

Podstawy Biometrii – zagadnienia teoretyczne






1. Algorytm kompresji JPEG

JPEG (ang. Joint Photographic Experts Group) jest to algorytm stratnej kompresji statycznych obrazów cyfrowych. Oficjalnie został on zdefiniowany jako standard: ISO/IEC IS 10918-1. Głównym celem powstania tego algorytmu było wsparcie kompresji obrazu dla różnych zastosowań. Jako jedno z podstawowych założeń wskazano, że algorytm musi mieć możliwość kompresji zdjęć różnych rozmiarów o dowolnej przestrzeni kolorów jak również powinna istnieć możliwość uzyskania kompresji o bardzo dużym współczynniku. Co więcej to użytkownik powinien mieć możliwość regulacji wartości współczynnika kompresji. Jedną z cech tego algorytmu jest możliwość uzyskania obrazu o wysokiej jakości rekonstrukcji.

Przebieg algorytmu:

- a) **Transformacja przestrzeni barw obrazu z RGB (Red, Green, Blue) do przestrzeni YUV** (znaną także jako YCbCr – opisuje ona obraz poprzez jasność (luminancję - Y) oraz dwa kanały barwy (chrominancję – Cb, Cr)). Należy podkreślić, że w przypadku przestrzeni YUV, to luminancja musi być przechowywana w dużej rozdzielczości (ze względu na to, że ludzkie oko bardzo łatwo zauważa zmiany w jasności) natomiast chrominancje mogą być przechowywane w mniejszej rozdzielczości. Krok ten jest opcjonalny.
- b) **Podpróbkowanie składowych chrominancji** – wstępne odrzucenie części pikseli kanałów barwy. Standard JPEG wykorzystuje podpróbkowanie zgodne ze schematem 4:2:0. Składowe chrominancji są podpróbkowane i ich rozdzielczość jest dwukrotnie mniejsza od składowej luminancji. Krok ten również jest opcjonalny.
- c) **Podział obrazu na bloki 8x8 pikseli.**
- d) **Wykonanie transformaty kosinusowej DCT na wszystkich wydzielonych w kroku poprzednim blokach.** Dzięki wykonaniu tego kroku pozyskujemy informacje o średniej wartości wewnątrz bloku a także o częstotliwości zmian w bloku. Obydwie wartości są wyrażane w formie liczb zmiennoprzecinkowych. Transformata kosinusowa jest odwracalna, więc w tym kroku nie tracimy żadnych danych.
- e) **Kwantyzacja współczynników transformaty.** Jest to etap w ramach którego następuje utrata części danych. Kwantyzacja polega bowiem na zastąpieniu liczb zmiennoprzecinkowych przez wartości całkowite. Liczba utraconych danych jest zależna od parametrów kompresora.
- f) **Kodowanie współczynników transformaty.**
- g) **Utworzenie wektorów z tablic współczynników po kwantyzacji.** Po kwantyzacji linearyzujemy macierz współczynników, zgodnie z algorytmem Zig-Zag i kodujemy entropijnie. Ten krok powoduje, że współczynniki zerowe leżą obok siebie.
- h) **Kodowanie niezerowych współczynników zgodnie z algorytmem Huffmana.**

Przykłady obrazów

Obraz	Jakość (Q)
	100
	50
	25
	10
	1

Źródło obrazów: <https://pl.wikipedia.org/wiki/JPEG>

Struktura pliku JPEG

Plik JPEG składa się z sekwencji segmentów (które reprezentują poszczególne części obrazu), z których każdy rozpoczyna się od tzw. markera. Każdy marker zaczyna się od bajtu 0xFF po którym następuje bajt wskazujący jego rodzaj. Niektóre markery składają się tylko i wyłącznie z tych dwóch bajtów natomiast inne mogą posiadać dodatkowe informacje w formie dodatkowych dwóch bajtów wskazujących na długość danych specyficznych dla danego markera. Z kolei innym rodzajem są markery, po których następują dane zakodowane z wykorzystaniem metody entropii. W ramach Tabeli 1 przedstawione zostały standardowe markery opisujących każdy z segmentów obrazu.

Tabela 1. Standardowe markery

Dane	Opis
SOI	Start of Image – początek obrazu
SOF0	Start of Frame – wskazuje na wykorzystanie podstawowej formy DCT i określa szerokość, wysokość, liczbę komponentów oraz formę ich podpróbkowania.
SOF2	Start of Frame – obecność danych w tym miejscu wskazuje na użycie progresywnej formy DCT i podobnie jak SOF0 określa szerokość, wysokość, liczbę komponentów oraz schemat ich podpróbkowania.
DHT	Define Huffman Table – definicja tablic Huffmana, które są stosowane podczas kodowania elementów w procesie kompresji JPEG.
DQT	Define Quantization Table – definicja tablic kwantyzacji, stosowanych w algorytmie kompresji JPEG.
DRI	Define Restart Interval – specyfikuje przerwę pomiędzy kolejnymi markerami typu RST w jednostkach MCU. Po tym elemencie podawane są wartości dwubajtowe, które wskazują na ustandaryzowany rozmiar. Powoduje to, że może być on traktowany jak każda inna zmienna wskazująca na rozmiar segmentu.
SOS	Start Of Scan – rozpoczyna zstępujące skanowanie obrazu. W przypadku podstawowej formy DCT jest to zazwyczaj pojedynczy skan, gdy rozważana jest progresywna forma tej transformaty wtedy wskazuje się na wiele skanów. Ten element wskazuje, który fragment danych będzie zawierać, jego bezpośrednim następnikiem są dane zakodowane z wykorzystaniem metody entropii.
RSTn	Restart – wprowadzany co kolejne r mikrobloków, gdzie r jest wartością wskazywaną przez marker DRI. Nie jest on używany w momencie w którym wartość DRI nie została ustawiona.
APPn	Application-Specific – są to dane specyficzne dla pewnych aplikacji, przykładowo mogą być tutaj przechowywane pewne metadane skorelowane z obrazem (np. data ostatniej modyfikacji, właściciel etc.).
COM	Comment – zawiera tekstowy komentarz do obrazu
EOI	End Of Image – zakończenie obrazu.

