

## **KOMPRESJA OBRAZÓW RUCHOMYCH Z WYKORZYSTANIEM KODOWANIA SUBPASMOWEGO**

**Andrzej Popławski**

**Instytut Informatyki i Elektroniki, Uniwersytet Zielonogórski  
65-246 Zielona Góra, ul. Podgórna 50**

*e-mail: A.Poplawski@iie.uz.zgora.pl*

### **STRESZCZENIE**

W artykule przedstawiono zagadnienia kompresji obrazów video z wykorzystaniem techniki kodowania falkowego. Wprowadzono czytelnika w ogólne zagadnienia stratnej kompresji obrazów ruchomych. Zaprezentowano trójwymiarowy kodek falkowy IMC-3DEZBC, wykorzystujący kompensację ruchu, ze skalowalnością typu SNR. Przedstawiono wyniki eksperymentalne badań kodeka z wykorzystaniem wybranych sekwencji testowych.

### **1. WPROWADZENIE**

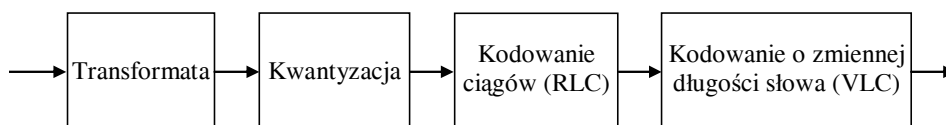
Nieustanny postęp techniczny, zwiększające się moce obliczeniowe współczesnych komputerów, zapotrzebowanie na różne usługi multimedialne, stymulują rozwój nowych technik kodowania, zarówno obrazów nieruchomych jak i sekwencji wizyjnych. Cyfrowa telewizja satelitarna, telekonferencje, videotelefonie, transmisja obrazów w sieci Internet są dzisiaj codziennością. Powstanie nowych, bardziej efektywnych technik kompresji obrazu, może przyczynić się do zaistnienia nowych usług audiowizualnych.

Standard kompresji cyfrowych danych audiowizualnych ISO/IEC IS 13818 [5], znany głównie pod nazwą MPEG-2, jest obecnie wykorzystywany m.in. w cyfrowej telewizji satelitarnej, zapisie materiału filmowego na nośniki DVD, jak również w transmisji informacji wizualnej poprzez Internet. Opiera się on o dyskretną transformatę kosinusową. Korzystanie z transformaty DCT, pomimo jej wielu pozytywnych właściwości, prowadzi do powstawania artefaktów (efekty blokowania), szczególnie przy małych przepływnościach strumienia danych.

W ostatnich latach wzrosło zainteresowanie koderami wykorzystującymi dyskretną transformatę falkową (DWT), głównie ze względu na naturalną skalowalność, braku efektu blokowego oraz większą efektywność kompresji obrazów. Skalowalność jest cechą szczególnie cenną w przypadku przesyłania danych przez sieci heterogeniczne, umożliwia przesłanie sekwencji video z zadaną przepływnością, co wpływa bezpośrednio na jakość obrazu. Standard MPEG-2 określa cztery typy skalowalności: przestrzenną, czasową, SNR oraz skalowalność przez podział danych [2], [5].

## 2. METODY STRATNEJ KOMPRESJI OBRAZÓW

Współczesne techniki stratnej kompresji obrazów opierają się na kodowaniu z wykorzystaniem transformaty. Obecnie w praktyce używa się dyskretnej transformaty kosinusowej (DCT – *discrete cosine transform*) oraz dyskretnej transformaty falkowej (DWT – *discrete wavelet transform*). Ogólny schemat kodowania transformatowego przedstawiono na rys. 1.



Rys. 1. Ogólny schemat kodowania transformatowego

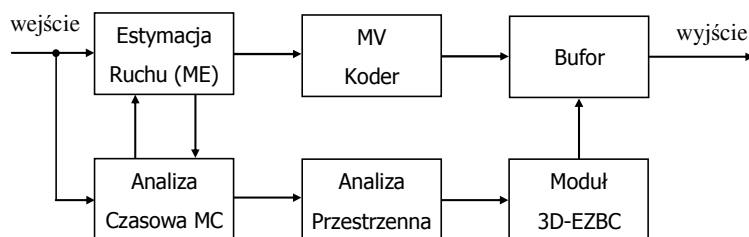
Dane po przetworzeniu transformatą są kwantowane celem zmniejszenia ilości informacji w nich zawartych. Na tym etapie dokonuje się właściwa kompresja, zawiązana z nieodwracalną utratą informacji, zaś otrzymane wartości zaokrągla się najczęściej do najbliższej liczby całkowitej. Następnie dane zostają poddane kodowaniu ciągów, liczby określające wartości RLC są przetwarzane za pomocą kodów Huffmana lub kodów arytmetycznych. Dekodowanie tak zapisanych informacji przebiega w odwrotnym kierunku, czyli: dekodowanie kodu o zmiennej długości słowa, dekodowanie ciągów, dekwantyzacja i wykonanie odwrotnej transformaty.

