

# Automatyczna ocena wielomodalnych zniekształceń w obrazach cyfrowych

Andrzej Głowacz, Michał Grega, Lucjan Janowski, Mikołaj Leszczuk, Piotr Romaniak  
(e-mail: {Głowacz, Grega, Janowski, Leszczuk, Romaniak}@kt.agh.edu.pl)  
Katedra Telekomunikacji Akademii Górniczo-Hutniczej – Kraków

## STRESZCZENIE

Artykuł przedstawia nową metodę jakościowej oceny obrazów zawierających zniekształcenia wielomodalne. Idea oceny polega na określeniu jakości obrazu postrzeganej przez użytkownika końcowego w sposób automatyczny, w celu uniknięcia czasochłonnej i niepowtarzalnej metody charakterystycznych dla metod eksperymentalnych. Efekt osiągnięty jest poprzez obliczanie miar poszczególnych zniekształceń cząstkowych oraz mapowanie otrzymanych wyników na miary liczbowe. Dla celów analizy opracowano siedem skutecznych miar zniekształceń cząstkowych w stosunku do obrazu oryginalnego. W dalszej części pracy zaproponowano użyteczne modele przygotowane na bazie GLZ (ang. Generalized Linear Model), służące do mapowania miar ilościowych na oceny subiektywne, w celu otrzymania ogólnej oceny jakości obrazu. Dla opracowania i weryfikacji modeli przeprowadzono testy subiektywne zbierając 12.000 indywidualnych ocen jakości dla zniekształconych obrazów.

## ABSTRACT

### **Automated Assessment of Multi-Modal Distortions in Digital Images**

This paper presents a new method of qualitative assessment of images affected by multi-modal distortions. The idea is to assess the image quality perceived by an end user in an automatic way in order to avoid the usual time-consuming, costly and non-repeatable method of collecting subjective scores during a psychophysical experiment. This is achieved by computing quantitative image distortions and mapping results on qualitative scores. Overall qualitative image distortion is computed based on partial quantitative distortions from component algorithms operating on specified image features. Seven such algorithms are applied to successfully analyze the seven image distortions in relation to the original image. Useful mapping models have been proposed and constructed using the Generalized Linear Model (GLZ). These models allow for mapping the results of quantitative assessments on the subjective scores in order to obtain an overall quality score of the image. For research and validate the models, a survey of over 12,000 subjective quality scores has been carried out.

## 1. Wprowadzenie

Operacje przetwarzania i przesyłania są obecnie powszechnie stosowane w przypadku zdjęć cyfrowych. Przykładem może być kompresja, która umożliwia zmniejszenie rozmiaru zdjęcia, lub transmisja w sieci telekomunikacyjnej w oparciu o protokoły bezpołączeniowe. Operacje te mogą spowodować zakłócenia obrazu i, w konsekwencji, nieidealną rekonstrukcję w stosunku do obrazu pierwotnego. W rezultacie mogą powstać zakłócenia mono-modalne (np. szum lub rozmycie) lub raczej multimodalne (np. kombinacja szumu oraz rozmycia). Niniejszy artykuł przedstawia jednolite podejście, pozwalające na określenie niezależnych ocen ilościowych pojedynczych typów zakłóceń i odwzorowywanie ich na jakościowe oceny reprezentujące zarówno pojedyncze zakłócenia, jak i ogólną jakość. Większość systemów oceny jakości obrazów przedstawia tylko jedną ocenę reprezentującą ogólną jakość obrazu, podczas gdy proponowany niezależny system oceny pozwala na szczegółowe określenie źródła degradacji obrazu.

Parametry jakości obrazu można sklasyfikować przy użyciu trzech ortogonalnych systemów klasyfikacji: w zależności od ilości informacji potrzebnych do określenia jakości, w zależności od metody obliczeń i w zależności od sposobu, w jaki jakość jest wyrażona. Jeżeli pod uwagę brana jest ilość informacji wymaganej w celu określenia jakości, zdefiniować można trzy scenariusze:

- 1) w pełni referencyjny (FR, *Full Reference*),
- 2) zredukowany referencyjny (RR, *Reduced Reference*),
- 3) bezreferencyjny (NR, *No Reference*).

Jeśli pod uwagę brana jest metoda obliczeń, możliwych jest wiele klas metryk sparametryzowanych skalarными parametrami, a opartych na algorytmach, począwszy od prostego porównywania danych (piksel do piksela) do zaawansowanej analizy obrazu. Proste metryki porównawcze określają wierność odtworzenia sygnału bez uwzględnienia jego treści. Przykładami

takich metryk są: PSNR (ang. *Peak Signal to Noise Ratio*), MSE (ang. *Mean Squared Error*) i podobne. Istnieje również kilka miar działających w oparciu o zaawansowane analizy obrazu. Metryki takie analizują informacje wizualne, a także obejmują szeroki zakres skalarnych parametrów systemu wizyjnego człowieka (HVS, ang. *Human Visual System*), analizując widmo cyfrowego obrazu w celu odtworzenia ludzkiego postrzegania. Przykładowo, autorzy metryki Picture Quality Scale (PQS) [1] zdefiniowali ogólną metrykę z połączenia błędów skalarnych. W rozwiązaniu tym jednak nie można uzyskać szczegółowych informacji na temat poszczególnych zakłóceń obrazu.

Parametry jakości obrazu mogą być również klasyfikowane na podstawie tego, w jaki sposób jakość jest wyrażona, a w szczególności – czy wyrażona jest w podejściu jakościowym, czy ilościowym. Kryteria ilościowe wyrażane są przez wartość liczbową. Z drugiej strony, kryteria jakościowe przedstawiane są graficznie (np. wykresy Hosaki [2]), tekstowo (np. Mean Opinion Score – MOS [3]) lub też numerycznie (np. R-wartość). Ważne jest, aby pamiętać, że choć miary ilościowe można obliczyć, to zazwyczaj jednak nie istnieją ich proste odwzorowania na skale jakości obrazujące rzeczywistą jakość obrazu odbieraną przez użytkownika.

Przykładem metryki jakości podającej ogólny wynik jakości (MOS) jest metryka Perceptual Evaluation of Video Quality (PEVQ), stworzona na podstawie Zalecenia ITU-T J-144 [4]. Metryka ta została zaprojektowana tak, by oszacować jakość sekwencji wizyjnych, przy degradacjach występujących przy okazji transmisji za pośrednictwem sieci. Metryka ta jednak może zostać uproszczona do jakości obrazu statycznego, ponieważ działa na poziomie zdekompresowanych klatek sekwencji wizyjnych.

Ideą przewodnią podejścia zaprezentowanego w niniejszym artykule jest opracowanie zestawu algorytmów dla niezależnej oceny wybranych zakłóceń obrazu. Opracowane algorytmy są niewrażliwe na inne zakłócenia wprowadzone do obrazu. Ocena zakłóceń jakości obrazu nie jest zbyt trudna, gdy bada się obrazy zniekształcone mono-modalnie (tylko jeden rodzaj zakłócenia). Zadanie staje się dużo bardziej skomplikowane, gdy obraz jest zniekształcony wielo-modalnie (np. zarówno szum i jak i rozmycie).

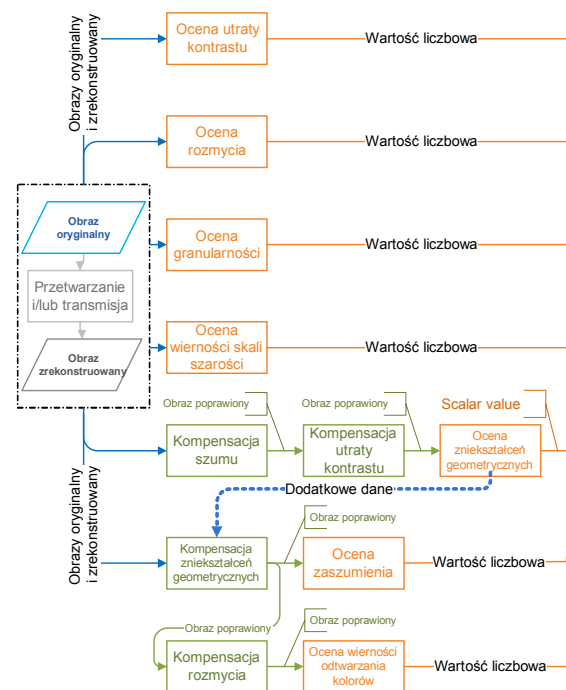
Ostatnim krokiem w prezentowanym systemie jest automatyczne odwzorowanie uzyskanych wartości ilościowych na jakościowe odpowiedzi symulowanego użytkownika. Odwzorowanie to oparte jest na doświadczeniach psychofizycznych (testach subiektywnych) wykonanych wcześniej.

