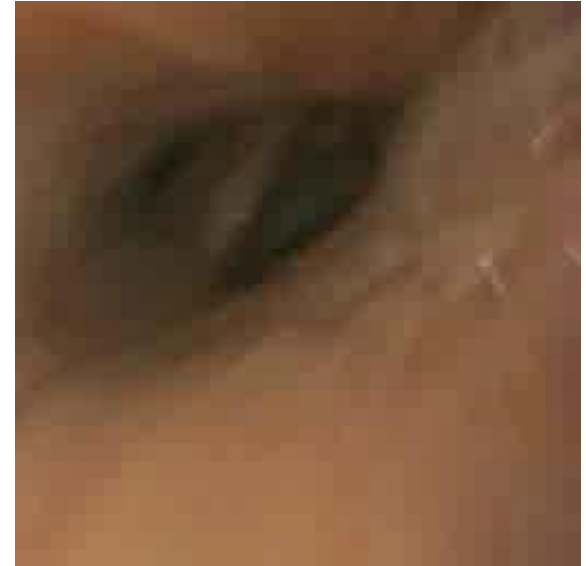
Przykładowe sekwencje testowe



Oryginał



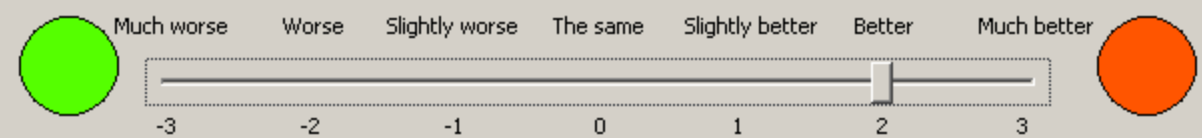
Przetworzony 1



Przetworzony 2

Give your mark! (SCACJ method)

Please, choose your opinion about the quality of the LEFT picture compared to the quality of the RIGHT picture (for example, choosing -2 or -3 means that the LEFT picture is slightly worse than the RIGHT one).



Circles symbolize your opinion on left and right video correspondingly. Red circle means that video is bad, and green means that video is good.

Your choice: 2

Watch again

OK

Test SCASJ

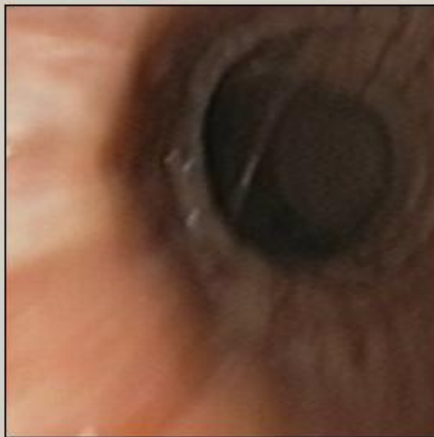


Test controls
Status: Paused
Expert: a
Task name: ocena1
Test number: 2 of 9



Hot keys
Vote against left frame: left arrow
Vote against right frame: right arrow
Play/Pause: space, mouse L
3 seconds back: backspace
Restart sequence: Ctrl + R

Sequence A



Test controls
Status: Paused
Expert: s
Task name: comvideBRONCHO
Test number: test 1 of 9, repetition 2 of 2

Hot keys
Vote against left frame: left arrow
Vote against right frame: right arrow
Play/Pause: space, mouse L
3 seconds back: backspace
Restart sequence: Ctrl + R

Test DSCQS I

Sequence B



Test controls
Status: Paused
Expert: s
Task name: comvideBRONCHO
Test number: test 1 of 9, repetition 2 of 2

Hot keys
Vote against left frame: left arrow
Vote against right frame: right arrow
Play/Pause: space, mouse L
3 seconds back: backspace
Restart sequence: Ctrl + R

Give your mark! (DSCQS method)

Please, vote on both videos.

Impairments are:



A: 5



B: 5

5, Imperceptible

4, Perceptible, but not annoying

3, Slightly annoying

2, Annoying

1, Very annoying

Watch again

OK

Test DSCQS II

Give your mark! (DSCQS method)

Please, vote on both videos.

Impairments are:



A: 5

B: 5

5, Imperceptible

4, Perceptible, but not annoying

3, Slightly annoying

2, Annoying

1, Very annoying

Watch again

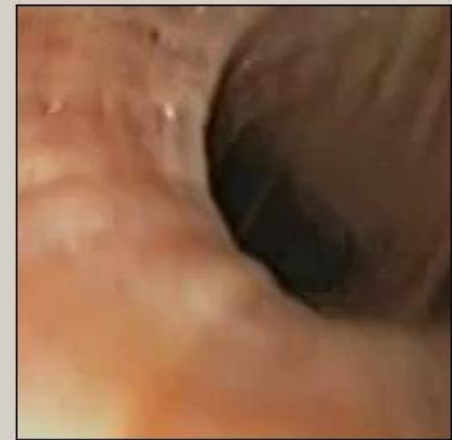
OK

eo Quality tool ver. 1.0 - player

Sequence A



Sequence B



Test controls

Status: Playing test
Expert: a
Task name: ocenavideo
Test number: test 1 of 9, repetition 1 of 2



Hot keys

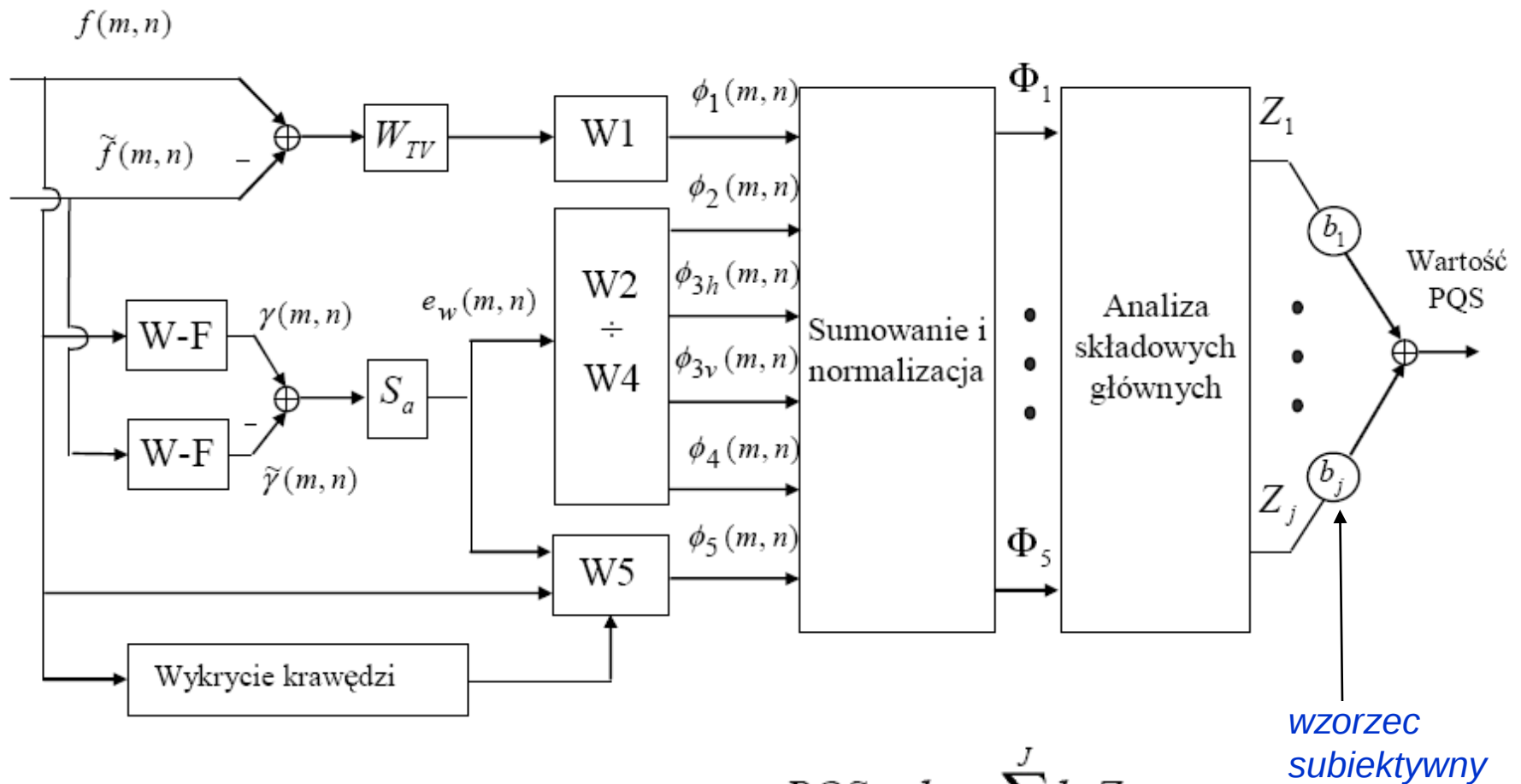
Vote against left frame: left arrow
Vote against right frame: right arrow
Play/Pause: space, mouse L
3 seconds back: backspace
Restart sequence: Ctrl + R

Wyniki oceny subiektywnej

Koder/Transkoder	bronchoskopia (95 kB)					
	Procedury oceny subiektywnej					miejsce
	SQAMVI	SCACJ	DSCQS-I	DSCQS-II	Oceny z wszystkich testów	
	MOS _{SQ}	MOS _{SC}	MOS _{D_{SI}}	MOS _{D_{SII}}	MOS _{all}	
MPEG-2	2,64	0,71	3,25	3,25	2,46	3
MPEG-4	3,06	2,11	3	3,25	2,86	2
h.264	4,31	3,29	4,2	4,1	3,98	1
MPEG-2 → MPEG-4	1,53	0,29	2	2,5	1,58	8
MPEG-2 → h.264	1,81	1,5	2,5	2,75	2,14	7
MPEG-4 → MPEG-2	1,39	0,62	2,25	2,3	1,64	9
MPEG-4 → h.264	2,36	0,86	2,5	3,1	2,21	6
h.264 → MPEG-2	2,22	0,7	3,5	2,75	2,29	5
h.264 → MPEG-4	2,78	1	2,67	3,25	2,43	4

**MIARY OBLICZENIOWE
USTALANE SUBIEKTYWNIE**

Skala jakości obrazu (PQS)



$$PQS = b_0 + \sum_{j=1}^J b_j Z_j$$

wzorec
subiektywny

PQS (zniekształcenia losowe – percepcja)

Współczynnik Φ_1

$$\Phi_1 = \frac{\sum_{m,n} \phi_1(m,n)}{\sum_{m,n} f^2(m,n)}.$$

Mapa zniekształceń w tym przypadku uwzględnia funkcję ‘ważenia’ szumu telewizyjnego $w_{TV}(\cdot)$ zdefiniowaną w standardzie CCIR 567-1 [2], która jest splatana (*) z obrazem różnicowym $e_f(\cdot)$:

$$\phi_1(m,n) = [e_f(m,n) * w_{TV}(m,n)]^2,$$

gdzie $e_f(m,n) = f(m,n) - \tilde{f}(m,n)$, a waga $w_{TV}(\cdot)$ definiowana jest w dziedzinie częstotliwościowej jako

$$W_{TV}(v) = \frac{1}{1 + (v/v_c)^2}$$

z trzy-decybelową częstotliwością graniczną $v_c = 5.56$ cykl/stopień przy odległości obserwacji równej czterokrotnej wysokości obrazu; $v = \sqrt{u^2 + v^2}$, u, v – poziome i pionowe częstotliwości przestrzenne.