2. Algorytm kompresji bezstratnej PNG

PNG (Portable Network Graphics) – jest to rastrowy format plików graficznych oraz system bezstratnej kompresji danych graficznych. Został on przygotowany w 1995 roku jako następca GIF'a po złożeniu roszczeń patentowych dotyczących kompresji LZW, która była używana w ramach poprzednika PNG.

PNG obsługuje stopniowaną przezroczystość (kanał alfa) oraz 48-bitową głębię kolorów (zwróćmy uwagę, że oznacza to 16 bitów przeznaczonych na każdy z kanałów koloru). Technika, która jest stosowana w formacie PNG jest wprowadzanie filtracji obrazu tuż przed dokonaniem jego kompresji. Pozwala to na znaczące uproszczenie obrazu co z kolei prowadzi do zwiększenia wydajności algorytmu kompresji. Odwrotnie niż w formacie JPEG, format PNG domyślnie nie przewiduje wyboru stopnia kompresji (aczkolwiek niektóre aplikacje pozwalają na taki wybór).

Na czym zatem polega kompresja wykonywana w ramach formatu PNG? Otóż dokonuje się jednej transformacji każdej z linii obrazu a następnie wynik jest dodatkowo kompresowany z wykorzystaniem algorytmu deflate. Numer transformacji jest zapisywany przed linią, której dotyczył. Transformacje są skorelowane z dostosowaniem danych w celu uproszczenia procesu kompresji. Prowadzą one do zapisania różnicy pomiędzy wartością rzeczywistą badanego piksela a wartością obliczoną na podstawie wybranej funkcji przewidującej. Funkcje przewidujące mogą odnosić się do pikseli leżących po stronie lewej badanego piksela (np. najbliższy piksel leżący po stronie lewej), do średnich wartości najbliższych sąsiadów czy też ich sumy.

PNG został określony w ramach standardu RFC 2083, natomiast algorytm deflate, który jest w ramach niego stosowany jest przedstawiony w standardzie RFC 1951. Format PNG jest zalecany przez konsorcjum W3C do stosowania w ramach sieci WWW – dotyczy to takich obszarów jak chociażby grafika reprezentacyjna czy też schematy blokowe.

Struktura pliku PNG

W najbardziej ogólnym ujęciu plik PNG składa się z tzw. podpisu PNG po którym następują ciągi danych. Pierwsze osiem bajtów pliku PNG zawsze zawiera wartości dziesiętne. Te wartości zawsze wskazują na to, że dany plik reprezentuje pojedynczy obraz PNG. Ciągi danych zawsze zaczynają się od bloku IHDR a kończą blokiem IEND.

Każdy blok (ciąg danych) składa się z czterech elementów: długości bloku (wartość 4-bajtowa wskazująca na liczbę bajtów danych w bloku), rodzaju bloku (wartość 4-bajtowa wskazująca na rodzaj bloku, dla zachowania czytelności przyjmuje się, że rodzaj bloku może być opisany tylko i wyłącznie z wykorzystaniem małych i dużych znaków ASCII), danych zawartych w bloku oraz czterobajtowego CRC, który zapewnia że struktura pliku została zachowana (bierze pod uwagę typ bloku oraz pola danych zawartych w bloku ale nie uwzględnia długości bloku).

Pierwszym blokiem występującym w ramach każdego pliku PNG jest blok rozpoczynający IHDR. Jest to tzw. nagłówek obrazu. Zawiera on informacje o szerokości i wysokości obrazu, głębi kolorów, rodzaju wykorzystywanej przestrzeni kolorów, metodzie kompresji, metodzie filtracji oraz sposobie splotu. Do podstawowych bloków, które mogą wystąpić w pliku PNG jest blok PLTE, wskazuje on na paletę kolorów poprzez odpowiednie ustawienie wartości dla parametrów R,G,B.

Kolejnym blokiem, który możemy zaobserwować jest IDAT. Reprezentuje on dane skorelowane z obrazem a dokładniej rzecz ujmując posiada on dane z wyjściowego strumienia danych uzyskiwanego po wykonaniu kompresji obrazu. W pliku PNG może występować wiele bloków IDAT aczkolwiek nie mogą pomiędzy nimi występować żadne inne bloki.

Ostatnim elementem każdego pliku PNG jest blok IEND. Występuje on zawsze na końcu pliku i wskazuje na zakończenie strumienia danych PNG.

Do opcjonalnych bloków pliku PNG możemy zaliczyć:

- bKGD – specyfikuje on domyślny kolor tła, aplikacje wyświetlające pliki najczęściej nie korzystają z tego bloku.
- cHRM – podstawowe wartości składowych chromatycznych oraz punkt bieli
- gAMA – specyfikuje on podstawową wartość parametru gamma obrazu
- hIST – zawiera informacje o histogramie obrazu w aktualnie wykorzystywanej przestrzeni barw
- pHYS – fizyczny rozmiar piksela na obrazie
- sBIT – stanowi on uproszczenie pracy dekodatorów obrazu albowiem wskazuje on na najważniejsze bity na obrazie
- tEXt – dane tekstowe skorelowane z obrazem
- tIME – termin ostatniej modyfikacji obrazu
- tRNS – wskazuje na to czy obraz wykorzystuje prostą formę przezroczystości oraz przechowuje informacje o wartości parametru alpha
- zTXt – skompresowane dane tekstowe