W artykule zaprezentowane zostały dwa główne wkłady do stanu badań. Są to: (1) spójny zestaw odpornych na zakłócenia wzajemne algorytmów dla niezależnej oceny ilościowej wybranych zakłóceń obrazu oraz (2) odwzorowanie miar ilościowych na oceny jakościowe, które pozwala na wyeliminowanie trudnych do zorganizowania, niedokładnych i zużywających zasoby, testów subiektywnych.

Artykuł zredagowano w następujący sposób: rozdział 2 opisuje szczegóły metodologii konstrukcji metryk i kompensacji zakłóceń. Rozdział 3 przedstawia subiektywne oceny jakości, a rozdział 4 – odwzorowywanie odpowiedzi użytkowników. W rozdziale 5 wyniki są weryfikowane, natomiast rozdział 6 przedstawia implementację i podsumowuje artykuł.

## 2. Miary liczbowe zniekształceń oraz sposoby ich kompensacji

Sekcja ta zawiera dokładny opis siedmiu miar służących do oceny jakości obrazów oraz opis metod kompensacji, pozwalających na wyeliminowanie wpływu pewnych zniekształceń na ocenę pozostałych.



Rys. 1. Metodologia pomiaru parametrów jakości

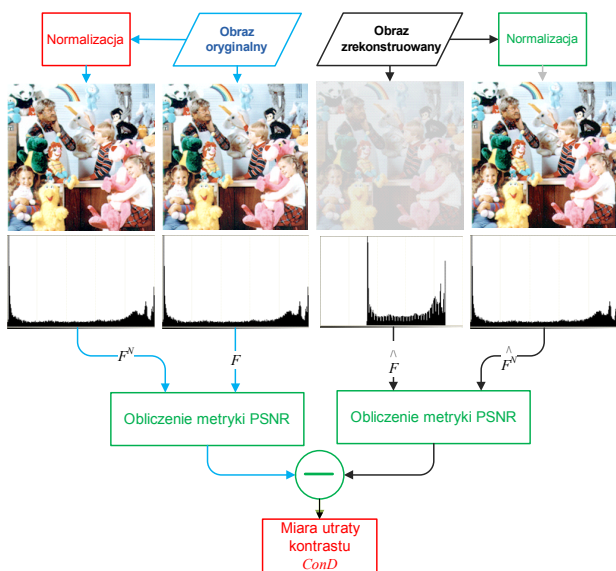
### 2.1. Miary ilościowe

Autorzy stworzyli miary do oceny poziomu następujących zniekształceń: utrata kontrastu, rozmycie, granularność, zniekształcenia geometryczne, zaszumienie, zniekształcenie kolorów oraz zmiana współczynnika gamma. Zestaw rozważanych parametrów jakości ob-

razu wyspecyfikowała, jako Zamawiający nieniejsze badania, Telekomunikacja Polska S.A. Wyboru dokonano z uwzględnieniem ludzkiego systemu postrzegania oraz parametrów mierzonych w istniejących miarach (PQS, PEVQ). Rysunek 1 przedstawia metodologię zastosowaną do wyznaczania poziomu poszczególnych zniekształceń oraz zastosowane metody kompensacji, pozwalające na wyeliminowanie niekorzystnego wpływu jednego zniekształcenia na ocenę pozostałych. Aspekt ten szczegółowo został opisany w dalszej części artykułu.

### 2.1.1. Utrata kontrastu

W celu oszacowania utraty kontrastu ConD wykorzystana została metoda zilustrowana na rysunku 2. W pierwszym kroku histogramy obrazu oryginalnego  $F$  i zniekształconego  $F'$  są normalizowane. Następnie dwie pary obrazów (obraz i jego znormalizowany odpowiednik) porównywane są z wykorzystaniem metryki PSNR. Metryka PSNR zwraca stopień podobieństwa w decybelach. Wynik odejmowania dwóch wartości podanych w decybelach (odejmowanie w dziedzinie decybeli jest równoważne dzieleniu w skali liniowej) określa poziom utraty kontrastu. Wyniki badań dowiodły, że zaproponowany pomiar utraty kontrastu jest nieczuły na występowanie pozostałych zniekształceń.

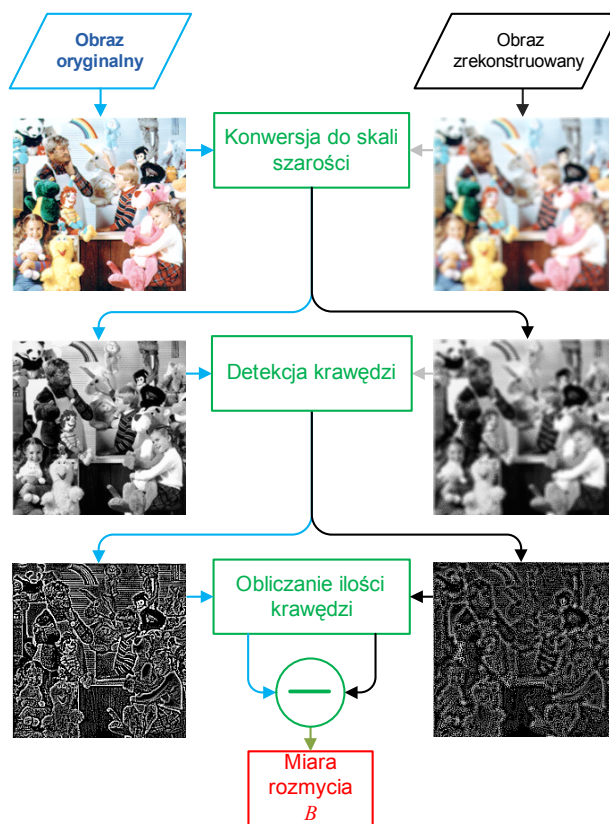


Rys. 2. Schemat oceny utraty kontrastu

### 2.1.2. Ocena rozmycia

Rozmycie obrazu jest jednym z najważniejszych zniekształceń, jakie wpływają na jakość odbieranych obrazów. Związane jest to ściśle z ilością detali, jakie prezentuje obraz i, definiowany jako najkrótszy, dystans pomiędzy obszarami o przeciwnych kolorach, np. biały i czarny. Ze względu na powyższe, algorytm do

wykrywania krawędzi obrazu wydaje się być odpowiednim narzędziem do oceny rozmycia. Im więcej krawędzi wykrytych na obrazie, tym większa jego ostrość.