PQS (zniekształcenia losowe – percepcja)

- Współczynnik Φ_2

$$\Phi_2 = \frac{\sum_{m,n} \phi_2(m,n)}{\sum_{m,n} \tilde{f}^2(m,n)},$$

przy czym mapa zniekształceń:

$$\phi_2(m,n) = I_T(m,n)[e(m,n) * s_a(m,n)]^2 \quad e_w(m,n) = e(m,n) * s_a(m,n)$$

prawo Webera-Fechnera:

$e(m,n) = \gamma(m,n) - \tilde{\gamma}(m,n)$, gdzie $\gamma(m,n) = k \cdot f(m,n)^{1/2.2}$ (k jest stałą skalującą pozwalającą dostosować dynamikę zmian wartości zmiennej γ)

$$S_a(u,v) = s(\omega)O(\omega,\theta),$$

$$\text{gdzie } s(\omega) = 1.5e^{-\sigma^2\omega^2/2} - e^{-2\sigma^2\omega^2} \text{ z } \sigma = 2, \quad \omega = \frac{2\pi v}{60}, \text{ a } O(\omega,\theta) = \frac{1 + e^{\beta(\omega-\omega_0)} \cos^4 2\theta}{1 + e^{\beta(\omega-\omega_0)}} \text{ z}$$

kątem $\theta = \tan^{-1}(u/v)$ do osi poziomej, przy czym $\beta = 8$, $v_0 = 11.13$ cykl/stopień.

Ponadto $I_T(\cdot)$ oznacza funkcję wskaźnikową dla percepcyjnego progu widzenia. Odcina ona mało znaczące wartości $e_w(\cdot)$ poniżej progu T , przyjmując dla nich wartość 0, podczas gdy dla pozostałych - wartość równą 1. Przyjęto $T = 1$.

PQS

- Współczynnik Φ_3 (dotyczy efektów blokowych)

$$\Phi_3 = \sqrt{\Phi_{3h}^2 + \Phi_{3v}^2}$$

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

- Współczynnik Φ_4 (błędy skorelowane)

$$\Phi_4 = \frac{1}{MN} \sum_{m,n} \phi_4(m,n)$$

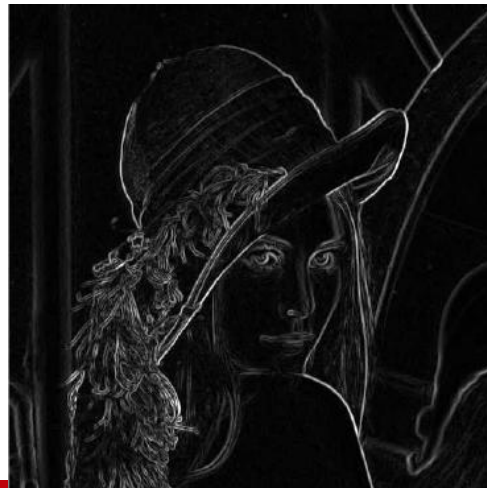
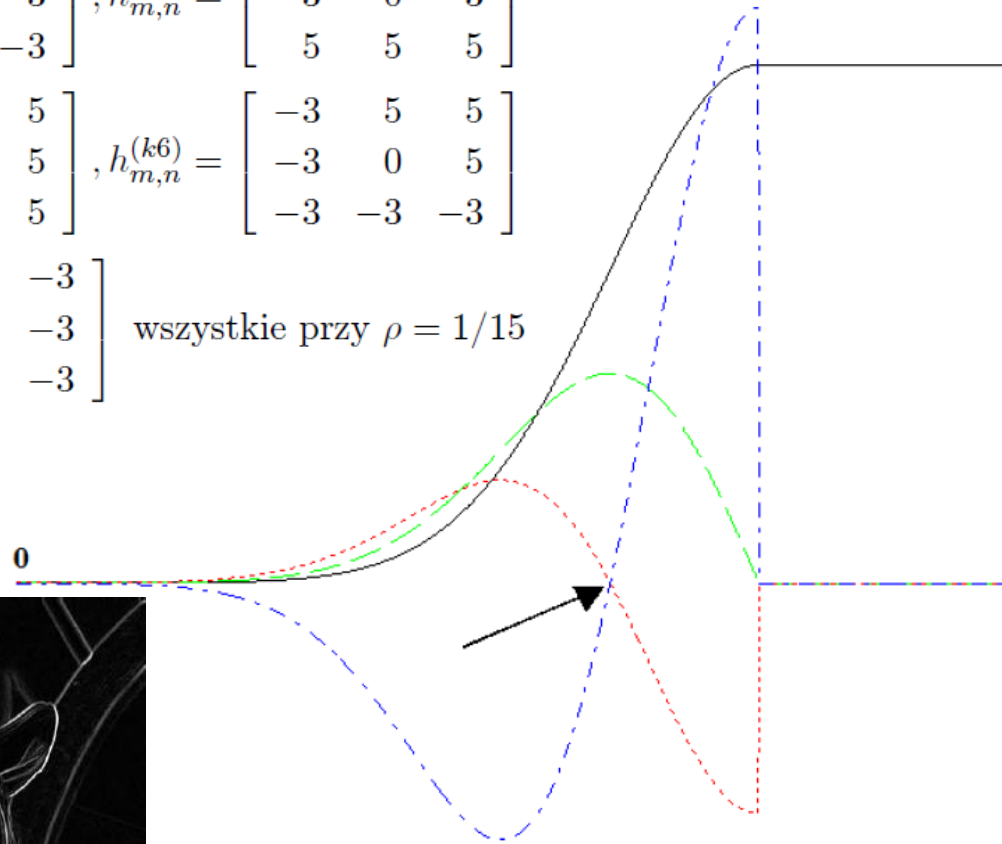
$$\phi_4(m,n) = \sum_{(k,l) \in W} |r(m,n,k,l)|^{0.25},$$

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

Sumowanie odbywa się po zbiorze pikseli, dla których (i,j) oraz $(i+k, j+l)$ leżą w oknie W o rozmiarach 5×5 i środku w punkcie (m,n) .

Wyznaczanie krawędzi filtrem Kirscha

$$\begin{aligned}
 h_{m,n}^{(k1)} &= \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}, h_{m,n}^{(k2)} = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}, h_{m,n}^{(k3)} = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \\
 h_{m,n}^{(k4)} &= \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}, h_{m,n}^{(k5)} = \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}, h_{m,n}^{(k6)} = \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \\
 h_{m,n}^{(k7)} &= \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, h_{m,n}^{(k8)} = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \text{ wszystkie przy } \rho = 1/15
 \end{aligned}$$



PQS (krawędzie)

- Współczynnik Φ_5 (dotyczy błędów w sąsiedztwie wyraźnych krawędzi):

$$\Phi_5 = \frac{1}{N_K} \sum_{m,n} \phi_5(m,n),$$

gdzie N_K jest liczbą pikseli, których odpowiedź krawędzi Kirscha o rozmiarze 3×3 jest większa lub równa stałej $K=400$. Współczynnik ten związany jest z psychowizualnym efektem występującym przy obserwacji zniekształconych obrazów. W sąsiedztwie struktur o dużych gradientach (czy ogólniej w obszarach o dużych zmianach kontrastu) występuje redukcja widoczności zniekształceń, co nazywane jest maskowaniem widzenia (ang. *visual masking*). Mapa zniekształceń opisana równaniem:

$$\phi_5(m,n) = I_M(m,n) \cdot |e_w(m,n)| \cdot (S_h(m,n) + S_v(m,n))$$

jest miarą poziomu zniekształceń w sąsiedztwie wyraźnych krawędzi struktur. Zawiera więc wskaźnik maskowania w kierunku poziomym:

$$S_h(m,n) = e^{\{-0.04V_h(m,n)\}},$$

gdzie miara lokalnej aktywności (w poziomie): $V_h(m,n) = \frac{|f(m,n-1) - f(m,n+1)|}{2}$ oraz

wskaźnik maskowania w kierunku pionowym $S_v(\cdot)$, zdefiniowany analogicznie. Funkcja $I_M(\cdot)$ wskazuje na piksele leżące blisko aktywnych regionów obrazu, określone za pomocą wspomnianego testu z odpowiedzią krawędzi Kirscha.

Miary jakości wideo

- Ocena numeryczna
 - Miary klasyczne (MSE, PSNR, AD, ...)
 - Dostosowanie SSIM
 - Miary z modelem HVS
 - Miary z modelem HVS dla wideo
-

V-SSIM (video structural similarity) - jakość lokalnych struktur, jasność, kontrast oraz estymacja ruchu

- Lokalna miara SSIM

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

- SSIM dla ramki (uśredniamy po 6% pikseli z najmniejszym SSIM(x,y))

$$S_{frame} = \frac{1}{|\psi|} \sum_{i \in \psi} SSIM(i)$$

- Bryła

$$\begin{aligned} & (i - \lfloor \alpha/2 \rfloor, j - \lfloor \beta/2 \rfloor, k - \lfloor \gamma/2 \rfloor), & (i - \lfloor \alpha/2 \rfloor, j - \lfloor \beta/2 \rfloor, k + \lfloor \gamma/2 \rfloor), \\ & (i - \lfloor \alpha/2 \rfloor, j + \lfloor \beta/2 \rfloor, k - \lfloor \gamma/2 \rfloor), & (i - \lfloor \alpha/2 \rfloor, j + \lfloor \beta/2 \rfloor, k + \lfloor \gamma/2 \rfloor), \\ & (i + \lfloor \alpha/2 \rfloor, j - \lfloor \beta/2 \rfloor, k - \lfloor \gamma/2 \rfloor), & (i + \lfloor \alpha/2 \rfloor, j - \lfloor \beta/2 \rfloor, k + \lfloor \gamma/2 \rfloor), \\ & (i + \lfloor \alpha/2 \rfloor, j + \lfloor \beta/2 \rfloor, k - \lfloor \gamma/2 \rfloor) & (i + \lfloor \alpha/2 \rfloor, j + \lfloor \beta/2 \rfloor, k + \lfloor \gamma/2 \rfloor). \end{aligned} \quad \text{i}$$

- Statystyki (w – wagi filtrów dobierane na podstawie estymacji ruchu)

