

Andrzej POPŁAWSKI

UNIwersytet Zielonogórski, Instytut Informatyki i Elektroniki

Ograniczenie opóźnienia kodowania w trójwymiarowych falkowych koderach sekwencji wizyjnych

Dr inż. Andrzej POPŁAWSKI

Tytuł doktora otrzymał na Politechnice Poznańskiej w roku 2006, autor lub współautor prac z zakresu kompresji sekwencji wizyjnych. Pracuje na stanowisku adiunkta w Instytucie Informatyki i Elektroniki Uniwersytetu Zielonogórskiego. Głównie zainteresowania naukowe dotyczą zagadnień trójwymiarowej falkowej kompresji sekwencji wizyjnych dla małych opóźnień kodowania.



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

Streszczenie

Analiza falkowa z kompensacją ruchu nabrała w ostatnich latach dużego znaczenia za względu na liczne zastosowania, między innymi w koderach skalowalnych. Skalowalność jest ważną cechą wykorzystywaną podczas transmisji obrazów ruchomych poprzez sieci telekomunikacyjne o niejednorodnej strukturze. Analiza falkowa w dziedzinie czasu wprowadza opóźnienie kodowania, które wyklucza kodery falkowe z wielu zastosowań. Przedstawione w pracy wyniki badań eksperymentalnych przedstawiają możliwości ograniczenia opóźnienia kodowania poprzez redukcję rozmiaru grupy obrazów.

Słowa kluczowe: analiza falkowa, kompresja sekwencji wizyjnych z małym opóźnieniem kodowania, falkowa filtracja w dziedzinie czasu, drabinkowe struktury filtracji.

Constraint of coding delay in three-dimensional wavelet video coders

Abstract

The significance of the wavelet analysis with motion compensation, due to numerous applications in the area of scalable video coders, has grown significantly in importance in the recent years. The scalability of the coders is a fundamental feature while deploying within the heterogeneous networks. Unfortunately the time domain wavelet processing is coupled with the delay, which efficiently eliminates its use in many applications. The paper presents possibilities of limitation of coding delay by decreasing size of group of pictures.

Keywords: wavelet analysis, low delay video coding, wavelet temporal filtering, lifting.

1. Wstęp

W współczesnych koderach sekwencji wizyjnych powszechnie wykorzystuje się techniki kodowania hybrydowego, dla których powstało wiele standardów, np.: MPEG-2, MPEG-4, H.263, H.264. W procesie kodowania hybrydowego wykorzystywana jest dyskretna transformacja kosinusowa (DCT – *discrete cosine transform*) [1]. W ostatnich latach zainteresowanie wielu badaczy zostało skierowane na kodery wykorzystujące dyskretną transformację falkową (DWT – *discrete wavelet transform*) [22, 23, 24]. Warto podkreślić, że o ile analiza falkowa w dziedzinie przestrzennej posiada długą historię, o tyle analiza falkowa w dziedzinie czasu dopiero za sprawą Ohma [13], który zaproponował efektywną analizę falkową z kompensacją ruchu stała się przedmiotem szerokiego zainteresowania. Późniejsze prace [4, 5, 17, 20, 25, 26] dotyczyły wykorzystania drabinkowych struktur filtracji czasowej (*lifting*), zwiększenia dokładności kodowania wektorów ruchu z 1 do 1/16 odstępu próbkowania oraz zastąpienie krótkich filtrów Haara dłuższymi filtrami LeGalla 5/3.

Współczesne kodery falkowe to przede wszystkim kodery trójwymiarowe, wykorzystujące analizę falkową (*wavelet analysis*) w trzech wymiarach (dwóch przestrzennych i jednym czasowym). Kodery falkowe stanowią interesującą alternatywę w stosunku do

klasycznych koderów hybrydowych, gdyż umożliwiają w naturalny sposób uzyskanie pełnej skalowalności (czasowej, przestrzennej oraz typu SNR), dzięki czemu możliwa jest regulacja prędkości transmisji [2]. Ponadto efektywność kodowania trójwymiarowych kodeków falkowych jest zbliżona do efektywności kodowania kodeków hybrydowych [5, 19].