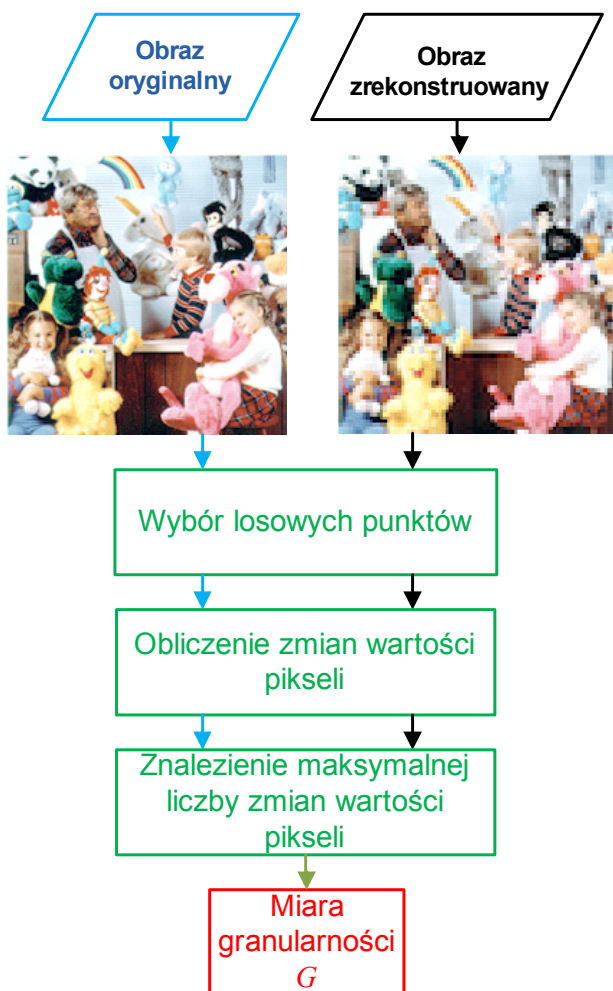
Rys. 3. Schemat oceny rozmycia

Pierwszym krokiem w ocenie rozmycia jest konwersja obrazu oryginalnego i zniekształconego do skali szarości (rys. 3). Następnie dokonywana jest detekcja krawędzi z wykorzystaniem filtru metody CED (ang. *Canny Edge Detection*) [5]. Kolejnym krokiem jest obliczenie liczby krawędzi wykrytych na obydwu obrazach. Różnica obliczonych wartości stanowi miarę rozmycia. Wyniki badań dowiodły, że zaproponowany pomiar rozmycia jest nieczuły na występowanie pozostałych zniekształceń.

### 2.1.3. Granularność

Obliczanie granularności obrazu (nazywanej także rozdzielczością efektywną) dokonywane jest w kilku krokach, tak jak pokazano na rysunku 4. Na początku wybierana jest pewna liczba losowych punktów na obrazie (w badaniach zdecydowano się na 100 punktów). Zaczynając od wybranych punktów, liczba zmian wartości pikseli (zmiana jednego ze współczynników RGB) obliczana jest dla pionowych i poziomych linii obrazu, każda o długości 100 pikseli. Obliczenia te wykonywane są dla obrazu oryginalnego i zniekształconego, rozpoczynając od tych samych punktów startowych. Maksymalna uzyskana

liczba zmian wartości pikseli dla danego obrazu wyznacza granularność. Stosunek tych wartości dla obrazu oryginalnego i zniekształconego określa stopień zwiększenia granularności na obrazie zniekształconym. Wyniki badań dowiodły, że zaproponowany pomiar granularności jest nieczuły na występowanie pozostałych zniekształceń.

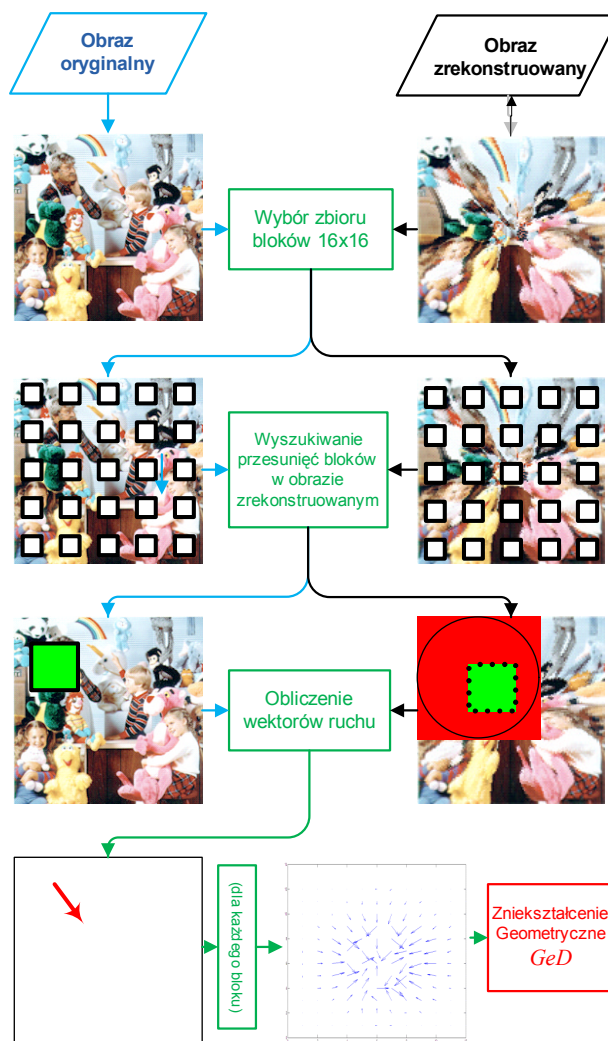


Rys. 4. Schemat oceny granularności

#### 2.1.4. Zniekształcenia geometryczne

Zniekształcenia geometryczne wprowadzone mogą zostać do obrazu w czasie przetwarzania analogowego, takiego jak na przykład skanowanie dokumentu. Prezentowana metoda analizy zniekształceń geometrycznych oparta jest na badaniu wektorów ruchu.

Pomimo że badanie zniekształceń przeprowadzane jest dla obrazów nieruchomych, w tej metodzie obraz oryginalny i zniekształcony traktowane są jako dwie kolejne klatki sekwencji wideo. Dla takiej sekwencji, składającej się z dwóch klatek, odbywa się detekcja ruchu wraz z obliczeniem wektorów ruchu. Tak więc zniekształcenie geometryczne traktowane jest jako forma ruchu pomiędzy dwoma kolejnymi klatkami wideo.



Rys. 5. Schemat oceny zniekształceń geometrycznych

Opisany poniżej i składający się z kilku prostych kroków algorytm detekcji ruchu jest bardzo podobny do algorytmu estymacji ruchu stosowanego w wielu rozwiązaniach problemu kompresji wideo z rodziny standardów MPEG. W pierwszym kroku wybierany jest zestaw  $n$  równomiernie rozłożonych kwadratowych bloków w obrazie oryginalnym. W drugim kroku dla każdego bloku w obrazie oryginalnym wyszukiwany jest najbardziej podobny blok w obrazie zniekształconym. Wyszukiwanie najbardziej podobnego bloku odbywa się na zasadzie porównania bloku oryginalnego ze wszystkimi możliwymi blokami na obrazie zniekształconym, leżącymi w promieniu nie większym niż zadany promień  $r$ . Porównanie to odbywa się przy użyciu miary błędu średniokwadratowego (ang. *Mean Square Error*, *MSE*). W kolejnym kroku dla każdego bloku z obrazu oryginalnego oraz najbliższego mu bloku z obrazu zniekształconego obliczany jest wektor ruchu  $V$ . W ostatnim kroku do mia-

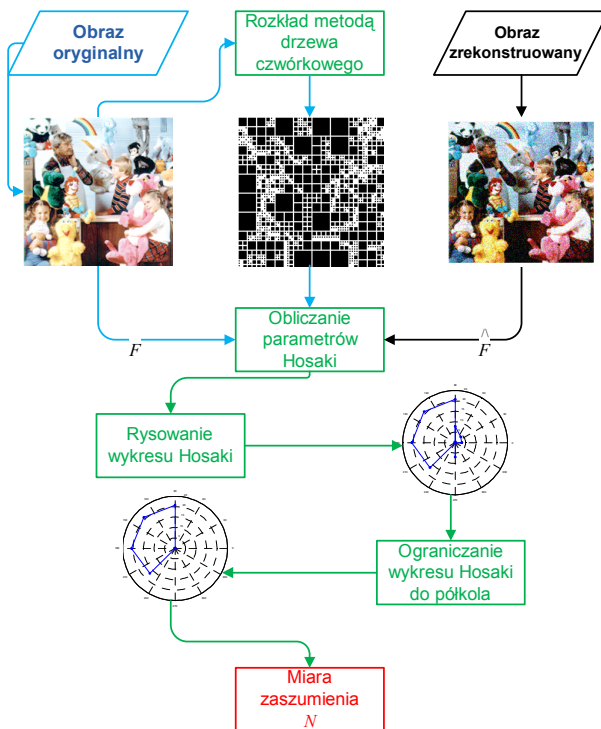


ry zniekształceń geometrycznych brana jest pod uwagę całkowita długość wszystkich wektorów ruchu  $GeD$ .