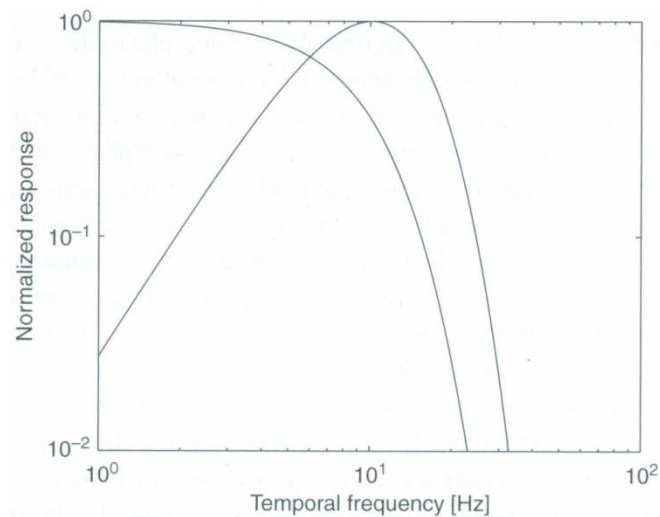
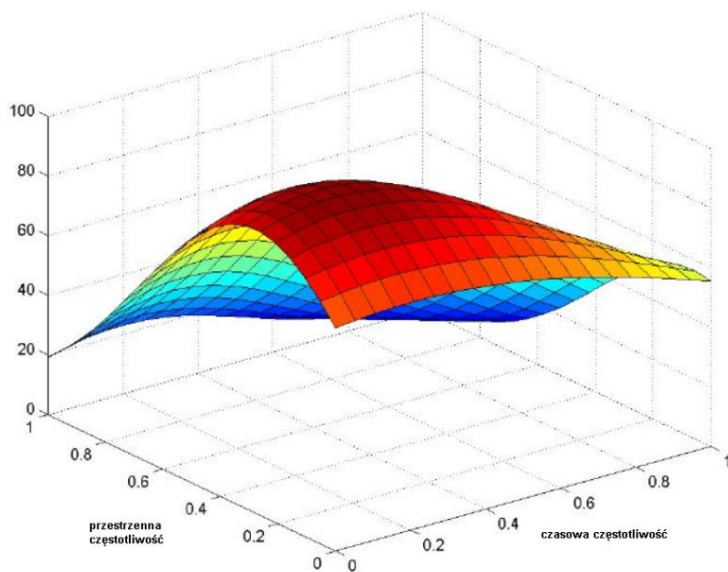
$$\mu_{x(i,j,k)} = \sum_{m=1}^{\alpha} \sum_{n=1}^{\beta} \sum_{o=1}^{\gamma} w(m, n, o) x(m, n, o) \quad \sigma_{x(i,j,k)}^2 = \sum_{m=1}^{\alpha} \sum_{n=1}^{\beta} \sum_{o=1}^{\gamma} w(m, n, o) (x(m, n, o) - \mu_{x(i,j,k)})^2$$

$$\sigma_{x(i,j,k)y(i,j,k)} = \sum_{m=1}^{\alpha} \sum_{n=1}^{\beta} \sum_{o=1}^{\gamma} w(m, n, o) (x(m, n, o) - \mu_{x(i,j,k)})(y(m, n, o) - \mu_{y(i,j,k)})$$

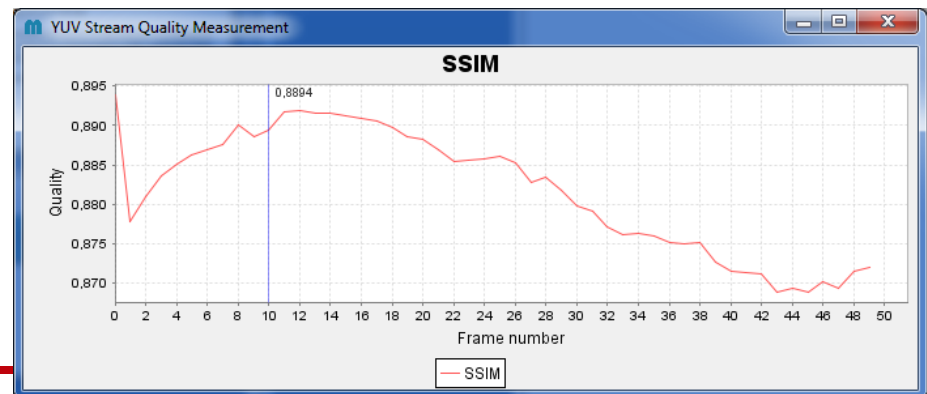
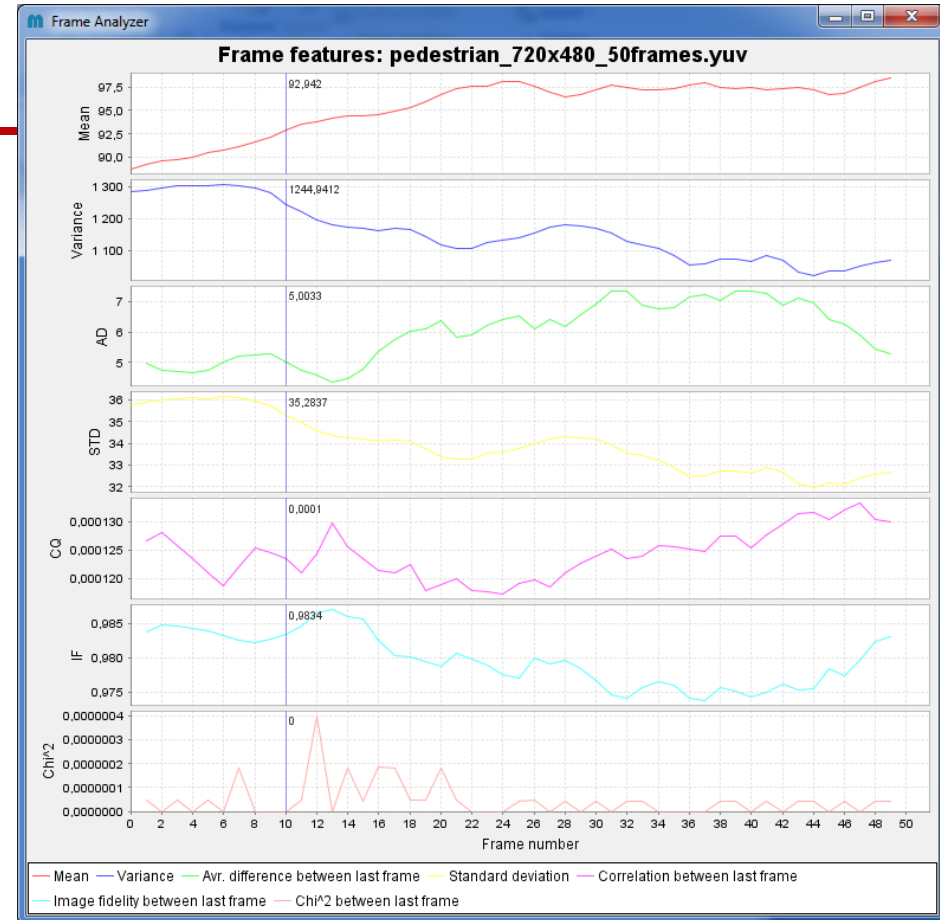
$$V\text{-SSIM} = \frac{(2\mu_{x(i,j,k)}\mu_{y(i,j,k)} + C_1)(2\sigma_{x(i,j,k)y(i,j,k)} + C_2)}{(\mu_{x(i,j,k)}^2 + \mu_{y(i,j,k)}^2 + C_1)(\sigma_{x(i,j,k)}^2 + \sigma_{y(i,j,k)}^2 + C_2)}$$

Zaawansowane miary jakości

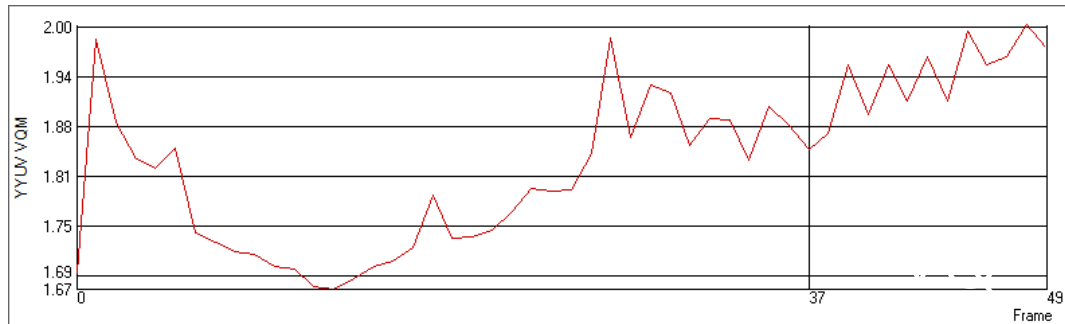
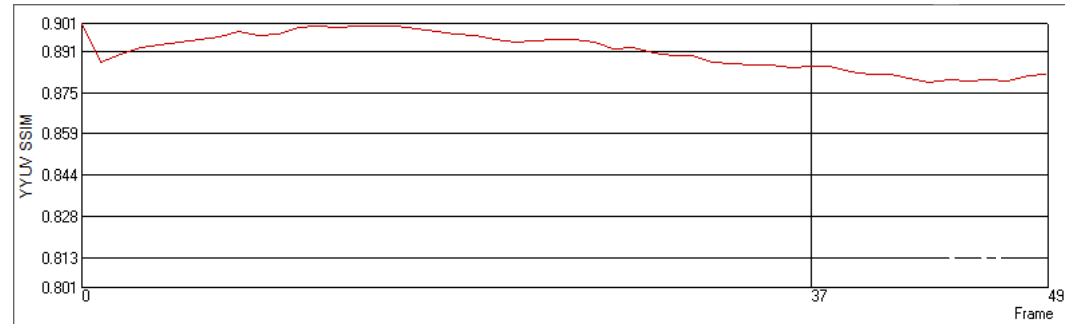
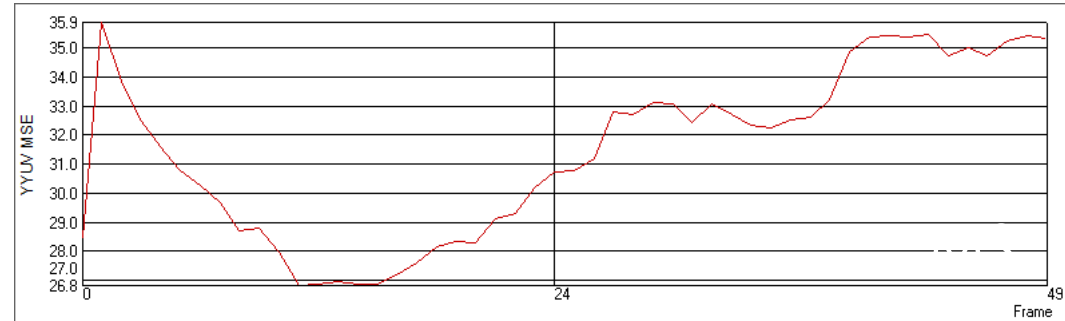
- PDM (perceptual distortion metric) – rozbudowane wielokanałowe modele percepcji
- VQM (video quality measure) – modele w dziedzinie DCT