Skalowalność jest istotna w wielu zastosowaniach np. w systemach bezprzewodowych o zmieniającej się w czasie przepustowości kanału transmisyjnego, w transmisji sekwencji wizyjnych poprzez sieci komunikacyjne o niejednorodnej strukturze [21, 25]. Niestety poważną wadą koderów falkowych jest powstawanie dużego opóźnienia kodowania spowodowanego wykonywaniem analizy falkowej w dziedzinie czasu.

Opóźnienie kodowania związane z wykonywaniem analizy falkowej w dziedzinie czasu jest tym większe, im większy jest rząd wykorzystywanych filtrów, zwiększa się również znacznie wraz ze wzrostem liczby poziomów analizy czasowej. Opóźnienie powstaje w wyniku konieczności przetwarzania jednocześnie pewnej liczby obrazów sekwencji wizyjnej, zwanej grupą obrazów (GOP – *Group Of Pictures*). Ta grupa obrazów może obejmować np. szesnaście obrazów. W celu przetworzenia tych obrazów, koder musi zgromadzić wszystkie obrazy grupy, co powoduje powstanie istotnego opóźnienia.

W wielu zastosowaniach jak np.: wideotelefony, systemy wideokonferencyjne, systemy zdalnej obserwacji i nadzoru wizyjnego, wymagane jest niewielkie opóźnienie kodowania. Przykładowo zalecenie IUT–T H.261 [9] określa wartość opóźnienia kodowania dla usługi wideokonferencji na poziomie nie większym niż 150 ms. To ograniczenie wymusza stosowanie bądź to innych technik kompresji sekwencji wizyjnych, bądź też modyfikacje istniejących algorytmów falkowej analizy w dziedzinie czasu.

Opisane w najnowszej literaturze techniki filtracji z kompensacją ruchu, poprawiając efektywność kompresji, przyczyniają się do zwiększenia opóźnienia kodowania związanego z wykorzystaniem dłuższych filtrów oraz z wykonywaniem wielu (zazwyczaj pięciu) poziomów analizy falkowej w czasie.

Artykuł przedstawia wyniki badań eksperymentalnych efektywności kompresji w zależności od liczby poziomów analizy falkowej w dziedzinie czasu dla dłuższych filtrów LeGalla 5/3 [10]. Badania przeprowadzono dla reprezentatywnej grupy testowych sekwencji wizyjnych, szeroko stosowanych i akceptowanych przez innych badaczy.

2. Analiza falkowa w dziedzinie czasu

W falkowych koderach sekwencji wizyjnych, obrazy wejściowe w pierwszej kolejności są grupowane a następnie wspólnie przetwarzane. Rozmiar grupy obrazów (GOP) ustalany jest na podstawie liczby poziomów analizy falkowej w dziedzinie czasu i wpływa bezpośrednio na wielkość opóźnienia kodowania. Wielkość grupy obrazów wynosi 2^K , gdzie K jest liczbą poziomów analizy falkowej w czasie. Korzystając z drabinkowych układów filtracji [17], które są obecnie powszechnie wykorzystywane w koderach falkowych, równania analizy dla filtrów LeGalla 5/3 [10] można zapisać w postaci (patrz rys. 1):

$$\begin{aligned} h_t &= x_{2t+1} - \frac{1}{2}(MC_t(x_{2t}) + MC_t(x_{2t+2})), \\ l_t &= x_{2t} + \frac{1}{4}(MC_{t-1}^{-1}(x_{2t+2}) + MC_{t-1}^{-1}(x_{2t})), \end{aligned} \quad (1)$$

gdzie:

x_t – obraz wejściowy w czasie t ,

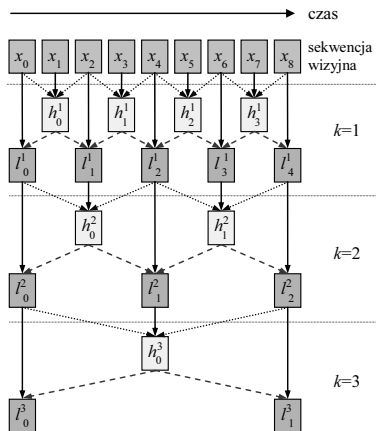
h_t – próbka sygnału subpasma wysokich częstotliwości w czasie t ,

l_t – próbka sygnału subpasma wysokich częstotliwości w czasie t ,

$MC_i(x_{2i})$ – operator kompensacji ruchu między obrazami x_{2i+1} i x_{2i} ,

$MC_i(x_{2i+2})$ – operator kompensacji ruchu między obrazami x_{2i+1} i x_{2i+2} .

Syntezę wykonuje się poprzez wykonanie tych samych operacji, lecz w odwrotnej kolejności, przy jednoczesnej zmianie znaków operacji na przeciwnie. Drabinkowe struktury filtracji zapewniają idealną rekonstrukcję, niezależnie od stosowanych filtrów analizy.



Rys. 1. Trzy poziomy analizy falkowej w czasie w oparciu o filtry LeGalla 5/3
Fig. 1. Three levels of temporal wavelet analysis based on LeGall 5/3 filters

Opisane w literaturze techniki redukcji opóźnienia kodowania skupiają się głównie na eliminacji kroków predykcji lub uaktualnienia w drabinkowych strukturach filtracji w dziedzinie czasu, dla pięciu (lub czterech) poziomów analizy czasowej. W pracach [3, 18] badano efektywność kompresji dla zredukowanego ograniczenia kodowania uzyskanego poprzez ograniczenie liczby poziomów analizy czasowej dla krótkich filtrów Haara. Brak jest natomiast takiego zestawienia dla filtrów LeGalla 5/3. Celem pracy jest uzupełnienie tego braku. Ponadto badania miały dać odpowiedź, czy możliwe jest uzyskanie dużej efektywności kodowania przy obniżonej wartości opóźnienia kodowania osiągniętej poprzez zmianę rozmiaru GOP.

W pracy nie uwzględnia się opóźnienia kodowania wynikającego z czasu potrzebnego na przetworzenie danych przez jednostkę przetwarzającą oraz wynikające z czasu potrzebnego na przesłanie danych do odbiorcy. Skupiono się na opóźnieniach powstających w zespołach filtrów analizy/syntezy. Dla filtrów 5/3 opóźnienie kodowania D , wyrażone w liczbie obrazów (odstępów próbkowania w dziedzinie czasu) dane jest równaniem [12]:

$$D = 3 \cdot (2^K - 1), \quad (2)$$

gdzie K oznacza liczbę poziomów analizy falkowej w dziedzinie czasu.

W tab. 1 przedstawiono wartości opóźnienia kodowania w zależności od liczby poziomów falkowej analizy w dziedzinie czasu dla filtrów LeGalla 5/3.

Tab. 1. Wartości opóźnienia kodowania w zależności od rozmiaru grupy obrazów
Tab. 1. Coding delay values depending on size of group of pictures

Liczba poziomów falkowej analizy w dziedzinie czasu (K)	1	2	3	4	5
Rozmiar grupy obrazów (GOP)	2	4	8	16	32
Opóźnienie D [liczba obrazów]	3	9	21	45	93

3. Opis badań eksperymentalnych

Do badań wykorzystano zestaw dziewięciu sekwencji wizyjnych (*City*, *Crew*, *Harbour*, *Ice*, *Soccer*, *Football*, *Silent*, *Mobile*, *Foreman*) o rozdzielczości przestrzennej CIF (352x288) i częstotliwości występowania obrazów równej 30Hz. Sekwencje te są szeroko

akceptowane i stosowane przez innych badaczy do oceny efektywności kompresji sekwencji wizyjnych, charakteryzują się zróżnicowaną dynamiką, zarówno na pierwszym jak i na drugim planie.