$$GeD = \frac{\sum_{i=1}^n |\vec{v}_i|}{n \cdot r}$$

### 2.1.5. Zaszumienie

Ocena zaszumienia oparta jest na idei wykresów Hosaki [2]. Algorytm rozpoczyna się od rozkładu obrazu metodą drzewa czwórkowego, a następnie zaszumienie ocenia się w kwadratowych blokach pikseli podzielonych na kilka klas  $C_i$ ,  $i = 0, \dots, n$  (gdzie zazwyczaj  $n = 4$ ), co oznacza bloki od  $1 \times 1$ , poprzez  $2 \times 2$ , do (zwykle)  $16 \times 16$ . Jednakowa liczba parametrów Hosaki reprezentuje zarówno zaszumienie, jak i nieścistości rekonstrukcji. Parametry te to:  $DS(C_i)$  i  $DM(C_i)$ , gdzie  $(DS(C_0) \equiv 0)$ . Wykresy Hosaki rysowane są na wykresie biegunowym, gdzie jedno półkole związane jest ze współczynnikami zaszumienia, a drugie ze współczynnikami nieścistości rekonstrukcji. Kształt wykresu Hosaki określa, czy szum został wprowadzony dla szczegółów (reprezentowanych przez małe bloki), czy do większych, jednorodnych obszarów (duże bloki).



Rys. 6. Schemat oceny zaszumienia obrazu

Wartość zakłócenia  $N$  została określona jako proporcjonalna do powierzchni wykresu Hosaki odpowiedzialnego za zaszumienie, a więc jako suma powierzchni trójkątów  $O(0,0)S_i(\theta_i, DS(C_i))S_{i+1}(\theta_{i+1}, DS(C_{i+1}))$ , gdzie współrzędne punktów podane są w biegunowym ukła-

dzie współrzędnych, a  $\theta_i$  jest kątem, dla którego obliczona została wartość  $DS(C_i)$

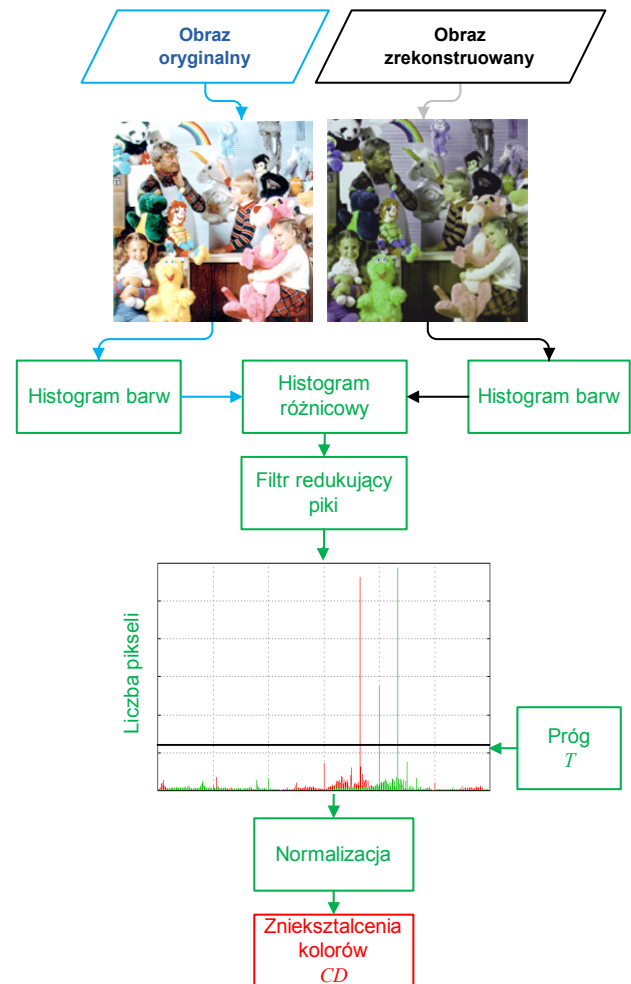
$$N = \sum_{i=0}^{n-1} \frac{1}{2} \sin|\theta_{i+1} - \theta_i| DS(C_i) DS(C_{i+1})$$

Biorąc pod uwagę, że  $|\theta_{i+1} - \theta_i| \equiv \pi / (n - 1)$ , jak również fakt, że wartość zakłócenia nie jest znormalizowana, wszystkie stałe wartości można wykluczyć, z czego wynika uproszczona notacja  $N$  jako

$$N = \frac{1}{2} \sin \frac{\pi}{n-1} \sum_{i=0}^{n-1} DS(C_i) DS(C_{i+1}) \alpha \sum_{i=0}^{n-1} DS(C_i) DS(C_{i+1})$$

Na rysunku 6 przedstawiono graficzny opis algorytmu oceny poziomego zaszumienia.

### 2.1.6. Wierność odtwarzania kolorów

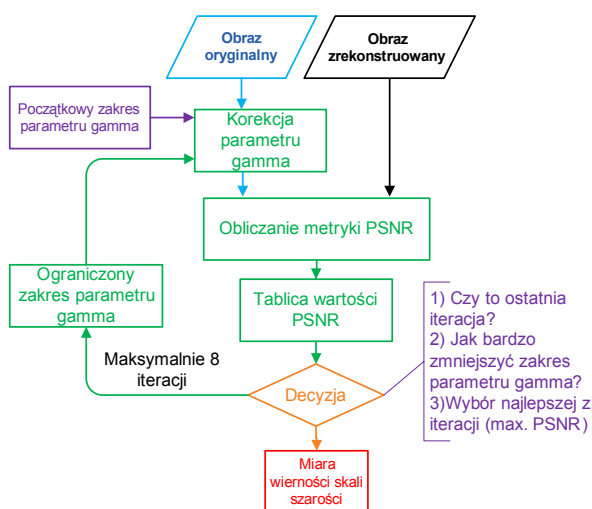


Rys. 7. Schemat oceny wierności odtwarzania kolorów

Zniekształcenia odwzorowania kolorów postrzegane są poprzez zniekształcenia składowej barwy (ang. *Hue*) w przestrzeni kolorów HSV (ang. *Hue-Saturation-Value*). Rozważane są wszystkie pary odpowiadających sobie pikseli w obrazie oryginalnym i zniekształconym. Na podstawie różnic barw każdej pary

tworzony jest histogram różnic. Badania wykazały, że niektóre rodzaje zniekształceń dają lokalne maksima o dużej wartości w charakterystyce histogramu (rys. 7). Przykładem mogą być zniekształcenia kontrastu lub zniekształcenia geometryczne. Dla zmniejszenia ewentualnej wariancji miary oceny jakości, wskazane jest złagodzenie charakterystyki histogramu poprzez eliminację maksimów o dużych wartościach. Dlatego też wprowadzono próg o wartości  $T=1/360 \cdot M \cdot N$ . Do histogramu różnicowego wliczane są różnice mniejsze niż  $T$ . Miara oceny zniekształceń kolorów jest sumą wszystkich wartości tak otrzymanego histogramu różnicowego, podzieloną przez łączną liczbę pikseli w obrazie. Miara przyjmuje wartości ciągłe z zakresu  $0..1$  i rośnie wraz z poziomem zniekształceń kolorów.

### 2.1.7. Wierność odtwarzania skali szarości



Rys. 8. Schemat oceny wierności odtwarzania skali szarości

Zniekształcenia odwzorowań skali szarości powstają wskutek zmian poziomu współczynnika gamma obrazu. Ten typ zniekształcenia mierzony jest poprzez stopniową degradację obrazu oryginalnego oraz porównywanie z obrazem zniekształconym. Na początku określamy granice parametru gamma, tzn. najwyższą i najniższą wartość, poza którymi zmiany parametru nie wpływają na postrzeganie stopni szarości. Przedział ten został określony na podstawie wyników uzyskanych w testach subiektywnych. Następnie, w określonej liczbie kroków, obraz oryginalny jest pogarszany i obliczana jest miara PSNR w stosunku do obrazu zniekształconego (rys. 8). Przedział parametru gamma jest w kolejnych krokach zawężany. Najwyższa wartość miary PSNR wskazuje największe podobieństwo. Określenie indeksu pośrednich miar PSNR pozwala stwierdzić, w którym kroku metoda daje najlepsze

przybliżenie parametru gamma. Miara końcowa leży w zakresie wartości od 1 do 2 i rośnie wraz z poziomem zniekształceń szarości.