Przedstawiony schemat kodowania jest wykorzystywany dla zapamiętania informacji o jednej klatce obrazu – jest to tak zwane kodowanie wewnątrzobrazowe (*ang. intraframe*). Uzyskanie większych stopni kompresji wymaga kodowania międzyobrazowego (*ang. interframe*), polegającego na wykorzystywaniu podobieństw pomiędzy sąsiadującymi obrazami. Wykorzystuje się w tym celu technikę predykcji z kompensacją ruchu (*ang. motion compensated prediction*), polegającą na wyznaczeniu estymacji ruchu dla bloków o zadanym rozmiarze (np.: 8x8, 16x16). W obrazie poprzednim poszukuje się najbardziej podobnych bloków, różnice położenia bloków wyznaczają wektory ruchu (MV – *motion vector*). Wykorzystując kompensację ruchu wyznaczany jest przewidywany obraz, który porównuje się z obrazem na wyjściu kodera – różnica tych obrazów wyznacza błąd predykcji, który następnie jest kodowany podobnie jak w przypadku kodowania wewnątrzobrazowego.

## 3. KODEK IMC-3DEZBC

Badania mające na celu zwiększenie efektywności kompresji obrazów ruchomych prowadzone są na bazie kodeka IMC-3DEZBC (*ang. Invertible Motion Compensation – 3-Dimensional Embedded Zero Block Coding*) autorstwa S.-T. Hsiang-a oraz J.W. Woods-a. Kodek ten został opisany w [4]. Jest to trójwymiarowy kodek falkowy wykorzystujący

kompensację ruchu, ze skalowalnością typu SNR (*ang. SNR scalability*). Trójwymiarowe kodowanie subpasmowe z kompensacją ruchu zostało zapoczątkowane przez Ohm-a [6], który połączył estymację ruchu z trójwymiarowym kodowaniem subpasmowym. W oparciu o tą technikę powstało wiele koderów, jednym z nich jest omawiany IMC-3DEZBC. Ogólna struktura kodera została przedstawiona na rys. 2.



Rys. 2. Schemat blokowy kodera IMC-3DEZBC

Wejściowy strumień video jest rozkładany czasowo przez dwa kanały analizy kompensacji ruchu. W bloku estymacji ruchu wykorzystywane jest hierarchiczne dopasowanie ze zmiennym rozmiarem bloku HVSMB (*ang. hierarchical variable size block matching*). Do filtracji w dziedzinie przestrzennej wykorzystane zostały filtry typu Daubechies 9/7 opisane w [1]. Moduł 3D-EZBC jest rozszerzeniem kodera obrazu EZBC [3] do trzech wymiarów.

W literaturze brak jest szczegółowych informacji na temat możliwości kodera w zależności od różnych ustawień parametrów jego pracy, dla różnych sekwencji użytych do kodowania. Praca [4] przedstawia jedynie porównanie kodera IMC-3DEZBC w stosunku do innych koderów, m.in.: MPEG-2, gdzie okazał się on lepszy. Jednak testy porównawcze wykonywane były przy ściśle określonych parametrach kodowania, dostosowanych do parametrów pracy kodera MPEG-2. Z tego też względu zdecydowano się na przeprowadzenie serii testów, mających pokazać zachowanie się kodeka w różnych trybach jego pracy.

Koder posiada możliwość ustawienia wielu parametrów mających wpływ na przebieg kodowania sekwencji video. W zależności od tych parametrów można uzyskać różną jakość obrazu wyjściowego, okupioną mniejszą lub większą złożonością obliczeniową. Najważniejsze parametry kodera to:

- *accuracy* – parametr określający dokładność kodowania wektorów ruchu (dopuszczalne wartości: 1, 1/2, 1/4, 1/8, 1/16 piksela),
- *tPyrLev* – parametr określający liczbę poziomów czasowej subpasmowej dekompozycji obrazu (dopuszczalne wartości od 1 to 7); na bazie tego parametru wyznaczana jest grupa obrazów  $GOP = 2^{tPyrLev}$ ,
- *framerate* – parametr określający liczbę wejściowych klatek obrazu, przypadających na jedną sekundę.