Eksperyment polegał na zakodowaniu a następnie zdekodowaniu testowych sekwencji dla założonych wielkości grupy obrazów (a tym samym wartości opóźnienia kodowania) przy ustalonej prędkości transmisji i ocenie jakości tak zdekodowanych sekwencji. Wykorzystano sekwencje testowe zapisane w formacie YUV, reprezentującym składową luminancji (Y) oraz dwie składowe chrominancji (UV) w schemacie próbkowania 4:2:0 [6].

Jako miarę jakości wykorzystano impulsowy współczynnik sygnału do szumu PSNR (*Peak Signal-to-Noise Ratio*) zdefiniowany równaniem [2, 14]:

$$PSNR = 10 \log \frac{255^2 N^2}{\sum_i e_i^2} \quad (3)$$

gdzie:

255 – zakres dynamiczny sygnału,

N – liczba punktów w obrazie,

e_i – różnica wartości między i -tym punktem obrazu oryginalnego i zdekodowanego.

Ponieważ przeprowadzenie oceny jakości sekwencji wizyjnych metodami subiektywnymi [7] jest żmudne i czasochłonne, zrezygnowano z tej metody pomiarowej. Ze względu na rodzaj badań (wykorzystano jedną wersję kodeka we wszystkich badaniach), charakter wprowadzanych zniekształceń jest zbliżony, zaś zalecenie [8] dopuszcza stosowanie współczynnika PSNR do porównywania kodowania o zbliżonych metodach kompresji. Pomiaru dokonywano dla składowej luminancji, wyznaczając średnią arytmetyczną współczynnika PSNR wszystkich obrazów sekwencji. Taka metoda oceny efektywności kodowania sekwencji wizyjnych jest powszechnie stosowana przez wielu badaczy [14, 15, 16].

Ponieważ jak dotąd nie opracowano standardu dla kodeków falkowych, dlatego istotnym zagadnieniem jest wybór odpowiedniego kodeka do przeprowadzenia badań. W pracy wykorzystano kodek MC-EZBC (*Motion Compensation Embedded Zero Block Coding*) [5, 19]. Kodek ten jest powszechnie uznany i wykorzystywany w wielu pracach jako falkowy koder referencyjny.

W tab. 2 zaprezentowano wyniki przeprowadzonych badań eksperymentalnych dla dziewięciu testowych sekwencji wizyjnych. Wizualizacja tych wyników została przedstawiona na rys. 2. Dla większości badanych sekwencji GOP równy 32 jest wartością najkorzystniejszą, dla części sekwencji dla szesnastu obrazów w grupie uzyskano największe wartości współczynnika PSNR. Jedyne dla dwóch sekwencji (*Ice*, *Football*) grupa składająca się z ośmiu obrazów zapewnia największe wartości współczynnika PSNR. W tab. 3 przedstawiono optymalny rozmiar grupy obrazów dla poszczególnych sekwencji wizyjnych ze względu na jakość kompresji mierzonej wartością współczynnika PSNR.

Tab. 2. Wartości współczynnika PSNR w zależności od rozmiaru grupy obrazów, prędkość transmisji 1024kbps

Tab. 2. PSNR values depending on size of group of pictures, bitrate 1024kbps

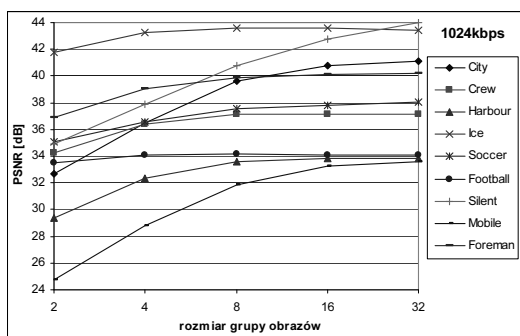
Sekwencja	Rozmiar grupy obrazów / PSNR [dB]				
	2	4	8	16	32
<i>City</i>	31,24	34,66	37,72	39,42	39,82
<i>Crew</i>	33,02	34,92	35,82	35,83	35,84
<i>Harbour</i>	27,85	30,57	32,14	32,65	32,65
<i>Ice</i>	39,59	41,30	41,64	41,59	41,38
<i>Soccer</i>	33,53	35,02	35,83	36,10	36,28
<i>Football</i>	32,09	32,69	32,72	32,68	32,70
<i>Silent</i>	33,50	36,28	38,81	41,07	42,05
<i>Mobile</i>	23,51	26,69	30,13	31,92	32,53
<i>Foreman</i>	35,28	37,54	38,61	38,95	39,10