Przykłady - jakość wideo

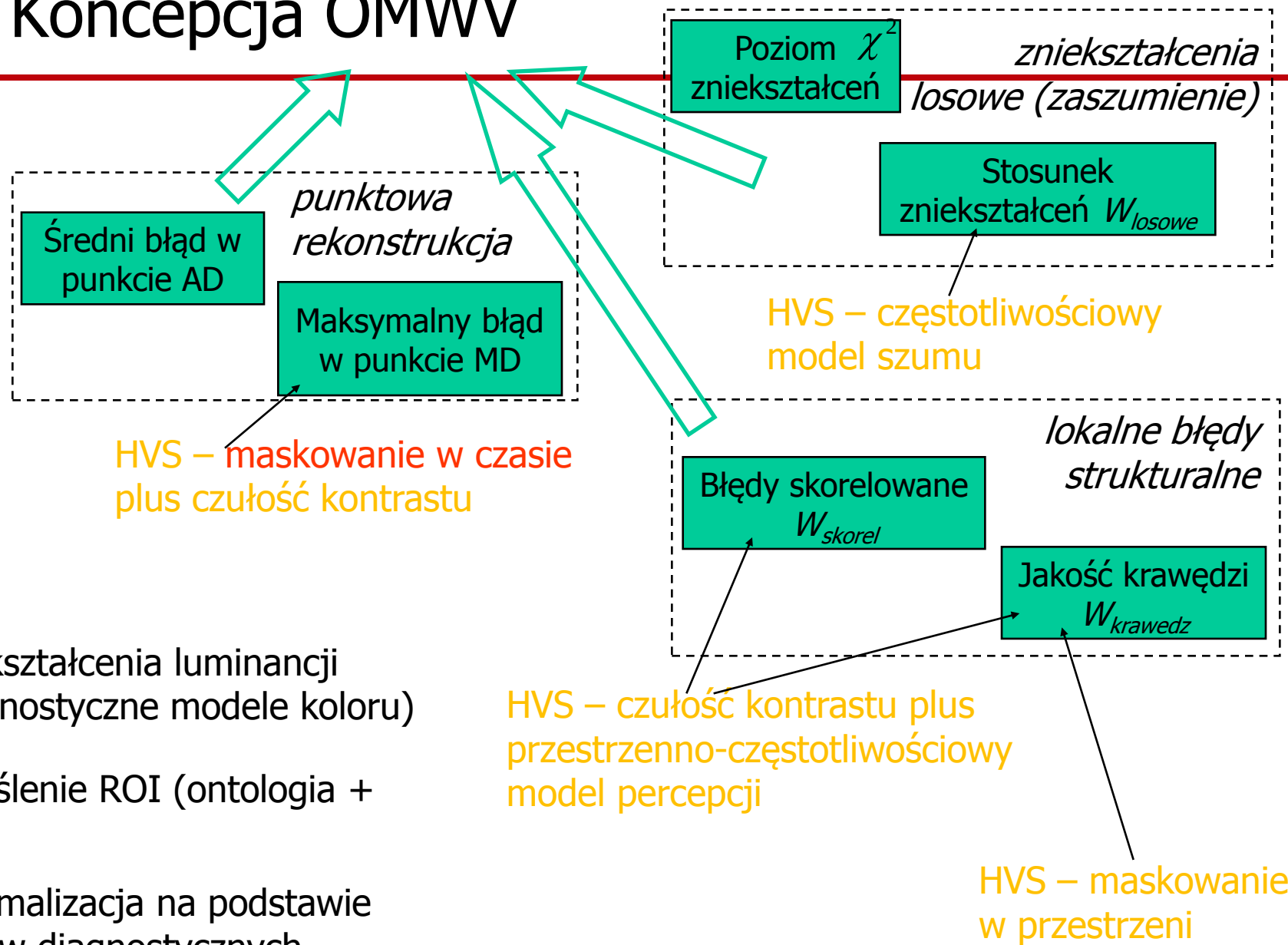


Zróżnicowanie ocen obliczeniowych



A.Zubala

Koncepcja OMWV



Zniekształcenia luminancji
(diagnostyczne modele koloru)

Określenie ROI (ontologia +
CAD)

Optymalizacja na podstawie
testów diagnostycznych

Miara OMWV

■ 3 w 1

$$OMWV = \alpha_1 \cdot AD + \alpha_2 \cdot (\beta_1 + \beta_2) \cdot MD + \alpha_3 \cdot \chi^2 + \alpha_4 \cdot W_{losowe} + \alpha_5 \cdot W_{skorel} + \alpha_6 \cdot W_{krawedz}$$

Przykładowo (dla badań bronchoskopowych):

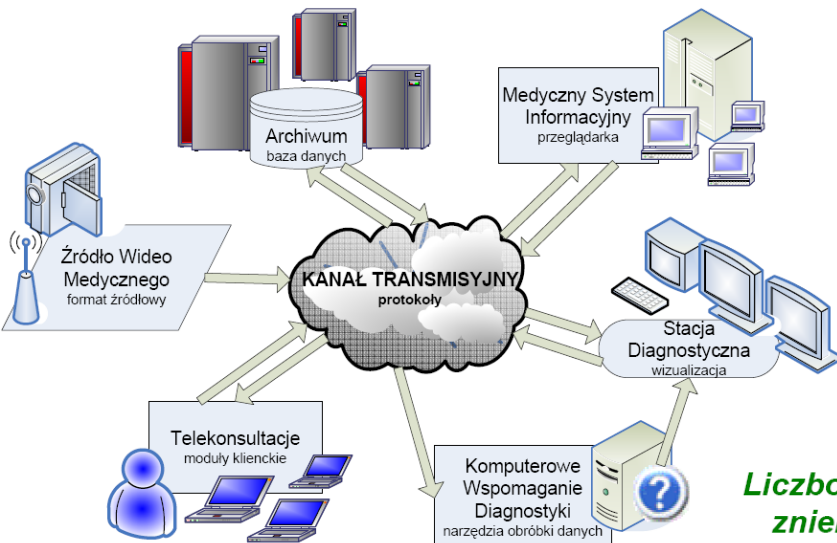
$$\alpha_1 = 3, \alpha_2 = 1, \alpha_3 = 5, \alpha_4 = 500, \alpha_5 = 1, \alpha_6 = 0,2$$

Model uwzględniający zależności czasowe (wideo):

$$\beta_1 = 0,1 \cdot \bar{x}_i / |\bar{x}_i - \bar{x}_{i-1}|$$

$$\beta_2 = 0,1 \cdot \sigma_i / |\sigma_i - \sigma_{i-1}|$$

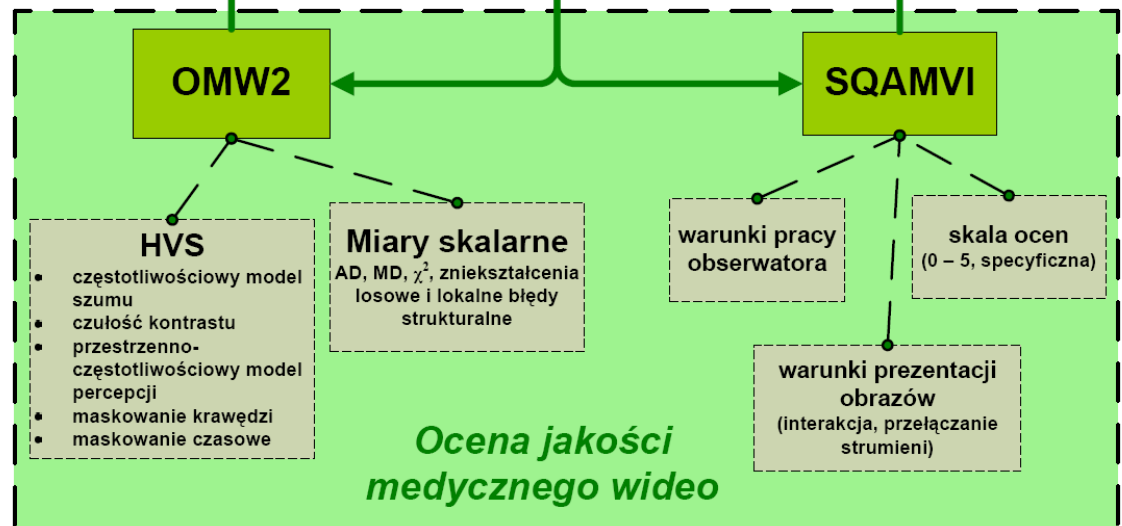
Koncepcje i metody



Liczbowy poziom zniekształceń
(automatyczna kontrola jakości, optymalizacja kodowania)

Wideo
(źródłowe/ kodowane/ transkodowane)

MOS
(weryfikacja, optymalizacja OMW2)



Wyniki badań oceny jakości

Kod/ Trans	bronchoskopia (95 kB)						USG z kontrastem (330 kB)					
	MOS	PSNR	SSIM	VQM	OMW	OMWV	MOS	PSNR	SSIM	VQM	OMW	OMWV
2	2,64	38,18	,9139	0,982	60,99	64,09	1,67	30,67	0,7541	3,0298	149,0765	156,2585
4	3,06	39,08	,9371	0,872	54,63	56,33	4,31	32,21	0,8067	2,6845	124,1832	128,3887
264	4,31	41,05	,9657	0,737	43,67	43,30	3,61	32,74	0,8177	2,6381	117,2832	119,6674
2_4	1,53	35,85	,9087	1,185	69,73	72,30	2,64	30,31	0,7493	3,2564	153,7465	160,0517
2_264	1,81	36,33	,9251	1,110	67,51	71,46	1,94	30,49	0,7590	3,1818	151,9815	159,8256
4_2	1,39	35,91	,8967	1,212	69,96	72,51	4,03	30,80	0,7624	3,1876	145,2072	150,2352
4_264	2,36	36,32	,9372	1,090	65,10	67,87	3,75	31,55	0,7932	2,9454	132,6354	136,473
264_2	2,22	36,68	,9063	1,127	67,09	70,34	5	30,88	0,7563	3,1810	142,7504	145,2974
264_4	2,78	36,35	,9313	1,105	64,55	66,05	3,47	31,73	0,7920	2,9213	129,2177	130,7699
Korelacja z MOS		,9095	,8717	-,9354	-,9557	-,9569	→	,4467	,3847	-,2477	-,5117	-,5648

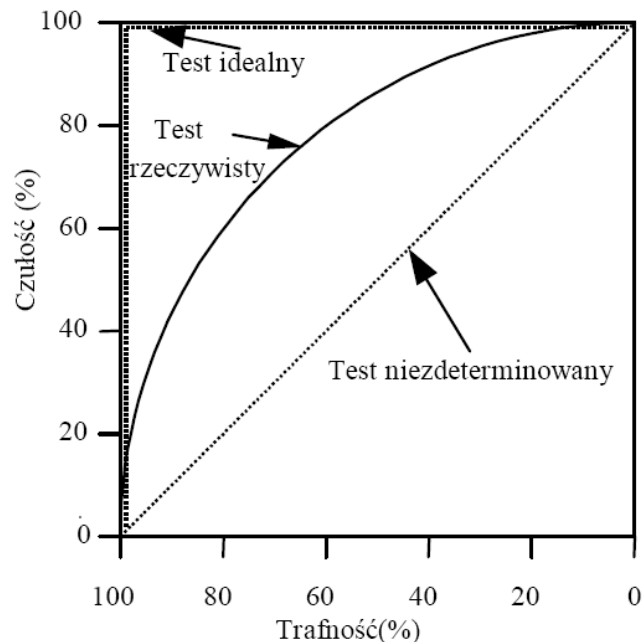
WIARYGODNOŚĆ OBRAZÓW – OCENA

Ocena wiarygodności diagnostycznej obrazów

- Potrzeba dużych zbiorów wiarygodnych danych testowych (wysokie koszty)
 - Konieczny jest wiarygodny zespół ekspertów do weryfikacji (wysokie koszty)
 - Ustalenie 'złotego standardu' (wysokie koszty)
 - Projektowanie (uwarunkowania) i realizacja testu (wysokie koszty)
 - Statystyczne metody analizy, weryfikacja hipotez (trudne zadanie)
 - Formułowanie wniosków
-

Testy oceny klinicznej

- Reprezentatywny zbiór badań testowych
- Detekcja patologii
- Czułość, specyficzność
- Krzywa ROC
- Weryfikacja hipotez statystycznych



- 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.

