

## Krótki przegląd pierwszych standardów kompresji obrazów

Najstarszymi (1980 rok) i szeroko stosowanymi obecnie standardami kompresji obrazów cyfrowych są międzynarodowe standardy kodowania cyfrowych faksów, odpisów (ang. facsimile) Grupy 3 i Grupy 4 opracowane przez grupę konsultacyjną CCITT (*Consultative Committee of the International Telephone and Telegraph*). Standardy te dotyczą jedynie binarnych obrazów zawierających teksty i dokumenty. Wzrost liczby zastosowań wielopoziomowych obrazów cyfrowych w szerokiej gamie zastosowań nieuchronnie prowadzi do opracowywania nowych standardów kompresji. Udogodnienia związane z wprowadzeniem tych standardów dotyczą nie tylko łatwiejszej wymiany obrazów pomiędzy różnymi systemami i aplikacjami, lecz także pozwalają na znaczące ograniczenie kosztu budowy wyspecjalizowanych urządzeń cyfrowych niezbędnych w wielu systemach kompresji obrazów w czasie rzeczywistym. W ostatnich latach prace nad nowymi standardami kompresji obrazów prowadzone były w trzech zasadniczych kierunkach:

### **- obrazy binarne**

W 1988 roku został uformowany komitet znany jako JBIG (*Joint Bilevel Imaging Group*) pod auspicjami ISO-IEC/JTC1/SC2/WG8 i CCITT SG VIII NIC, w celu opracowania standardu kompresji i dekompresji obrazów binarnych. Grupa skoncentrowała swoje wysiłki na poszukiwaniu efektywniejszego algorytmu od opracowanych wcześniej przez CCITT, w zastosowaniu do klasycznych aplikacji (n.p. ośmiu binarnych obrazów odniesienia zaproponowanych przez CCITT), a także rozszerzenia ich stosowalności do nowych aplikacji. Chodziło głównie o opracowanie algorytmów progresywnych i adaptacyjnych.

Należy też wspomnieć o opracowanej przez naukowców z IBM w 1988 roku binarnej wersji koda arytmetycznego o nazwie Q-koder i zastosowaniu go do kompresji obrazów binarnych - technika ABIC (*Arithmetic Binary Image Compression*).

- CCITT Group 3 i 4 - kodowanie długości sekwencji do kompresji obrazów binarnych
- JBIG/JBIG2 (Joint Bilevel Imaging Group) - kodowanie arytmetyczne do kompresji danych binarnych

### **- pojedyncze obrazy wielopoziomowe, monochromatyczne i kolorowe**

Komitet znany powszechnie pod nazwą JPEG (*Joint Photographic Experts Group*), działający jako ISO-IEC/JTC1/SC2/WG10 przy bliskiej nieformalnej współpracy z CCITT SG VIII NIC, uformowany został pod koniec 1986 roku w celu opracowania międzynarodowego standardu dla pojedynczych, wielopoziomowych, monochromatycznych i kolorowych obrazów. Zadaniem zespołu było zdefiniowanie standardu dla tak różnych zastosowań jak foto- i telegazeta, grafika komputerowa, skład komputerowy, mała poligrafia, kolorowe faksy, systemy medyczne i wiele innych. Pomimo tego, iż w tej dziedzinie nie istniały wcześniejsze standardy, członkowie JPEG byli silnie przekonani, że wymagania zdecydowanej większości tych zastosowań winny być uwzględnione w standardzie. Ostateczna propozycja standardu, która została opublikowana w 1992/1993 r. jako standard międzynarodowy ISO/IEC zawiera trzy główne składniki: 1) system podstawowy, który zawiera prosty i efektywny algorytm, adekwatny w stosunku do większości zastosowań kompresji obrazów, 2) zbiór rozszerzeń systemu zawierający przede wszystkim algorytm progresywnego kodowania rozszerzający pole zastosowań, 3) niezależna bezstratna metoda kodowania dla zastosowań wymagających tego typu kompresji. Bardziej dokładne omówienie tego standardu, ze względu na przewidywane przez twórców jego zastosowanie także w systemach obrazowania medycznego, zostanie przedstawione w następnej części tego podrozdziału.

## Standard JPEG

Specyfikacja normy standardu JPEG zawiera:

- opis procesu przetwarzania źródłowych danych obrazowych w dane obrazowe skompresowane;
- opis procesu przetwarzania skompresowanych danych obrazowych w zrekonstruowane dane obrazu;
- wskazania dotyczące praktycznych implementacji standardu;
- opis zakodowanej reprezentacji skompresowanych danych obrazowych.