Warto jednak zwrócić uwagę na stosunkowo niewielkie różnice w wartości współczynnika PSNR dla grupy składającej się z ośmiu obrazów w stosunku do rozwiązania optymalnego (sekwencje *Crew*, *Harbour*, *Soccer*, *Foreman*). Innymi słowy, jedynie dla trzech testowych sekwencji wizyjnych (*Silent*, *Mobile*, *City*) wraz ze wzrostem rozmiaru grupy obrazów zanotowano wyraźny wzrost wartości współczynnika PSNR. Omówione zależ-

ności występują niezależnie od badanej prędkości transmisji. Zarówno przy prędkości transmisji 1024kbps (rys. 2) jak i 512kbps (rys. 3) można zaobserwować podobne tendencje.

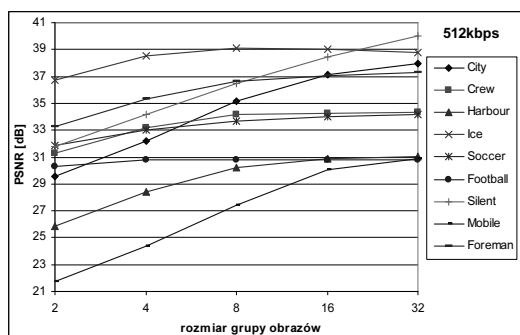
Tab. 3. Optymalny rozmiar grupy obrazów dla poszczególnych sekwencji
Tab. 3. Optimal size of GOP for given sequences

Sekwencja	Rozmiar grupy obrazów	Opóźnienie [liczba obrazów]	PSNR [dB]
City	32	93	39,82
Crew	16	45	35,83
Harbour	16	45	32,65
Ice	8	21	41,64
Soccer	32	93	36,28
Football	8	21	32,72
Silent	32	93	42,05
Mobile	32	93	32,53
Foreman	32	93	39,10



Rys. 2. Efektywność kodowania w zależności od rozmiaru grupy obrazów, prędkość transmisji 1024kbps

Fig. 2. Coding efficiency for various size of group of pictures, bitrate 1024kbps



Rys. 3. Efektywność kodowania w zależności od rozmiaru grupy obrazów, prędkość transmisji 512kbps

Fig. 3. Coding efficiency for various size of group of pictures, bitrate 512kbps

Na podkreślenie zasługuje również fakt, że im mniejsza jest liczba poziomów falkowej analizy w dziedzinie czasu tym mniejsza jest wartość nakładu obliczeniowego. Wynika to z tego, że na każdym kolejnym poziomie analizy wyznaczane są wektory ruchu, których uzyskanie zajmuje około 70% ogółu czasu potrzebnego na zakodowanie sekwencji wizyjnej.

4. Podsumowanie

W pracy zaprezentowano wyniki badań eksperymentalnych, które pokazują, że możliwe jest ograniczenie opóźnienia kodowania w falkowych koderach sekwencji wizyjnych poprzez ograniczenie liczby poziomów falkowej analizy w dziedzinie czasu. W większości badanych przypadków możliwe jest ograniczenie tego opóźnienia z 93 do 21 obrazów przy niewielkim spadku efektywności kodowania. Badania pokazują, że korzystne jest stosowanie dużego rozmiaru grupy obrazów jedynie w przypadku sekwencji wizyjnych mało dynamicznych (*City*, *Mobile*, *Silent*).