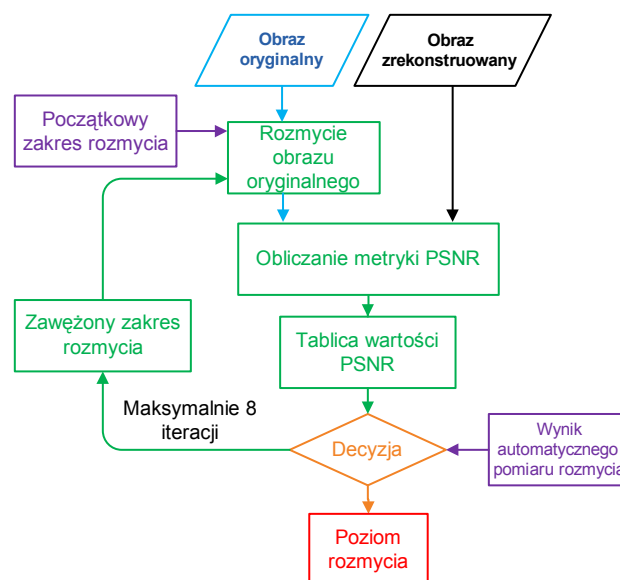
## 2.2. Kompensacja zniekształceń obrazu

Ocena pojedynczego zniekształcenia obrazu nie stanowi dużego wyzwania w przypadku, gdy dane zniekształcenie występuje samodzielnie. Trudność zadania znacznie wzrasta, gdy jednocześnie pojawiają się dodatkowe zniekształcenia. Wyniki badań pokazały, że tylko ocena spadku kontrastu, rozmycia i granularności są nieczułe na występowanie pozostałych zniekształceń. Oznacza to, że ocena pozostałych parametrów jakości obrazu może zostać przekłamana z powodu występowania innych zniekształceń. Aby rozwiązać ten problem, zaproponowano szereg kompensacji, pozwalających na wyeliminowanie niepożądanego wpływu. W przypadku zniekształceń odwracalnych kompensacja polega na usunięciu danego zniekształcenia z obrazu zniekształconego, natomiast w przypadku zniekształceń nieodwracalnych kompensacja polega na wprowadzeniu takiego samego zniekształcenia do obrazu oryginalnego.

### 2.2.1. Kompensacja utraty kontrastu

Kompensacja utraty kontrastu pozwala na dokładniejsze oszacowanie pozostałych parametrów. Jest to zniekształcenie odwracalne, a kompensacja polega na normalizacji histogramów obrazu oryginalnego i zniekształconego.

### 2.2.2. Kompensacja rozmycia



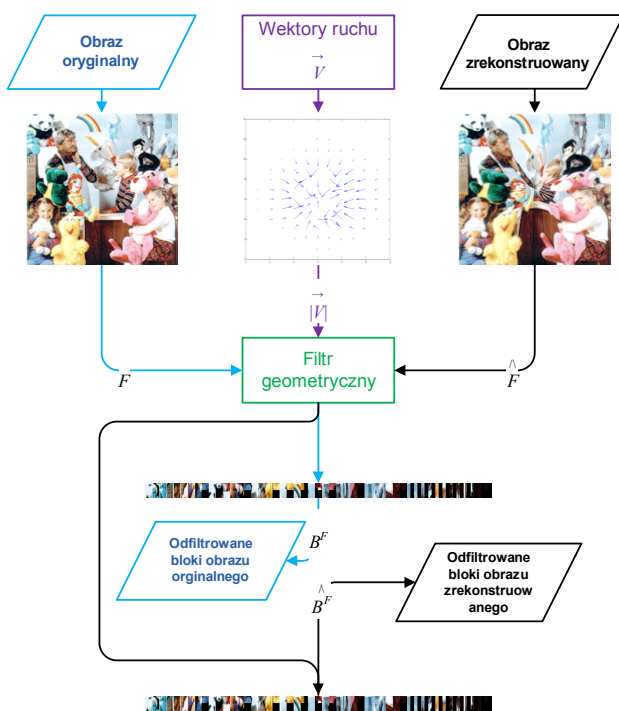
Rys. 9. Schemat kompensacji rozmycia

Rozmycie jest zniekształceniem nieodwracalnym (powoduje utratę detali obrazu), dlatego kompensacja polega na wprowadzeniu rozmycia o takim samym

poziomie do obrazu oryginalnego. W celu oszacowania poziomu rozmycia, jaki należy wprowadzić do obrazu oryginalnego, wykorzystano metodę numeryczną z ustaloną liczbą wykonań, tak jak pokazano na rysunku 9. W badaniach zdecydowano się na 8 wykonań, z czego każde kolejne pozwala na oszacowanie poziomu rozmycia z większą dokładnością. Przy takich założeniach metoda pozwala oszacować poziom rozmycia z dokładnością niemniejszą niż 2% wartości maksymalnej.

### 2.2.3. Filtrowanie zniekształceń geometrycznych

Po tym, kiedy zniekształcenia geometryczne zostaną ilościowo ocenione, są one, według zaproponowanego algorytmu, odfiltrowywane – w ten sposób obszary dotknięte takimi zniekształceniami nie są dalej przetwarzane i oceniane. Filtrowanie oparte jest na wynikach oceny zniekształceń geometrycznych. Pasujące bloki z oryginalnego i zrekonstruowanego obrazu przekazywane są do dalszych ocen.



Rys. 10. Schemat filtrowania zniekształceń geometrycznych

Kryterium kwalifikowania danej pary bloków ( $b$  oraz  $\hat{b}$ ) w przypadku dalszego przetwarzania (zestawy  $B^F$  i  $\hat{B}^F$ ), opiera się bezpośrednio na  $|\vec{V}|$  (długość wektora ruchu bloków). Tylko pary bloków mające  $|\vec{V}| = 0$  są kwalifikowane.

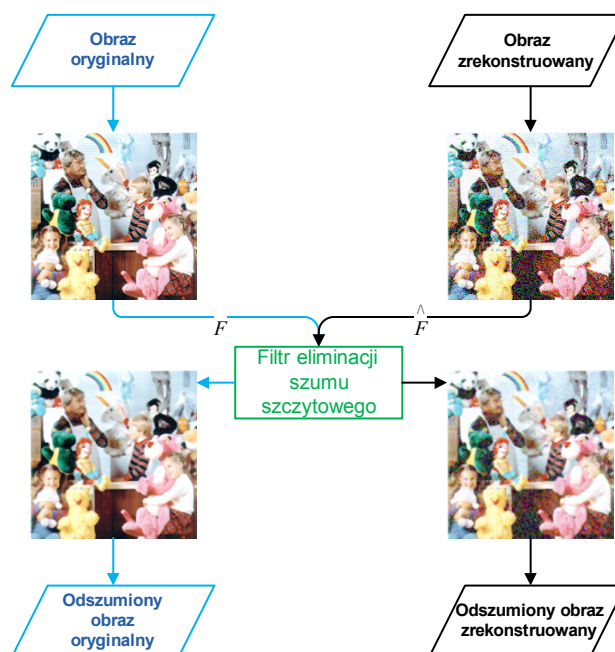
$$B^F = \{b \in B : |\vec{V}|(b, \hat{b}) = 0\}$$

$$\hat{B}^F = \{\hat{b} \in \hat{B} : |\vec{V}|(b, \hat{b}) = 0\}$$

Dwa nowe obrazy zapisywane są jako kompozycja oryginalnych i zrekonstruowanych bloków, które przeszły przez filtr. Bloki są dostosowane w jednym wierszu w każdym z obrazów jako poziomy pasek o wysokości równej wysokości pojedynczego bloku i szerokości równej szerokości wszystkich zakwalifikowanych bloków (rys. 10). Nowe obrazy są przekazywane do dalszego przetwarzania.