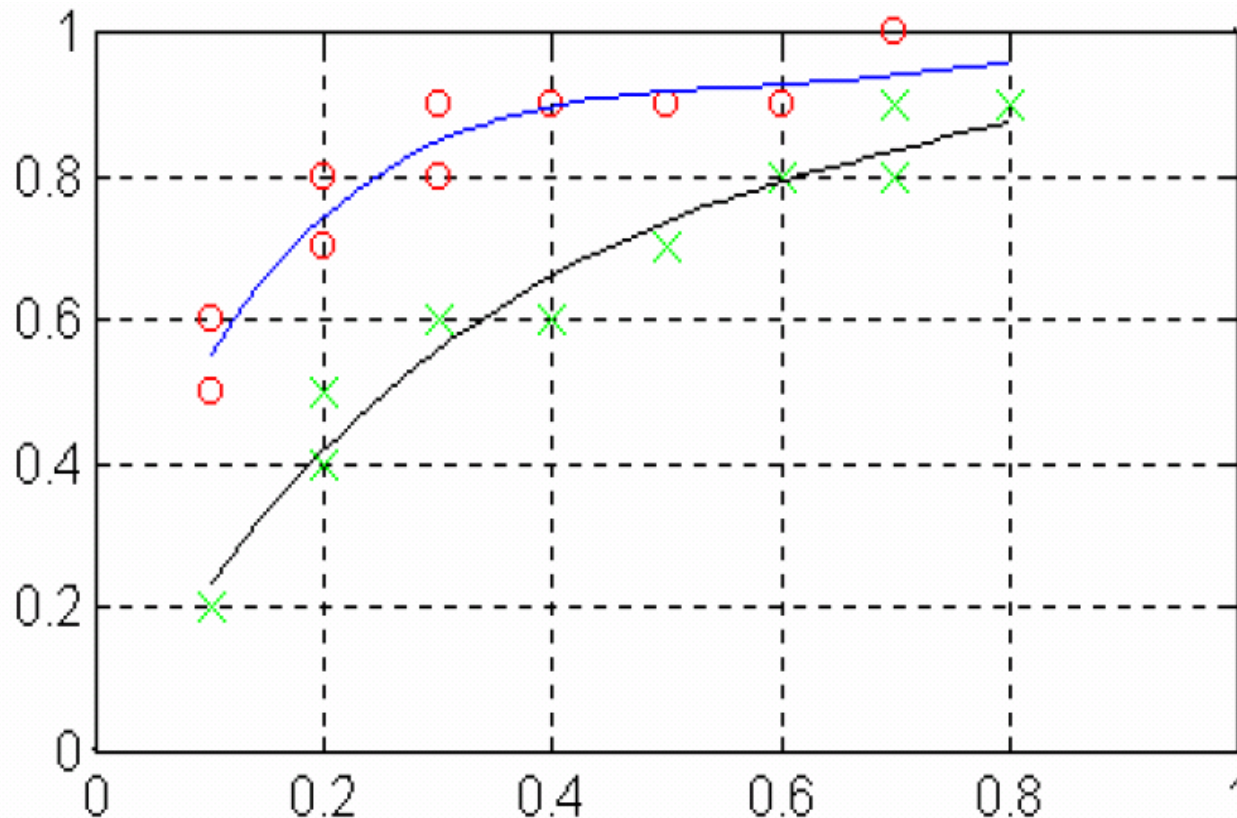
Dalsza charakterystyka

Cechy:

- częste wykorzystywanie krzywej ROC, stosowanie wielostopniowej skali ocen np. pięciostopniowej (pewna cecha jest zdecydowanie obecna, prawdopodobnie obecna, może obecna, prawdopodobnie nieobecna lub też definitywnie nieobecna),
 - specjaliści 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.
-

Przypomnienie – wykreślanie krzywej

	lekarze	1	2	3	4	5	6	7	8	9	10
CR=10	Npp	9	5	6	9	9	10	8	8	9	7
	Nfp	5	1	1	3	4	7	2	3	6	2
CR=25	Npp	6	2	5	7	9	9	8	6	8	4
	Nfp	4	1	2	5	7	8	6	3	7	2

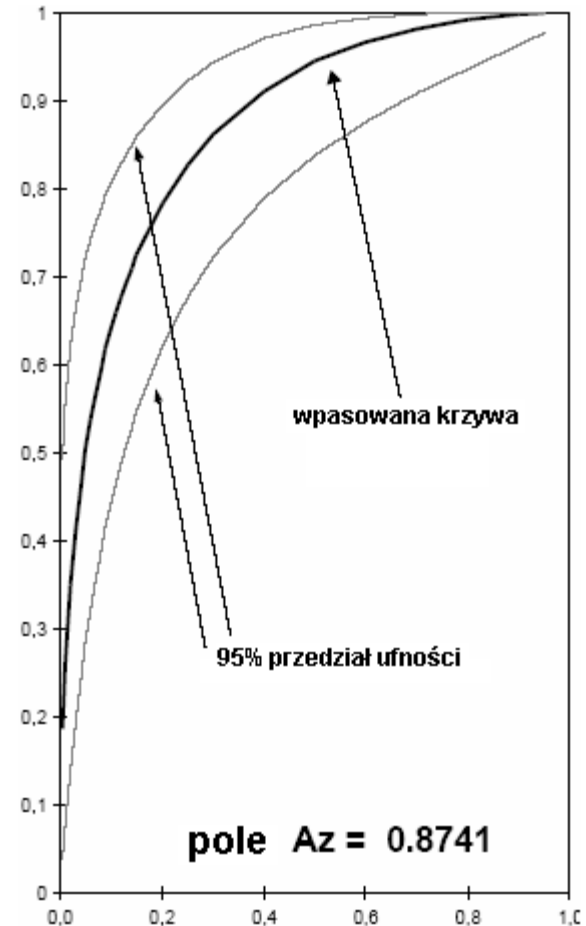


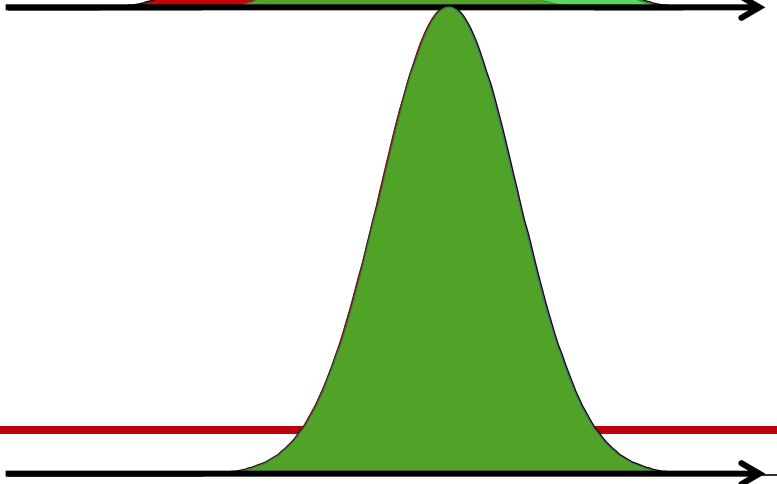
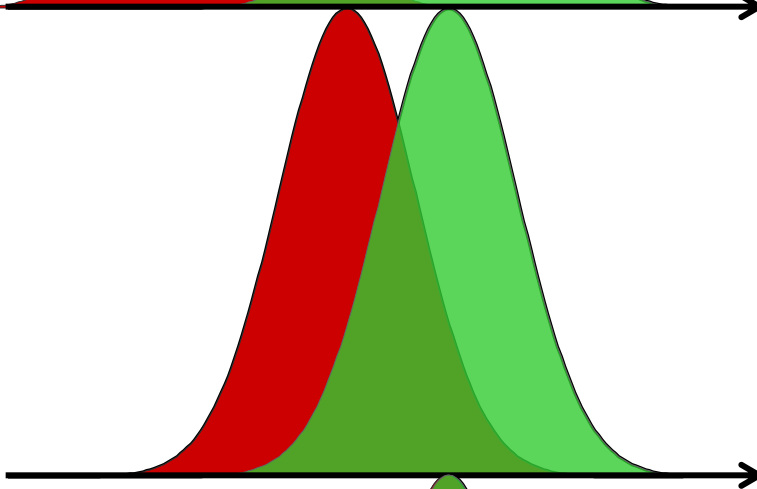
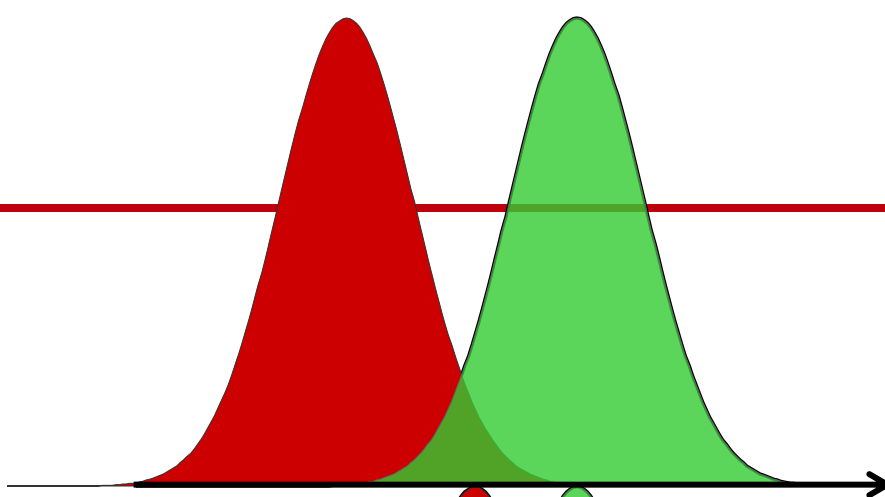
$N_{pat}=10$