Specyfikacja nie opisuje kompletnej zakodowanej reprezentacji obrazu, może ona zawierać pewne parametry zależne od aplikacji. W normie znajdują się cztery techniki kompresji, a mianowicie:

1. Podstawowy proces kodowania (ang. baseline process).
2. Rozszerzony, bazujący na DCT proces kodowania (ang. extended DCT- based process).
3. Bezstratny proces kodowania (ang. lossless process).
4. Hierarchiczny proces kodowania (ang. hierarchical process).

Kodowanie każdego składnika obrazu (luminancja, składowe chrominancji) przebiega analogicznie. W podstawowym procesie kodowania dane wejściowe muszą być ośmiobitowe. Obraz jest dzielony na bloki 8x8 i każdy blok jest transformowany za pomocą DCT (ang. discrete cosine transform), przy czym kodowanie obrazu przebiega sekwencyjnie, tzn. z lewej strony na prawą zaczynając od góry obrazu i przemieszczając się na dół.

W standardzie JPEG zastosowano zbiór funkcji bazowych transformacji DCT, 2-D dyskretna transformacja kosinusowa zdefiniowana jest w sposób następujący:

$$k(u, v) = \frac{1}{\sqrt{2N}} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}, \quad (1)$$

a transformacja odwrotna:

$$f(x, y) = \frac{1}{\sqrt{2N}} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)k(u, v) \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N}, \quad (2)$$

gdzie  $C(u), C(v) = 1/\sqrt{2}$  dla  $u, v = 0$ ;  $C(u), C(v) = 1$  wpp. oraz  $k(u, v)$  są wartościami współczynników transformaty w jej dziedzinie,  $f(x, y)$  są wartościami pikseli w dziedzinie obrazu,  $N$  zaś określa rozmiar bloku  $N \times N$  obrazu, w którym liczona jest transformata.

Obliczone 64 współczynniki z każdego bloku podlegają kwantyzacji. **Każdy ze współczynników jest dzielony przez odpowiadającą mu wartość w tablicy kwantyzacji. Tablicę tę można dobrać w zależności od aplikacji.** Metoda pozwala połączyć progowanie i równomierną kwantyzację wartości współczynników transformaty zorientowanych przestrzennie i oznaczonych  $k(u, v)$  według następującego wyrażenia:

$$\hat{k}(u, v) = NINT\left[\frac{k(u, v)}{z(u, v)}\right], \quad (3)$$

gdzie  $\hat{k}(u, v)$  jest skwantowaną aproksymacją  $k(u, v)$ , a  $z(u, v)$  oznacza element tablicy kwantyzacji, określającej progi i jednocześnie elementarne przedziały kwantyzacji. *NINT* - zaokrąglenie do najbliższej liczby całkowitej.

Oznacza to, że dla przykładowej wartości  $z(u, v)$  równej  $c$  (zobacz rysunek 1a) wartość  $\hat{k}(u, v)$  przyjmuje wartość  $p$  wtedy i tylko wtedy, jeśli

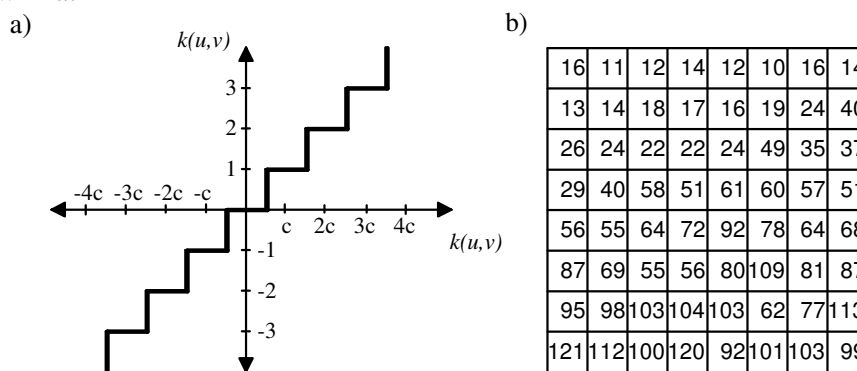
$$pc - \frac{c}{2} \leq k(u, v) < pc + \frac{c}{2}. \quad (4)$$

Jeśli  $z(u, v) > 2|k(u, v)|$ , wówczas  $\hat{k}(u, v) = 0$ , czyli następuje zupełne usunięcie tego współczynnika ze zbioru kompresowanych danych. Przykładowa postać tablicy kwantyzacji została przedstawiona na rys. 1 b).

W celu osiągnięcia różnych stopni kompresji tablica ta jest modyfikowana w następujący sposób:

$$z'(u, v) = \text{NINT}[z(u, v) \cdot c], \quad u, v = 0 \dots 7 \quad (5)$$

gdzie  $z'(u, v)$  jest nową wartością w tablicy zamiast  $z(u, v)$ , a  $c$  jest stałą dobieraną przez użytkownika.