### 2.2.4. Kompensacja zaszumienia

Dla różnych metryk zakłóceń musiały być zastosowane procedury kompensacji zaszumienia. W większości przypadków stosowany był filtr eliminacji szumu szczytowego [7]. Biorąc pod uwagę, że każda kompensacja zaszumienia obrazu wprowadza zmiany także w obrazach z usuniętym szumem, kompensacja była stosowana zarówno do pierwotnego, jak i do zrekonstruowanego obrazu (rys. 11).



Rys. 11. Schemat filtrowania szumu

Podstawową funkcją filtra eliminującego zaszumienie jest wygładzenie obiektów obrazu bez utraty informacji o krawędziach i bez tworzenia zbędnych struktur obrazu. Zasadniczym założeniem jest zastąpienie każdego piksela, oznaczonego jako zaszumiony, wartościami z sąsiednich pikseli. Piksel jest kwalifikowany jako zaszumiony tylko wtedy, gdy ma maksymalną lub minimalną wartość w oknie piksela –  $\omega(j, k)$ .

$$\omega(j, k) = \begin{bmatrix} F(j-r, k-r) & \dots & F(j+r, k-r) \\ \vdots & \ddots & \vdots \\ F(j-r, k+r) & \dots & F(j+r, k+r) \end{bmatrix}$$

Możliwym jest określenie promienia odległości sąsiednich pikseli, czyli szerokości sąsiedztwa ( $r$ ). Filtr-

---

wanie zaszumienia odbywa się przez zastosowanie filtrów cyfrowych do zaszumionego obrazu zrekonstruowanego. W większości przypadków wybór  $r = \{1,2\}$  dawał najlepsze wyniki.

### 3. Subiektywne testy jakości obrazów

Główną motywacją badań subiektywnych było zebranie wyników opinii (OS, ang. *Opinion Scores*) w zakresie jakości zrekonstruowanych obrazów, w celu określenia funkcji odwzorowania między jakością wyrażoną ilościowo (wyjście algorytmów ilościowej oceny jakości) i jakością postrzeganą przez typowego użytkownika. Subiektywne OS umożliwiają konstruowanie modeli, eliminując konieczność zaangażowania ludzi (testerów) w dalsze badania.

#### 3.1. Metodologia

Ogólne zalecenia w zakresie przeprowadzania testów subiektywnej oceny jakości przedstawiono w [8]. Zgodnie z zaleceniem, subiektywne badania jakości obrazu powinny być przeprowadzone na zróżnicowanych i licznych grupach ekspertów (testerów). Dla wszystkich odtworzonych obrazów powinien zostać zebrany szereg ocen subiektywnych (OS). W celu dokonania oceny, jak silnie zniekształcenie kilku parametrów wpływa na postrzeganie jakości, każda sesja testu powinna obejmować ocenę obrazów zniekształconych zarówno mono- jak i wielomodalnie.

Badanie wymaga metody z podwójnym bodźcem (ang. *double stimulus*) z pięcioma opisanymi poziomami klasyfikacji i bezwzględną oceną jakości obrazu. Stąd użyto DSIS [8] jako podstawowej metodologii badań subiektywnych, z jedną niewielką zmianą. Ocena jakości obrazu nie odnosi się do poziomu zakłóceń, ale do bezwzględnej jakości obrazu (tak jak opisano w metodologii DSCQS [8]). W związku z tym przedstawiona metodologia jest połączeniem obu rozwiązań opartych na podwójnym bodźcu. Zastosowane modyfikacje eliminują błąd, jaki pojawia się w wyniku przejścia między poziomem zakłóceń a jakością obrazu, która jest wymagana jako ostateczny wynik procesu oceny zakłóceń obrazów.

#### 3.2. Zbiór obrazów testowych

Testy subiektywne zostały przeprowadzone na podstawie badania materiału, przygotowanego przy użyciu narzędzia (oprogramowania) wprowadzającego zakłócenia, zaprojektowanego w ramach opisywanych badań naukowych. Narzędzie wprowadzające zakłócenia pozwalało niezależnie zadać dowolny poziom jednego z siedmiu badanych zakłóceń.

Jako podstawa do stworzenia całego zestawu testów wybrany został jeden obraz ze standardowego zestawu zdjęć cyfrowych [8] (rys. 12). Obraz przedstawia zróżnicowaną zawartość i wydaje się być reprezentatywny dla obrazów kolorowych. Zbiór obrazów testowych zawierał wiele obrazów wygenerowanych przy pomocy narzędzia wprowadzającego zakłócenia: 94 obrazy zniekształcone monomodalnie i 330 obrazów zniekształconych wielomodalnie.

#### 3.3. Testy subiektywne

Badania subiektywne składały się łącznie z ok. 200 prób. Każdy proces oceny składał się z 60 losowych obrazów (generowanych oddzielnie dla każdego procesu), wybranych z całego zbioru testowego (426 obrazów), co wyeliminowało błąd będący wynikiem stałej kolejności obrazów. W zbiorze tych 60 obrazów, w każdym teście 12 zostało zniekształconych monomodalnie, a 48 zniekształconych było wielomodalnie. Liczba zdjęć w jednym badaniu była ograniczona przez zdolności człowieka do udzielania wiarygodnych odpowiedzi w skończonym czasie (około 15 minut). W wyniku badań subiektywnych zebrano 2400 OS dla obrazów zniekształconych monomodalnie i 9600 OS dla obrazów zniekształconych wielomodalnie.

### 4. Odwzorowywanie odpowiedzi użytkowników

Znajomość funkcji obiektywnych jest niewystarczająca, ponieważ interesującym parametrem jest jakość postrzegana przez użytkowników. Należy więc wyznaczyć odpowiednią funkcję przekształcającą obiektywne wartości metryk na subiektywne wartości odpowiedzi testerów.

Najpowszechniej stosowaną techniką rzutowania pewnych wartości na inną wartość jest regresja liniowa, dlatego wydawałoby się, że można ją zastosować w tym przypadku. Niemniej jednak, podstawowym założeniem regresji liniowej jest normalność reszt, czyli możliwość przybliżenia przez rozkład normalny różnicy pomiędzy wartością modelu, a wartością obserwowaną. Ponieważ użytkownicy mogą wybierać jedynie jedną z pięciu możliwych odpowiedzi, otrzymany rozkład odpowiedzi oraz reszt nie może zostać przybliżony przez rozkład normalny. Zauważmy, że rozkład normalny jest ciągły i symetryczny (wokół wartości średniej). Przy tak małej liczbie odpowiedzi trudno mówić o przybliżeniu ciągłości, ponadto rozkład odpowiedzi jest często skośny. Co więcej, słowny opis zaproponowany przez DSIS jest zrozumiały dla ludzi, ale nie ma prostego matematycznego znaczenia. Dlatego w [8] zaproponowano przekształcanie odpowie-



dzi słownych na liczby od 1 do 5. Należy pamiętać, że mapowanie to jest jedynie umową i zmienna losowa opisująca odpowiedzi testerów to zmienna typu porządkowego, a nie liczbowego [10].

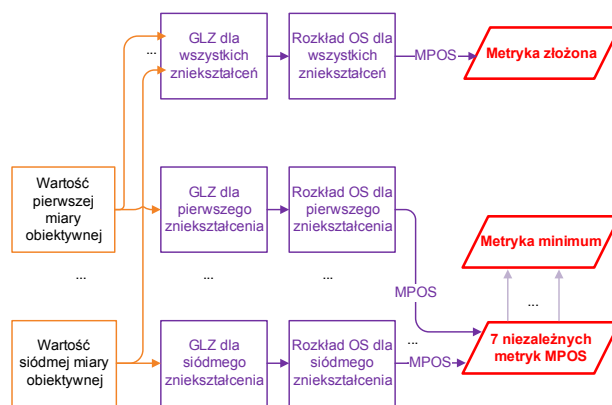
Zmienne porządkowe są zmiennymi, dla których istnieje zdefiniowany porządek, jednak odległość pomiędzy kolejnymi wartościami nie jest stała i znana. Zauważmy, że skala odpowiedzi testerów jest uporządkowana, ponieważ odpowiedzi oznaczają jakość, która może być lepsza lub gorsza. Jednak odległość pomiędzy kolejnymi zmiennymi jest nieznaną, nie wiemy bowiem w jaki sposób konkretny tester rozumie pojęcie „Dobry” lub „Bardzo dobry”. Bardzo częstym błędem popełnianym w analizie zmiennych porządkowych jest używanie narzędzi statystycznych zdefiniowanych jedynie dla zmiennych ilościowych (takim narzędziem jest regresja liniowa).