$N_{norm}=10$

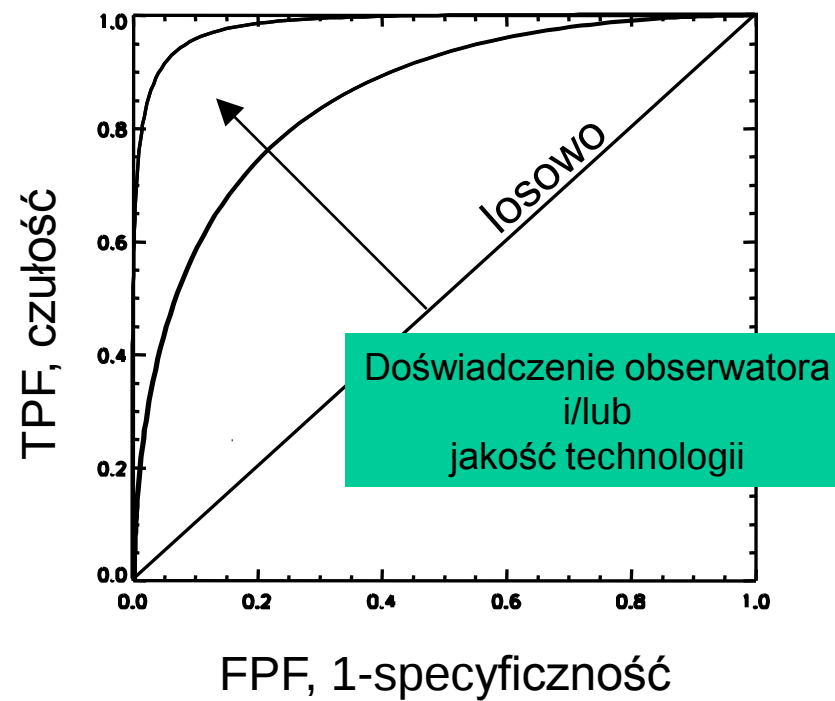
Krzywa ROC

- dwie populacje danych: z patologią i bez
- opracowany 'złoty standard'
- test różnicujący krzywe (poziomy istotności, przedziały ufności, testy jedno i dwuwymiarowe)
- parametry badane: czułość, specyficzność, pole pod krzywą
- weryfikacja hipotez statystycznych: testy parametryczne 1W i 2W z poziomem istotności (jakościowe) lub z przedziałami ufności (ilościowe)





krzywa ROC



Test z porównaniem wartości średniej

prosty test ze statystyką U

Mamy: dwie duże, niezależne próby pobrane z populacji niekoniecznie normalnych, o nieznanymi wartościach 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:

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

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

Weryfikacja hipotezy

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

$$|u_0| > u_\alpha,$$
$$u_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

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 odrzucaamy H_0 na korzyść H_2 jeśli $u_0 > u_\alpha$ itd.

Testy o innych założeniach

t-Studenta

W tych zastosowaniach można użyć następującego testu:

weryfikujemy hipotezę, że średnie dwóch małych prób nie różnią się istotnie, czyli

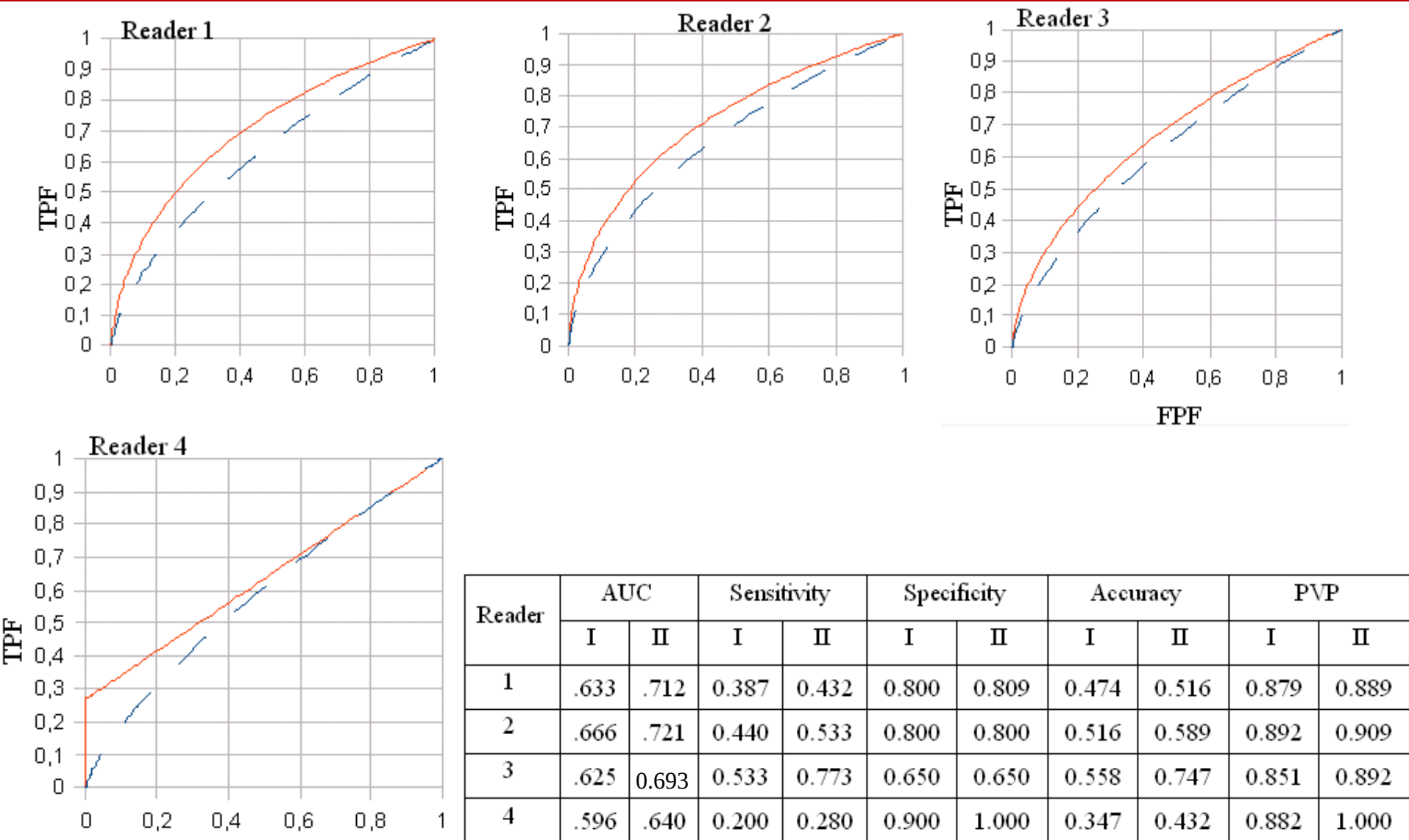
$H_0: m_1 = m_2$. Przyjmujemy, że próby są niezależne, a populacje w przybliżeniu normalne o nieznanym, lecz równym wariancjach (niestety niezbyt spełnione w tym przypadku). Podstawą testu jest statystyka:

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

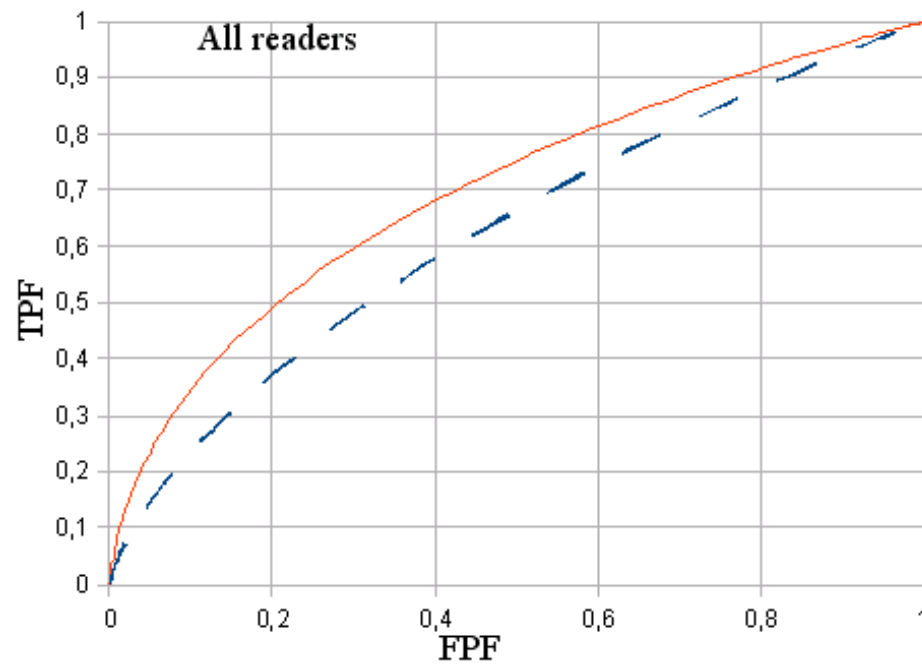
Ma ona rozkład t-Studenta o $n_1 + n_2 - 2$ stopniach swobody

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

Przykładowa ocena

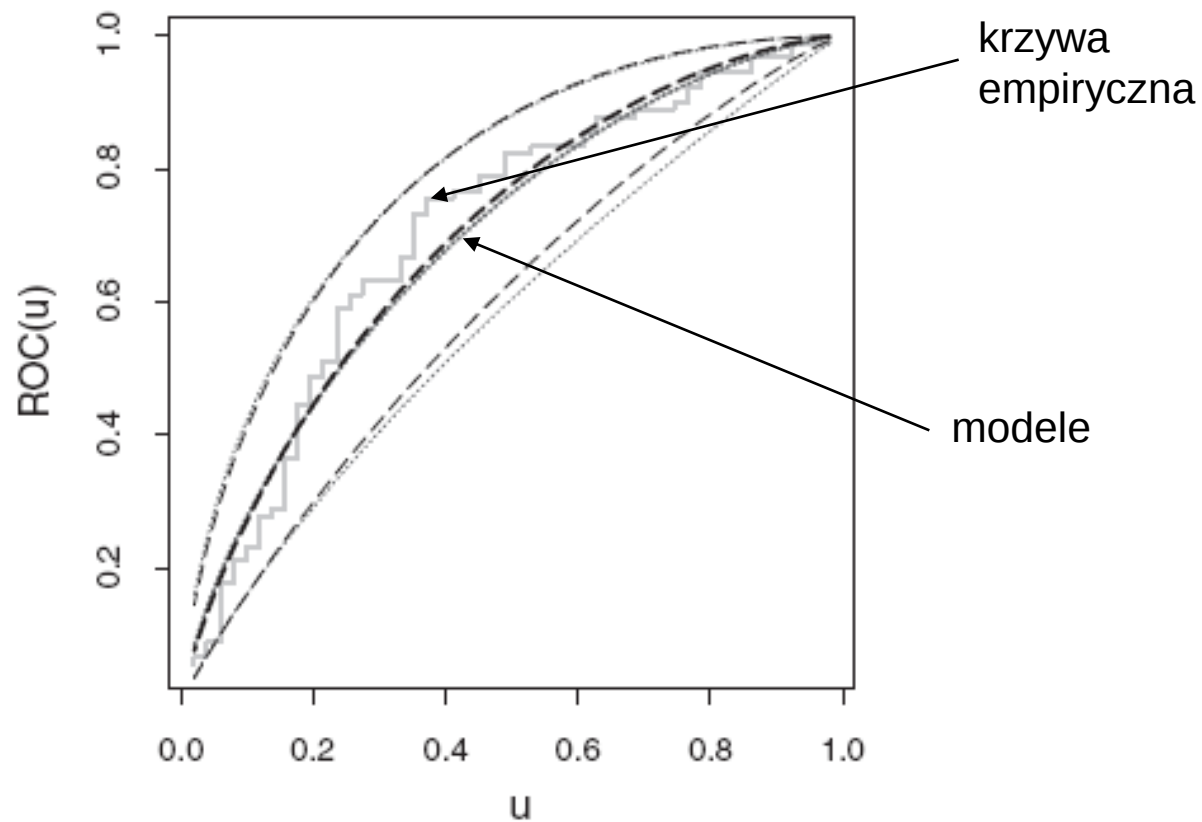


Przykładowa ocena



DBM MRMC analysis	AUC		p-value	
	I	II	CBM	TRAP
All test cases (95)	.630	.686	.0026	.0074
Cases without movement artifacts (85)	.615	.677	.0011	.0086
Cases without significant asymmetry (89)	.639	.687	.0067	.0118
Cases without scarring (59)	.653	.739	.1502	.0005