#### 4. BADANIE KODEKA IMC-3DEZBC

Testowanie pracy kodeka polega na zakodowaniu i zdekodowaniu wybranych sekwencji testowych przy określonych parametrach jego pracy, oraz na wyznaczeniu efektywności kodowania, (np. poprzez wyliczenie współczynnika PSNR) przy określonych przepływnościach strumienia danych. Do testów wybiera się sekwencje powszechnie znane w literaturze światowej, które są stosowane przez innych do porównywania algorytmów przetwarzania obrazów ruchomych. Bazowanie na powszechnie stosowanych sekwencjach testowych, pozwala na stworzenie wspólnej płaszczyzny odniesienia do wyników uzyskanych przez innych autorów.

Dotychczasowe badania prowadzone były z wykorzystaniem pięciu sekwencji testowych, znanych jako: *Mobile*, *Football*, *Garden*, *Tennis*, *Foreman*. Obrazy w rozdzielczości 352x240 przy 30 klatkach na sekundę zapisane są w formacie YUV 4:2:0. Charakteryzują się one różną dynamiką scen, występowaniem bądź brakiem nieruchomego tła. Przykładowo w sekwencji *Mobile* brak jest elementów nieruchomych, na pierwszym planie kolejką popycha piłkę, tło przesuwa się w niezbyt szybkim tempie. Sekwencję *Football* cechuje niemal nieruchome tło, natomiast główny plan przedstawia bardzo dynamiczną scenę na boisku sportowym. W sekwencji *Garden* tło porusza się jednostajnie w umiarkowanie szybkim tempie, przez pierwszych 50 klatek na głównym planie znajduje się drzewo przesuwane szybciej od tła (rys. 3). Wszystkie ww. sekwencje testowano dla następujących wartości przepływności: 384, 512, 768, 1024, 1536, 2048, 4096 kbps.

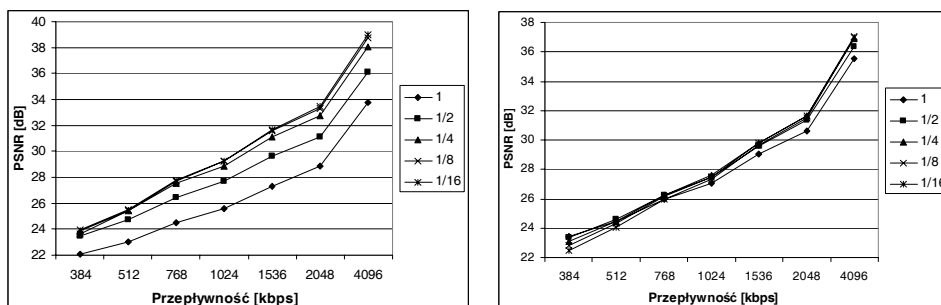


Rys. 3. Obrazy sekwencji testowych, od lewej: *Mobile*, *Football*, *Garden*

Do oceny jakości obrazu posłużono się powszechnie stosowaną w badaniach obrazów miarą jakości obrazu – współczynnikiem PSNR danym wzorem (1).

$$PSNR = 10 \log_{10} \left\{ \frac{255^2}{MSE} \right\} dB \quad (1)$$

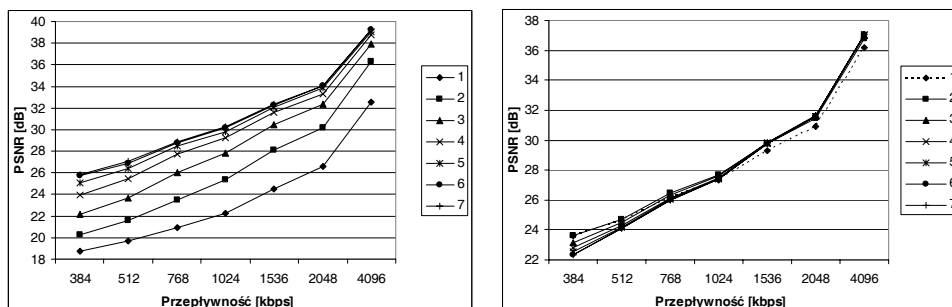
W pierwszej kolejności zbadano wpływ parametru *accuracy* na przebieg kodowania. Parametr ten odpowiada za dokładność wyznaczania wektorów ruchu podczas estymacji ruchu. Na rys. 4 przedstawiono wartości średnie współczynnika PSNR dla składowej Y, przy różnych przepływnościach dla parametru *accuracy* zmieniającego się w zakresie od 1 do 1/16 dla sekwencji *Mobile* oraz *Football*.