W celu poprawnego modelowania zmiennych porządkowych należy użyć modelu liniowego uogólnionego GLZ (ang. *Generalized Linear Model*) [10]. Model ten pozwala na estymację prawdopodobieństwa wystąpienia każdej z odpowiedzi w funkcji zmiennych niezależnych, którymi w tym przypadku są miary obiektywne. Znając te prawdopodobieństwa możemy obliczyć zarówno wartość średnią MOS (ang. *Mean Opinion Score*), jak i inne statystyki.

Przeprowadzone badania wykazały, że wartość średnia MOS jest często mylna i różni się od najbardziej prawdopodobnej odpowiedzi. Dla niektórych przypadków najbardziej prawdopodobna była odpowiedź 2 lub 4, ale nigdy 3. Jednocześnie wartość średnia jedynie nieznacznie oscylowała wokół wartości 3 dając mylne wrażenie, że większość użytkowników traktuje jakość serwisu jako średnią. Ponieważ opisywane badania były zamówione przez komercyjną instytucję, zdecydowano, że wartość średnia będzie dawała mylny obraz działania systemu i zaproponowano nowy współczynnik określający jakość systemu, będący odpowiedzią wybraną przez większość testerów. Współczynnik ten oznaczony jest jako MPOS (ang. *Most Probable Opinion Score*).

Dla każdego zniekształcenia wyznaczone zostały współczynniki modeli GLZ, pozwalające obliczyć prawdopodobieństwa każdej z odpowiedzi. Pozwoliło to na wyznaczenie siedmiu niezależnych metryk MPOS. Dodatkowo stworzony został model GLZ, oparty o wszystkie siedem wartości miar obiektywnych. Na podstawie tego modelu wyznaczono metrykę MPOS uwzględniającą wszystkie zniekształcenia. Ponieważ metryka ta jest dość złożona, rozważono dodatkowo zastosowanie metryki minimalnej, wynikającej z prostego założenia: jeżeli obraz zostanie zniekształcony wieloma zniekształceniami, największy wpływ na osta-

teczną ocenę jakości będzie miało zniekształcenie wpływające na ten obraz w sposób najbardziej negatywny. Dlatego też ostatnia zaproponowana metryka to metryka minimum, obliczona jako minimum z metryk MPOS, oceniających każde z siedmiu rozważonych zniekształceń osobno. Wszystkie metryki i ich zależności przedstawione są na rysunku 12.



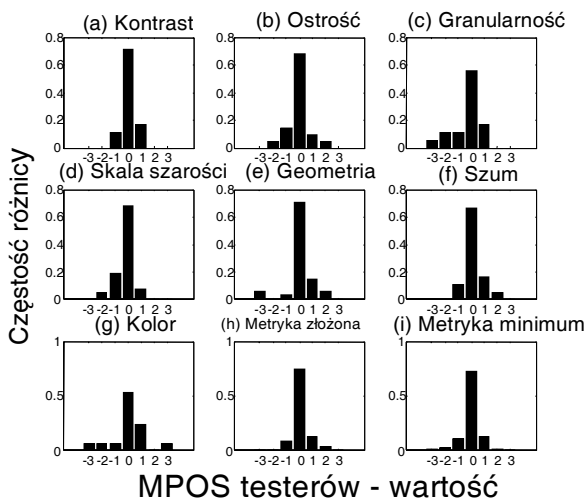
Rys. 12. Schemat analizy metryk

## 5. Weryfikacja wyników

W poprzednim rozdziale zaproponowano 9 różnych metryk MPOS. Siedem z nich pozwala obliczyć, jak bardzo jakość danego obrazu zmieniła się na skutek jednego z siedmiu rozważanych zniekształceń. Kolejne dwie metryki przedstawiają jakość obrazu zniekształconego w dowolny sposób, uwzględniając równocześnie wiele jednoczesnych zniekształceń.

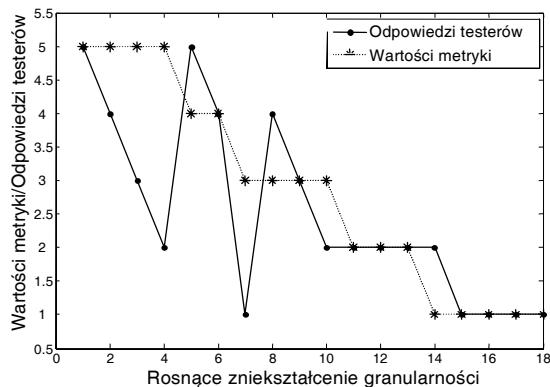
Dokładność zaproponowanych metryk oceniana jest na podstawie różnicy otrzymanych odpowiedzi, a odpowiedzi przedstawionej przez daną metrykę. Ponieważ odpowiedzi metryki są dyskretne (z definicji MPOS), różnica pomiędzy MPOS otrzymanym dla odpowiedzi testerów i metryką może przyjmować wartości od  $-4$  do  $4$ , gdzie ujemna wartość oznacza przeszacowanie jakości obrazu przez metrykę, a dodatnia oznacza niedoszacowanie jakości przez metrykę. Na rysunku 13 przedstawiono otrzymane wyniki.

Wykresy 13a-g przedstawiają pojedyncze zniekształcenia, które oceniane są w odniesieniu do obrazów zniekształconych danym zniekształceniem. Otrzymane wyniki wskazują, że precyzja oceny danego zniekształcenia jest różna. Dla przykładu metryka kontrastu przewiduje zachowanie testerów bardzo precyzyjnie (maksymalna różnica wynosi 1 i blisko 80% odpowiedzi zostało poprawnie przewidzianych). Z drugiej strony metryka granularności jest mniej precyzyjna. Dogłębna analiza pokazała, że prawdziwą przyczyną nieprecyzyjności tej metryki jest większe niezdecydowanie testerów.



**Rys. 13.** Częstości różnic MPOS otrzymanego dla odpowiedzi testerów i wartości metryki dla różnych metryk

Na rysunku 14 przedstawiono wartości metryki oraz odpowiedzi testerów dla rosnącego zniekształcenia granularności.



**Rys. 14.** Wartości MPOS otrzymane dla danych oraz metryki granularności w funkcji rosnącego zniekształcenia granularności

Zauważmy, że dla rosnącego zniekształcenia wartość MPOS nie powinna rosnąć i tak właśnie zachowuje się model. Jednak dane otrzymane przez analizę odpowiedzi testerów zachowują się bardzo chaotycznie. Jest to nieuniknione, ponieważ testerzy to ludzie, którzy nie udzielają w pełni jednoznacznych odpowiedzi i dodatkowo różnią się od siebie pod względem interpretacji postrzeganej jakości.

Najważniejsza z punktu widzenia praktycznego jest metryka umożliwiająca podanie jakości dla całego obrazu, a nie tylko dla konkretnego zniekształcenia. Na rysunkach 13h oraz 13i przedstawiono wyniki dokładności otrzymane dla dwóch rozważanych metryk. Prezentowane wyniki otrzymano zarówno dla obrazów ze zniekształceniami pojedynczymi jak, i wie-

lokrotnymi. Obie metryki przewidują odpowiedzi testerów z dużą precyzją. W obu przypadkach 95% odpowiedzi nie różni się od odpowiedzi testerów o więcej niż 1 OS. Co więcej, obie metryki dokładnie przewidują odpowiedzi użytkowników w 75% przypadków. Ponieważ dokładności metryki złożonej (rys. 13h) oraz metryki minimum (rys. 13i) są porównywalne, w komercyjnym rozwiązaniu będącym wynikiem projektu zaimplementowano miarę minimum wychodząc z założenia, że lepszy jest system prostszy koncepcyjnie niż bardziej skomplikowany, jeżeli wyniki otrzymywane z obu systemów są porównywalne.