Binormalny model krzywej ROC



$$\text{ROC}(u) = \Phi \left\{ a + b\Phi^{-1}(u) \right\}$$

Ograniczenia ROC

- konieczność zamiany normalnego trybu diagnozowania w praktyce klinicznej na wyrażenie opinii w pewnej skali ocen.
 - 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).
 - wiele praktycznych zadań diagnostycznych, jakie stoją przed specjalistami, nie sprowadza się do decyzji dwupoziomowej: tak lub nie; w niektórych patologiach występuje kilka nieprawidłowości w różnych fragmentach obrazu, a proces decyzyjny jest dużo bardziej złożony.
-

Modyfikacje ROC

- 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 rozszerzona na kilka, czy nawet kilkanaście poziomów.
- Analiza tak otrzymanych wyników przeprowadzana jest przy pomocy dwu parametrów: czułości i przewidywanej wartości pozytywnej PVP (*predictive value positive*) -

$$\text{positive}) - \frac{N_{pp}}{N_{pp} + N_{fp}}.$$

- Czułość i PVP wykreśla się osobno – łącznie dla wszystkich testowanych grup badań
 - Porównanie czułości i PVP dla różnych stopni kompresji przeprowadzono za pomocą testu t-Studenta z wykorzystaniem rozkładu permutacji dwuelementowych (dla danych, które nie mają charakteru gaussowskiego)
-

Przykłady: czułość

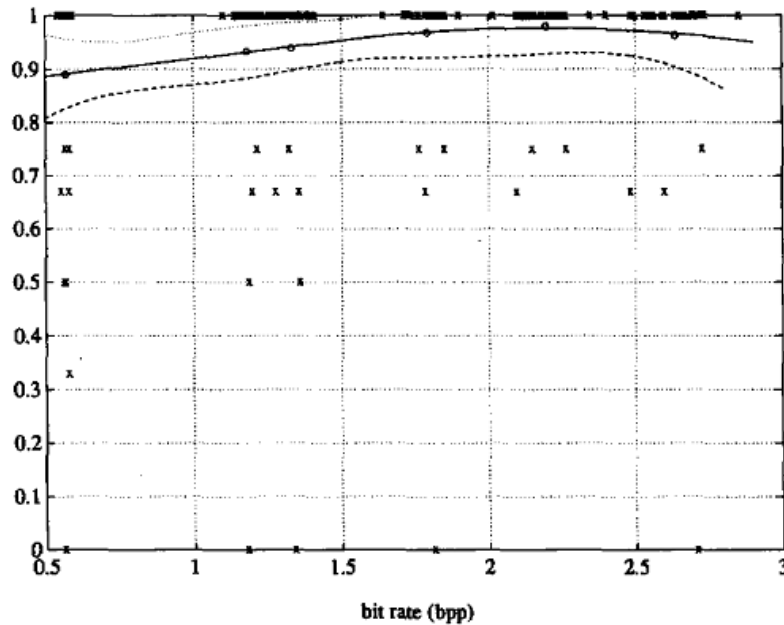


Fig. 8. Lung sensitivity: rms = 0.177.

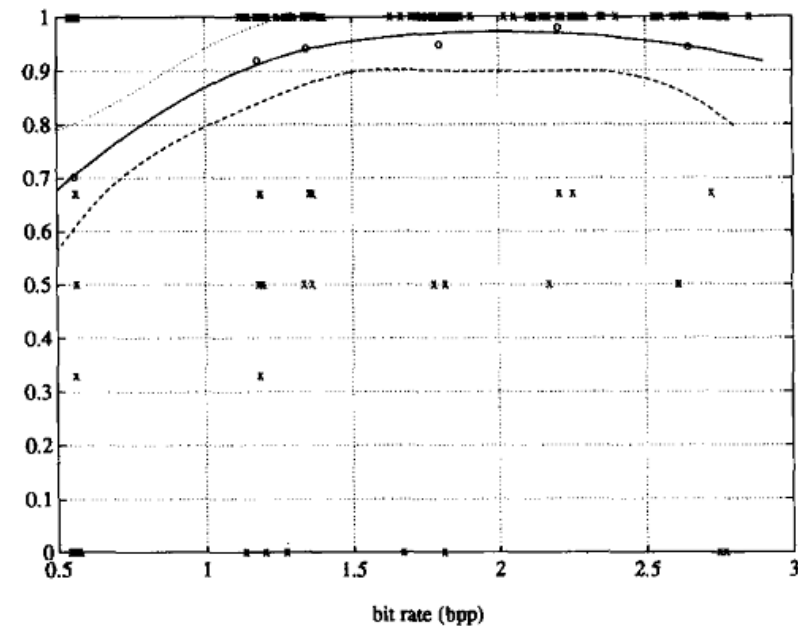


Fig. 10. Mediastinum sensitivity: rms = 0.243.

P. Cosman, R. Gray, R. Olshen Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy. PROCEEDINGS OF THE IEEE, VOL. 82, NO. 6, JUNE 1994

Przykłady: PVP

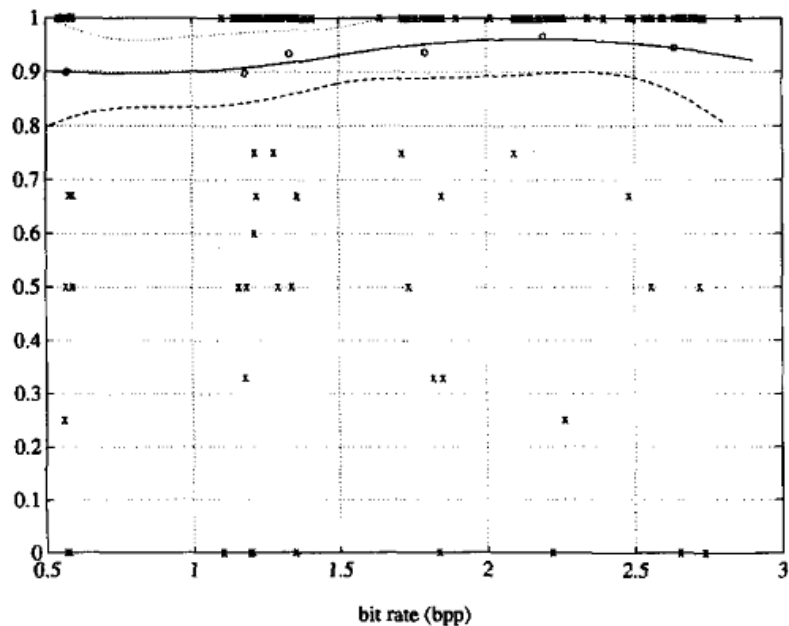


Fig. 9. Lung PVP: rms = 0.215.

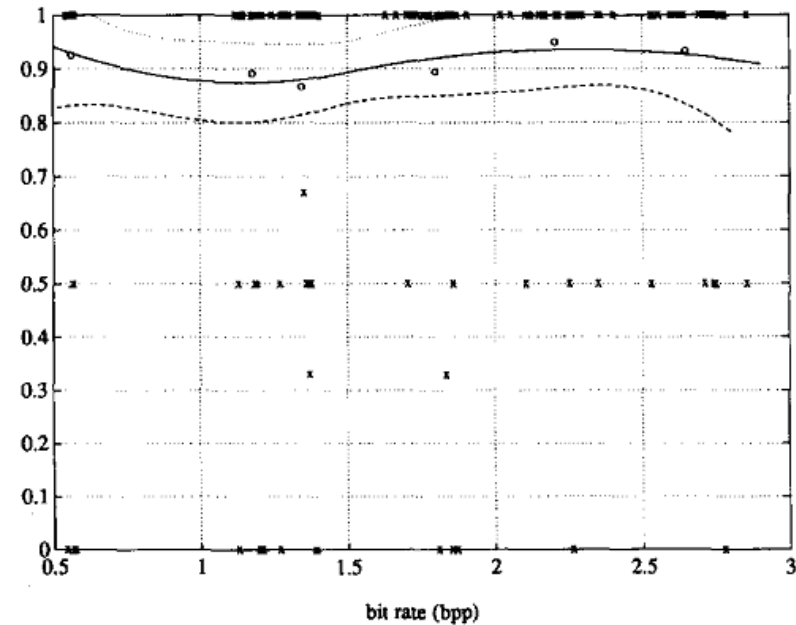
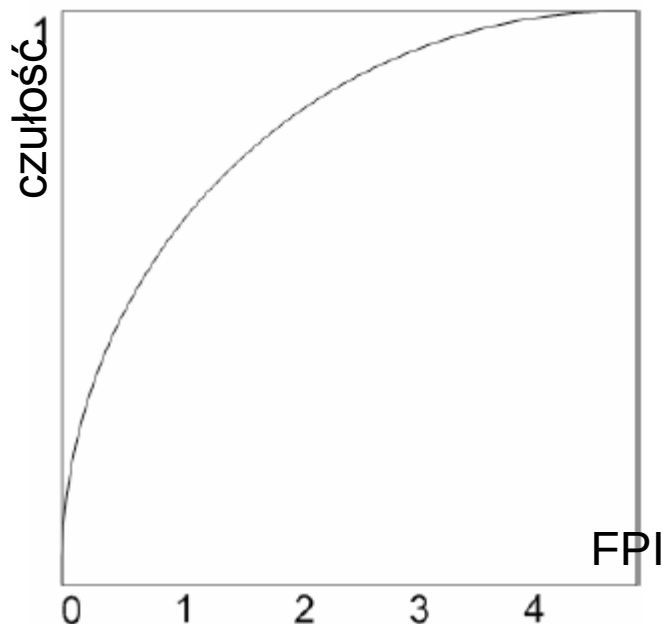


Fig. 11. Mediastinum PVP: rms = 0.245.