Rys. 4. Wartości współczynnika PSNR dla parametru  $accuracy=1$  do  $1/16$ , przy różnych przepływnościach, od lewej sekwencji: Mobile, Football

Test ten pozwolił zauważyć, że wraz ze wzrostem dokładności kodowania wektorów ruchu, wzrasta wartość współczynnika PSNR. Jednak dla niższych przepływności, przy dokładności  $1/16$ , można zauważyć gorsze wyniki w stosunku do dokładności  $1/8$  o około  $0,30$  dB.

Kolejnym etapem badań kodera było określenie wpływu zmian wartości parametru  $tPyrLev$  na przebieg kodowania. Parametr ten określa liczbę poziomów dekompozycji obrazu, jednocześnie wartość ta oddziałuje wprost na rozmiar grupy obrazów, zgodnie z zasadą podaną wcześniej.



Rys. 5. Wartości współczynnika PSNR dla parametru  $tPyrLev=1$  do  $7$ , przy różnych przepływnościach, od lewej sekwencji: Mobile, Football

Na rys. 5 przedstawiono wartości współczynnika PSNR przy różnych przepływnościach dla parametru  $tPyrLev$  w zakresie od  $1$  do  $7$  dla sekwencji *Mobile* oraz *Football*. Wraz ze wzrostem wartości parametru  $tPyrLev$ , rośnie jakość obrazu. Różnice między poziomem  $6$  i  $7$  badanego współczynnika są jednak niewielkie (od  $0,04$  do  $0,13$  dB dla sekwencji *Mobile*). Podobne, choć w mniejszym stopniu zależności można zauważyć dla pozostałych badanych obrazów, oraz w najmniejszym stopniu dla sekwencji *Football* (rys. 5).

Jak nietrudno zauważyć z przedstawionych dotychczas badań, z punktu widzenia jakości obrazu, najlepsze wyniki można uzyskać dla parametrów:  $tPyrLev=7$  oraz  $accuracy=1/16$ . Taka konfiguracja pociąga za jednak za sobą bardzo długie czasy kodowania. Należy mieć na

uwadze, że testowano sekwencje czterosekundowe, najkrótszy czas kodowania wyniósł 12 minut, najdłuższy zaś ponad 7 godzin. Złożoność obliczeniowa algorytmu kodowania jest duża i przyczynia się do znacznego utrudnienia w prowadzeniu badań. Testy przeprowadzono przy użyciu komputera z procesorem Intel Celeron 450MHz i 512MB RAM.

## 5. PODSUMOWANIE

Najbardziej dynamiczną testowaną sekwencją jest *Football*. W przypadku tego też obrazu wpływ badanych parametrów na wartość współczynnika PSNR jest najmniejszy (rys. 4, rys. 5). Nieco mniej dynamiczny obraz *Foreman*, również wykazuje mniejszą czułość na zmiany parametrów kodowania niż zdecydowanie mniej dynamiczna sekwencja *Mobile*. Stąd można wyciągnąć wniosek, że warto kodować sekwencje dynamiczne przy niższych wartościach parametrów *accuracy* oraz *tPyrLev*, zaś mniej dynamiczne dla większych wartości parametrów. Określenie dynamiki sekwencji nie jest sprawą łatwą i wymaga kodowania dwuprzebiegowego. Pierwszy przebieg służy do określenia dynamiki zadanej grupy obrazów. Kodowanie w drugim przebiegu polega na adaptacyjnym doborze parametrów pracy koderza, na podstawie informacji z pierwszego przebiegu.

Otrzymane wyniki testów koderza stanowią podstawę do dalszych badań. W pierwszej kolejności podjęte zostaną próby zoptymalizowania prędkości działania modułu odpowiedzialnego za kodowanie sekwencji. Obecnie długi czas potrzebny na wykonanie jednego testu, jest bardzo dużym utrudnieniem przy weryfikacji nowych algorytmów przetwarzania wprowadzonych w koderze.

## LITERATURA

- [1] A. Antonini, M. Barlaud, I. Daubechies: *Image coding using wavelet transform*, IEEE Transaction Image Processing, 1, 1992, 205-220
- [2] M. Domański: *Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych*, Poznań, WPP, 1998
- [3] S.-T. Hsiang, J.W. Woods: *Embedded image coding using zeroblocks of subband/wavelet coefficients and context modelling*, IEEE International Symposium on Circuits and Systems, 3, 2000, 662-665
- [4] S.-T. Hsiang, J.W. Woods: *Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank*, Signal Processing: Image Communication 16, 2001, 705-724
- [5] ISO/IEC International Standard 13818: *Information Technology – Generic Coding of Moving Pictures and Associated Audio Information*, 1994
- [6] J.-R. Ohm: *Three-dimensional subband coding with motion compensation*, IEEE Transactions Image Processing, 3, 1994, 559-571