## 6. Podsumowanie

W artykule przedstawiono rozwiązanie problemu automatycznego określania subiektywnej jakości obrazu. Autorzy opracowali narzędzie programistyczne dla oceny jakości obrazu, zaimplementowane w postaci pakietu w języku Perl. Dwustopniowa procedura pozwala użytkownikowi porównać dwa obrazy i uzyskać informację o ocenach jakości obrazu zniekształconego. W pierwszym kroku jest analizowanych siedem rodzajów zniekształceń, dobrze opisujących możliwe przypadki utraty jakości. Dokonywane jest obliczenie miar liczbowych dla poszczególnych zniekształceń na podstawie wybranych funkcji pakietu *ImageMagick* oraz oryginalnych algorytmów opracowanych przez autorów na potrzeby pracy. Zidentyfikowano problem powiązań pomiędzy niektórymi zniekształceniami oraz ich wpływu na uzyskiwane miary liczbowe. Dla rozwiązania tego problemu zaproponowano skuteczne algorytmy kompensacji zniekształceń.

W drugim kroku, zastosowano funkcję mapującą otrzymane miary liczbowe zniekształceń na oceny obrazu w sposób zbliżony do oceny subiektywnej dokonywanej przez człowieka. Przebieg funkcji mapującej, wraz z jego statystycznym uzasadnieniem, został wyznaczony za pomocą złożonych technik GLZ, na podstawie wyników wielu przeprowadzonych testów subiektywnych dla obrazu referencyjnego.

## Literatura

- [1] Miyahara, M., Kotani, K.: *Objective picture quality scale (pqqs) for image coding*. In: IEEE Transactions on Communications, vol. 46, no. 9, pp. 1215–1226. University of Utah, Salt Lake City, UT, USA (1998)
- [2] Hosaka, K.: *A new picture quality evaluation method*. In: Proc. International Picture Coding Symposium, pp. 17–18. Tokyo, Japan (1986)

- [3] ITU-T: *Methods for subjective determination of transmission quality*. Recommendation ITU-T P.800 (1996)
- [4] Perceptual Evaluation of Video Quality (2007). <http://www.opticom.de/technology/pevq.html>
- [5] Farias, M., Mitra, S.K.: *No-reference Video Quality Metric Based on Artifact Measurements*. IEEE International Conference on Image Processing, ICIP 2005 3, III – 141–4 (2005)
- [6] Canny, J.: *A computational approach to edge detection*. IEEE Trans. Pattern Anal. Mach. Intell. 8(6), 679–698 (1986)
- [7] Imme, M.: A noise peak elimination filter. CVGIP: Graph. Models Image Process. 53(2), 204–211 (1991). DOI [http://dx.doi.org/10.1016/1049-9652\(91\)90062-O](http://dx.doi.org/10.1016/1049-9652(91)90062-O)
- [8] ITU-T, Geneva, Switzerland: *Methodology for the Subjective Assessment of the Quality of Television Pictures*. Recommendation ITU-R BT.500-11 (1998)
- [9] ITU-T: *Standardized digitized image set*. Recommendation ITU-T T.24 (1998)
- [10] Agresti, A.: *Categorical Data Analysis*. 2<sup>nd</sup> edn. Wiley (2002.)
- [11] Bierens, H.J.: *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press (2004)
- [12] Aguirre-Torres, V., Rios-Curil, A.: *The effect and adjustment of complex surveys on chi-squared goodness of fit tests: Some montecarlo evidence*. In: Proceedings of the Survey Research Methods Section, pp. 602–607. American Statistical Association (ASA) (1994)

*Andrzej Głowacz ukończył studia magisterskie z Telekomunikacji w 2002 oraz studia doktoranckie z Informatyki w Akademii Górniczo-Hutniczej w Krakowie, uzyskując stopień doktora w 2007 roku. Obecnie pracuje jako adiunkt w Katedrze Telekomunikacji AGH. Doświadczenie zawodowe zdobywał realizując projekty europejskie: DAIDALOS, DAIDALOS 2, CARMEN, EuroNGI, EuroFGI, OASIS Archive, GAMA, grantach MNiSW oraz licznych projektach komercyjnych. Główne obszary zainteresowań to: przetwarzanie obrazów, budowa systemów informatycznych, zaawansowane programowanie w systemie Linux oraz jakość usług w sieciach bezprzewodowych. Jest autorem ponad 50 publikacji i raportów technicznych, jest również recenzentem międzynarodowych czasopism i konferencji naukowych.*

*Mgr inż. Michał Grega rozpoczął edukację na Akademii Górniczo-Hutniczej w Krakowie w 2001 roku. Uzyskał dyplom magistra inżyniera z wyróżnieniem w specjalności Telekomunikacja. Od 2006 roku jest słuchaczem studiów doktoranckich w Katedrze Telekomunikacji AGH. Od 2005 roku bierze czynny udział w pracach naukowo-badawczych Katedry Telekomunikacji w zespole „Sieci Szerokopasmowe i Usługi Multimedialne” pod kierownictwem prof. Zdzisława Papira. Brał udział w 4 projektach międzynarodowych i realizował liczne granty krajowe jak i zlecenia dla przemysłu. Jest współautorem ponad 25 publikacji i ponad 40 referatów, w tym artykułu w czasopiśmie „Multimedia Tools and Applications” wydawnictwa Springer oraz autorem artykułu opublikowanego na konferencji IEEE INFOCOM.*

*Dr inż. Lucjan Janowski jest adiunktem w Katedrze Telekomunikacji Akademii Górniczo-Hutniczej. Otrzymał tytuł zawodowy magistra inżyniera telekomunikacji w 2002 roku i stopień doktora telekomunikacji w 2006 roku – oba uzyskane w Akademii Górniczo-Hutniczej. W roku 2007 pracował na pozycji post-doc w Laboratory for Analysis and Architecture of Systems of CNRS we Francji, gdzie zajmował się problematyką wykrywania anomalii i ataków sieciowych. Jego głównym zainteresowaniem jest statystyka oraz modelowanie ruchu z wykorzystaniem modeli probabilistycznych. Brał też udział w kilku projektach, zarówno naukowych, jak i komercyjnych. Jest autorem kilku publikacji.*

*Dr inż. Mikołaj Leszczuk, adiunkt w Katedrze Telekomunikacji Akademii Górniczo-Hutniczej (Kraków). Jego aktualne zainteresowania badawcze koncentrują się wokół rozwoju cyfrowych bibliotek wideo, w szczególności podsystemów streszczania, indeksowania, kompresji i strumieniowania sekwencji wizyjnych. Wyniki tych badań zostały opublikowane w 2006 roku w formie rozprawy doktorskiej. Wybrane projekty, w których uczestniczył lub uczestniczył Mikołaj Leszczuk: INDECT (7PR, jako Administrator Projektu), GAMA (eContentPlus, jako lokalny kierownik), CONTENT (6PR), NOBEL (6PR), E-NEXT (6PR), PRO-ACCESS (5PR, jako członek Steering Committee), E-NET (5PR), ATAMS 2001 (5PR), LION (5PR), BTI (4PR), OASIS Archive (Kultura 2000, jako lokalny kierownik).*

*Mgr inż. Piotr Romaniak uzyskał tytuł zawodowy mgr inż. Telekomunikacji w roku 2006 w Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Obecnie jest uczestnikiem studiów doktoranckich w Katedrze Telekomunikacji AGH na kierunku Telekomunikacja. Od 2004 roku Piotr Romaniak był zaangażowany w międzynarodowe projekty realizowane w Katedrze Telekomunikacji w ramach programów Culture 2000, eContentPlus, FP6 i FP7. Uczestniczył także w projektach realizowanych na zamówienie polskich operatorów telekomunikacyjnych oraz we współpracy z innymi jednostkami edukacyjnymi. Obecnie jest zaangażowany w trzy projekty europejskie oraz trzy projekty krajowe.*