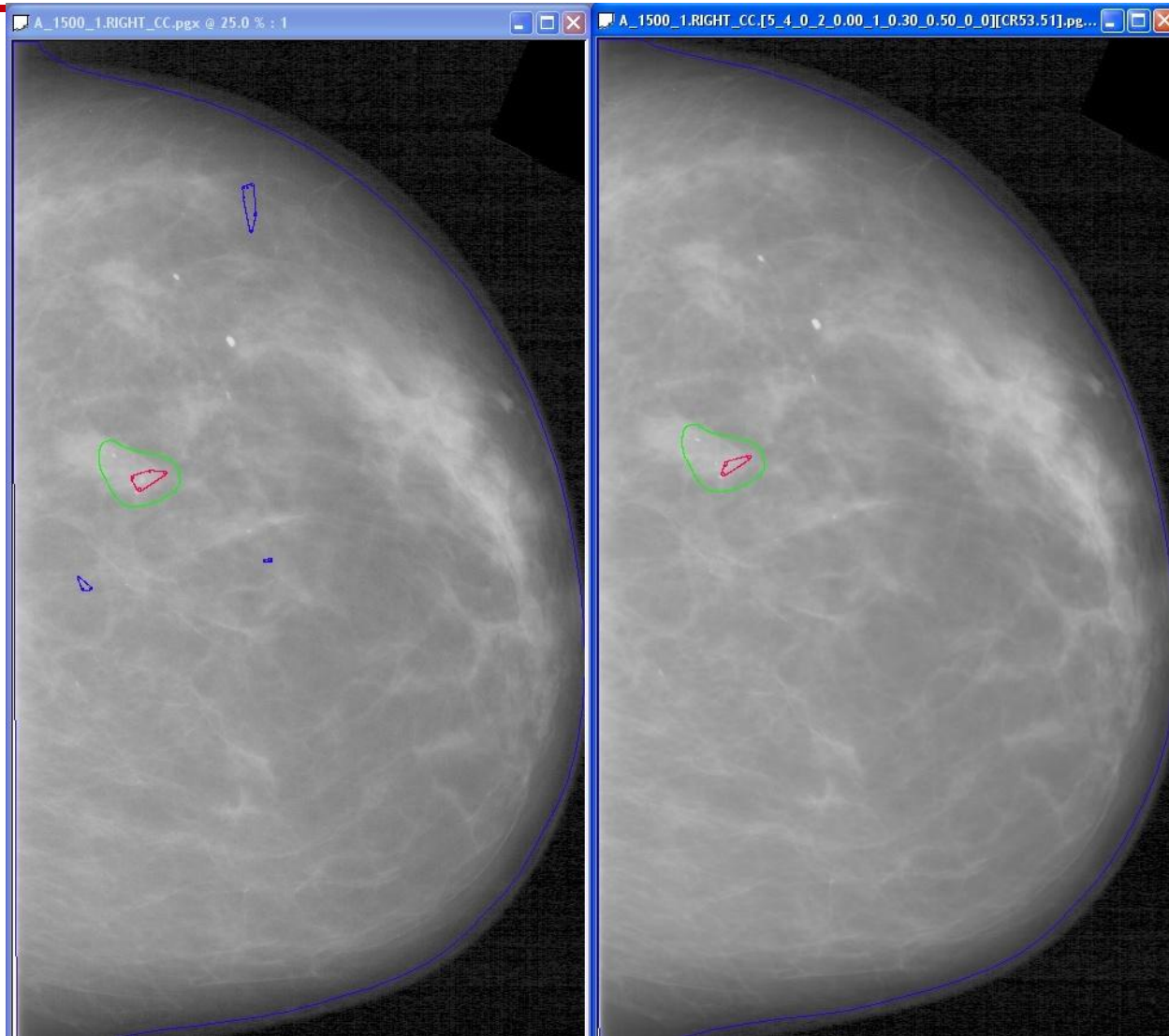
P. Cosman, R. Gray, R. Olshen Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy. PROCEEDINGS OF THE IEEE, VOL. 82, NO. 6, JUNE 1994

Ocena efektywności detekcji

- Czułość = TP (z poprawną lokalizacją)/liczba obiektów (patologii, zmian)
- FPI (*false positives per image*) = FP /liczba obrazów
- Krzywa FROC (*free-response receiver operating characteristic*)



Ocena wskazań detekcji

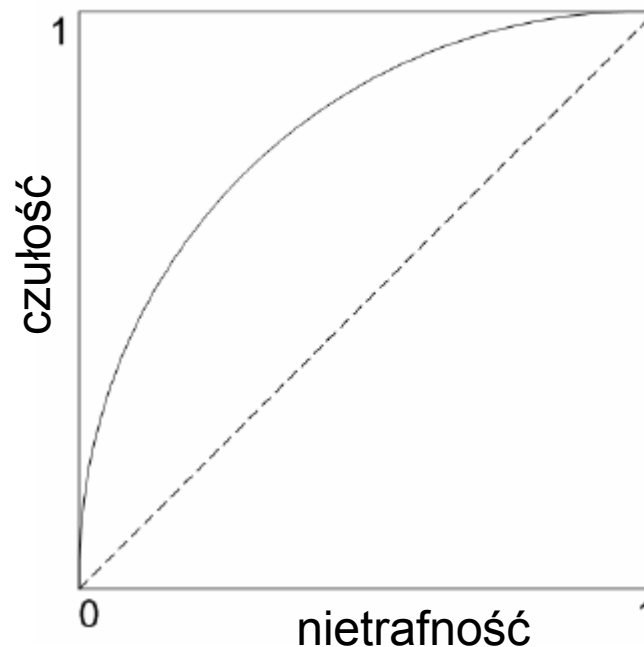


oryginał

efekt przetwarzania

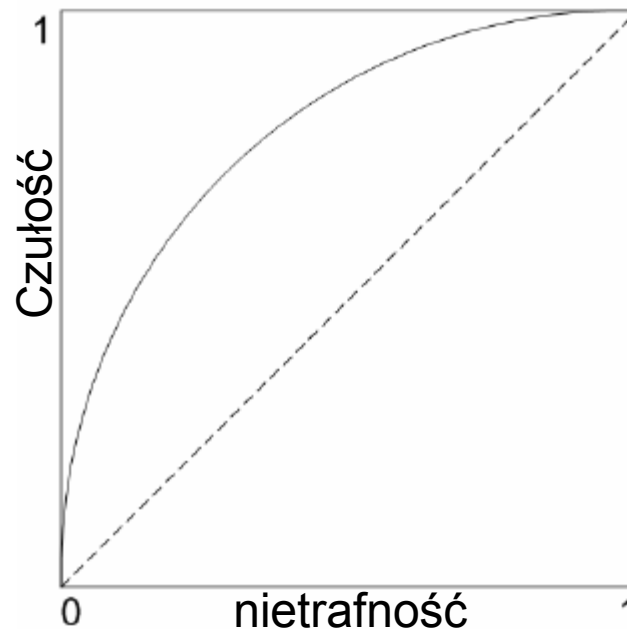
Ocena efektywności diagnozy (ROC)

- Czułość = $TP / \text{liczba zmian złośliwych}$
- Trafność (*false positives per image*) = $FP / \text{liczba zmian łagodnych}$
- Krzywa ROC (*receiver operating characteristic*)



Ocena efektywności diagnozy (LROC)

- Czułość = TP (z poprawną lokalizacją)/liczba zmian złośliwych
- Trafność (*false positives per image*) = FP/liczba zmian łagodnych
- Krzywa LROC (*localization receiver operating characteristic*)



Metod bez ROC

- 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.

Przykładowe oceny

II \ I	R	W
R	7	2
W	1	2

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	6	4
W	3	5

C/B

II \ I	R	W
R	14	2
W	2	8

BX

(A) Analog vs. Digital Original

II \ I	R	W
R	6	3
W	1	2

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	8	2
W	2	6

C/B

II \ I	R	W
R	14	2
W	1	9

BX

(B) Analog vs. Digital Lossy Compressed: 1.75 bpp

II \ I	R	W
R	6	3
W	0	3

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	6	4
W	2	6

C/B

II \ I	R	W
R	12	3
W	4	6

BX

(C) Analog vs. Digital Lossy Compressed: 0.4 bpp

II \ I	R	W
R	4	4
W	0	3

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	3	7
W	4	4

C/B

II \ I	R	W
R	11	4
W	4	6

BX

(D) Analog vs. Digital Lossy Compressed: 0.15 bpp

Table 9: Agreement 2×2 tables for radiologist A.

Analiza statystyczna

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

porównanie $N(1,2)$ i $(N(1,2)+N(2,1))/2$ z pomocą rozkładu dwumianowego o parametrach $N(1,2)+N(2,1)$ i 0.5

lub 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.

$$P(N(1,2) = k \mid N(1,2) + N(2,1) = n) = \binom{n}{k} 2^{-n}; \quad k = 0, \dots, n.$$

Dodatkowo tablice zgodności decyzji

	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.

Przykładowe oceny

	RTS	F/U	C/B	BX
RTS	11	0	5	1
F/U	0	0	0	0
C/B	3	0	11	7
BX	2	0	2	15

	RTS	F/U	C/B	BX
RTS	4	0	0	0
F/U	0	0	0	1
C/B	3	0	3	3
BX	1	0	7	35

	RTS	F/U	C/B	BX
RTS	8	0	6	1
F/U	0	0	0	0
C/B	1	0	10	1
BX	0	0	7	23

A: Analog versus Digital

	RTS	F/U	C/B	BX
RTS	11	0	6	0
F/U	0	0	0	0
C/B	2	0	15	4
BX	1	0	2	16

	RTS	F/U	C/B	BX
RTS	2	1	0	1
F/U	0	1	0	0
C/B	3	1	3	2
BX	1	0	4	37

	RTS	F/U	C/B	BX
RTS	11	0	4	0
F/U	0	0	0	0
C/B	1	1	8	2
BX	1	0	5	24

B: Analog versus Lossy Compressed Digital: 1.75 bpp

	RTS	F/U	C/B	BX
RTS	9	0	6	2
F/U	0	0	0	0
C/B	1	0	10	10
BX	1	0	2	15

	RTS	F/U	C/B	BX
RTS	1	0	2	1
F/U	0	0	0	1
C/B	2	0	2	5
BX	2	0	5	36

	RTS	F/U	C/B	BX
RTS	7	0	7	1
F/U	0	0	0	0
C/B	2	0	8	2
BX	1	0	4	25

C: Analog versus Lossy Compressed Digital: 0.4 bpp

	RTS	F/U	C/B	BX
RTS	8	0	7	1
F/U	0	0	0	0
C/B	3	1	9	8
BX	1	0	6	11

	RTS	F/U	C/B	BX
RTS	3	1	0	0
F/U	0	1	0	0
C/B	3	0	3	2
BX	1	1	5	35

	RTS	F/U	C/B	BX
RTS	7	0	7	0
F/U	0	0	0	0
C/B	0	0	9	3
BX	0	0	9	20

D: Analog versus Lossy Compressed Digital: 0.15 bpp

Radiologist A

Radiologist B

Radiologist C

Table 10: Radiologist agreement tables, Stanford judges.

Betts B.J., Li J., Cosman P.C., Gray R.M. et al.: Image quality in digital mammography. Revision of final report to the Army Medical Research and Material Command, Compression and Classification of Digital Mammograms for Storage, Transmission, and Computer Aided Screening, September 1998.