Rys. 1. a) krzywa równomiernej kwantyzacji z selekcją progową; b) tablica kwantyzacji optymalizująca jakość rekonstruowanych obrazów pod kątem oceny psychowizualnej (zamieszczona w normie standardu JPEG, jednak nie obligatoryjna).

Powyższy schemat kwantyzacji pozwala więc przy niewielkim błędzie kwantyzacji zachować współczynniki o wyższych częstotliwościach, a ponadto, poprzez dobranie odpowiednich wartości tablicy kwantyzacji, zgodnie z perceptualną wagą poszczególnych współczynników, wysoką psychowizualną jakość obrazu. Dodatkową zaletą jest zmniejszenie efektów blokowych poprzez zastosowanie tego samego przedziału kwantyzacji dla składowej stałej we wszystkich blokach.

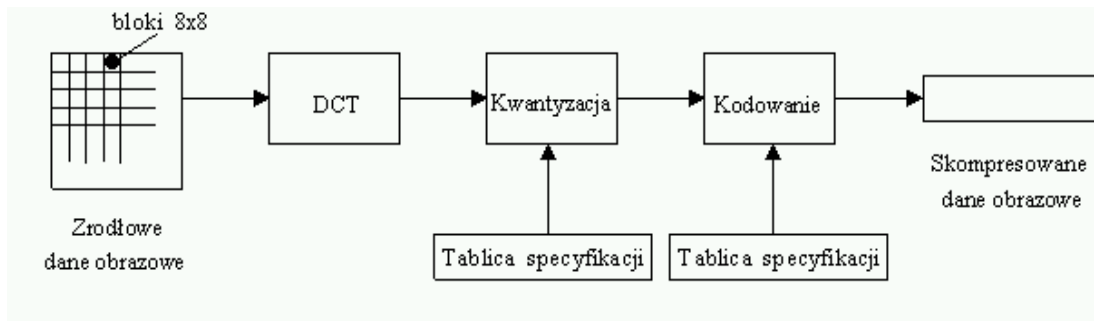
Standard JPEG dopuszcza dowolną postać tablicy kwantyzacji, jednak w normie podano przykładową postać tablicy kwantyzacji (1b)) dla luminancji, która zawiera doświadczalnie wyznaczone wartości (w wyniku wieloletnich badań nad psychowizualnym odbiorem obrazów naturalnych), odzwierciedlające perceptualną wagę poszczególnych współczynników.

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Rys. 2. Kolejność ustawienia wartości współczynników w bloku w jednowymiarowy ciąg danych według sekwencji zygzak.

W celu efektywniejszej kompresji wartości współczynników w blokach są ustawiane w jednowymiarowy ciąg danych według sekwencji zygzak (rys.2). Ze względu na charakterystyczny rozkład tych wartości w bloku (dominującej najczęściej, co do wartości bezwzględnej, składowej stałej i coraz mniejszych wartości współczynników wyższych częstotliwości, często zerowych), taki rodzaj dekompozycji danych w bloku pozwala na kolejne ustawienie danych o zbliżonych wartościach. Uzyskuje się w ten sposób jednowymiarowy ciąg wartości zakończony dużą ilością zer, który może być bardzo skutecznie kodowany przy pomocy techniki RLE.

Tak uszeregowane dane z poszczególnych bloków są następnie kodowane, przy czym składową stałą koduje się różnicowo, tzn. koduje się jedynie różnicę pomiędzy wartością składowej stałej obecnego bloku i poprzedniego. Do kodowania używa się kodu Huffmana, przy czym tablica Huffmana nie jest objęta normą. Norma dopuszcza także stosowanie kodera arytmetycznego.



Rys. 3. Schemat blokowy algorytmu kompresji wykorzystującego DCT.

Proces dekompresji przebiega dokładnie odwrotnie, przy czym dekoder musi posługiwać się dokładnie tymi samymi tablicami specyfikacji (tablica kwantyzacji, Huffmana). Umożliwia to format zapisu, w którym tablice specyfikacji poprzedzone odpowiednimi markerami umieszczone są w pliku razem z danymi skompresowanymi. Dokładny opis formatu znajduje się w normie.

Rozszerzony proces kompresji umożliwia kompresję zarówno 8- bitowych jak i 12- bitowych danych, przy czym kodowanie może być nie tylko sekwencyjne, ale i progresywne.

W modzie progresywnym poszczególne bloki współczynników są kodowane w tej samej kolejności, ale w wielu skanach dzielących współczynniki każdego bloku na poszczególne pasma. Związane jest to jednak z zapewnieniem dodatkowego bufora pamięci do przechowania wartości współczynników z całego obrazu po fazie kwantyzacji, a przed statystycznym kodowaniem tych wartości. Są one następnie kolejno kodowane w skanach zbierających wartości w określonym paśmie ze wszystkich bloków.

W normie występują dwa rodzaje procedur progresywnego kodowania. Pierwsza, nazywana selekcją widma, dzieli współczynniki ustawione według sekwencji zygzak na kolejne pasma, które zawierają poszczególne części częstotliwościowego spektrum każdego z bloków. Druga procedura związana jest z precyzją, z jaką kodowane są współczynniki w każdym z pasm i nazywana jest sukcesywną aproksymacją. Najpierw kodowana jest pewna liczba bardziej znaczących bitów wartości tych współczynników, a następnie mniej znaczące bity. Można progresywnie kodować współczynniki jedynie przy pomocy procedury selekcji widma, jak też z wykorzystaniem obu procedur. Wówczas mamy do czynienia z tzw. pełną progresją. Ze stwierdzeń zawartych w opisie normy wynika, że zastosowanie selekcji widmowej, jakkolwiek wygodne dla wielu zastosowań, daje porównywalne bądź nieco gorsze wyniki kompresji niż sekwencyjna metoda kodowania, podczas gdy przy pełnej progresji skuteczność kompresji może się okazać nieco większa.

W rozszerzonym procesie kodowania możliwe jest także arytmetyczne kodowanie, a tablica warunkowych zależności danych (ang. conditioning table) jest wówczas zapamiętywana jako tablica specyfikacji.

Proces kodowania bezstratnego ujęty w standardzie nie wykorzystuje DCT, lecz polega na predykcyjnym kodowaniu wartości piksela na podstawie trzech wartości pikseli sąsiednich. Dla danego skanu można wybrać jedną z ośmiu predykcji przedstawionych na rys. 4. Następnie oblicza się różnicę pomiędzy wartością przewidywaną  $P_x$ , a rzeczywistą wartością piksela i koduje się otrzymaną różnicę odpowiednio dostosowaną metodą Huffmana lub arytmetycznie. W metodzie Huffmana stosuje się 17 kategorii różnicowych wartości, a w kodowaniu arytmetycznym buduje się dwuwymiarowy model statystyczny.

						<u>Numer</u>	<u>Predykcja</u>
						0	Bez predykcji
						1	$P_x=R_a$
		c	b			2	$P_x=R_b$
		a	x			3	$P_x=R_c$
						4	$P_x=R_a+R_b-R_c$
						5	$P_x=R_a+((R_b-R_c)/2)$
						6	$P_x=R_b+((R_a-R_c)/2)$
						7	$P_x=(R_a+R_b)/2$

Rys. 4. Przestrzenna relacja pomiędzy pikselami (a,b,c), których wartości służą do wyznaczania wartości przewidywanej w pikselu x. Obok rodzaje predykcji dla bezstratnego kodowania.

Predykcja 0 może być używana dla różnicowego kodowania w modzie hierarchicznym, 1,2 i 3 to predykcje jednowymiarowe wykorzystywane głównie na skraju obrazów, podczas gdy predykcje 4-7 są dwuwymiarowe. Można tą metodą odwracalnie

kodować od 2 do 16 bitowe wartości pikseli obrazu. Można więc go z łatwością implementować do kompresji szerokiej klasy obrazów.

Hierarchiczny proces kodowania polega na kompresji obrazu poprzez zakodowanie sekwencji kadrów tego obrazu (różnicowo lub nieróżnicowo) o różnej rozdzielczości za pomocą rozszerzonego procesu kodowania (wykorzystującego DCT) lub metody bezstratnej. Można także połączyć te dwie metody i po kodowaniu opartym na DCT w ostatecznej fazie zastosować kodowanie bezstratne (bez fazy kwantyzacji). Wówczas zrekonstruowany obraz różni się od oryginalnego jedynie wskutek błędów przybliżeń wartości DCT i IDCT.

Tworzy się zrekonstruowane składniki odniesienia obrazu różnej rozdzielczości i koduje się różnice pomiędzy obrazem oryginalnym a tymi składnikami odniesienia. Stosowanie filtrów próbkujących obraz oryginalny z różną rozdzielczością tworzy charakterystyczną piramidę przestrzennej rozdzielczości.

Mod hierarchiczny może być stosowany alternatywnie, aby zwiększyć jakość rekonstruowanych składników obrazu o danej rozdzielczości. Oferuje progresywną prezentację czy transmisję, ale jest najczęściej użyteczna jedynie w systemach posługujących się wielorozdzielczymi wersjami danych obrazów.