W przypadku pozostałych sekwencji testowych, ograniczenie rozmiaru grupy do ośmiu obrazów nie wpłynęło znacząco na spadek efektywności kompresji. Niestety, opóźnienie o wartości 21 obrazów to wciąż zbyt dużo w wielu zastosowaniach, dlatego też opisana metoda redukcji opóźnienia kodowania nie jest uniwersalna i nie może być stosowana jako jedyna forma redukcji opóźnienia kodowania.

5. Literatura

- [1] Ahmed N., Natarajan T., Rao K.R.: Discrete cosine transform. – IEEE Trans. on Computer, str. 90–93, 1974
- [2] Domański M.: Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych, Poznań, WPP, 1998.
- [3] Huang L. M., Mei S. S., Honda Y.: Results on Scalable Video Coding in Low Delay Mode. – ISO/IEC JTC1/SC29/WG11, Doc.M9843, MPEG 2003
- [4] Hsiang S.-T., Woods J. W.: Embedded image coding using zeroblocks of subband/wavelet coefficients and contest modelling. – IEEE Int. Symp. on Circuits and Systems, Vol. 3, str. 662–665, 2000.
- [5] Hsiang S.-T., Woods J. W.: Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank. – Signal Processing: Image Communication, Vol. 16, str. 705–724, 2001.
- [6] ITU-R Rec. BT.470-3, Television systems, 1994.
- [7] ITU-R Rec. BT.500-6, Methodology for the subjective assessment of the quality of television pictures, 1994.
- [8] ITU-R Rec. BT.813, Methods for objective picture quality assessment in relation to impairments from digital coding of television signal, 1992.
- [9] ITU-T Rec. H.261, Video codec for audiovisual services at p_z64 kbit/s.
- [10] LeGall D., Tabatabai A.: Subband coding of digital images using symmetric short kernel filters and arithmetic coding techniques. – Proc. Int. Conf. on Acoustics, Speech and Signal Processing, IEEE, str. 761–765, 1988.
- [11] Jayant N., Noll P.: Digital Coding of Waveforms. – Prentice-Hall, 1984
- [12] Ohm J.-R.: Complexity and delay analysis of MCTF interframe wavelet structures, doc. m8520, Klagenfurt MPEG meeting, July 2002.
- [13] Ohm J.-R.: Three-dimensional subband coding with motion compensation, IEEE Transactions on Image Processing, vol. 3, 559-589, 1994.
- [14] Pau G., Pesquet-Popescu B., van der Schaar M., Viéron J.: Delay-performance trade offs in motion-compensated scalable subband video compression, doc. m11084, Redmond MPEG meeting, July 2004.
- [15] Pau G., Pesquet-Popescu B., van der Schaar M., Viéron J.: Delay-Performance Trade-Offs in Motion-Compensated Scalable Subband Video Compression, Proc. of Advanced Concepts for Intelligent Vision Systems (ACIVS), 2004.
- [16] Pau G., Viéron J., Pesquet-Popescu B.: Video Coding with Flexible MCTF Structures for Low End-to-End Delay, ICIP 2005, Genova, vol. 3, 241-244.
- [17] Pesquet-Popescu B., Bottreau V.: Three-dimensional lifting schemes for motion compensated video compression, Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal, 2001.
- [18] Popławski A.: Trójwymiarowa falkowa kompresja sekwencji wizyjnych z małym opóźnieniem kodowania. – Rozprawa doktorska, Wydział Elektryczny, Politechnika Poznańska, Poznań, 2006.
- [19] Ruser T., Hanke K., Wien M.: Optimization for locally adaptive MCTF based on 5/3 lifting. – Proc. Picture Coding Symposium, San Francisco, CA, USA. str. 210–220, 2004.
- [20] Secker A., Taubman D.: Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting, Proc. of the IEEE Int. Conf. on Image Processing, 1029-1032, 2001.
- [21] Sheldon T.: Wielka encyklopedia sieci komputerowych, Robomatic, Wrocław, 1999.
- [22] Vaidyanathan P.P.: Multirate Systems and Filter Banks, Englewood Cliffs, NJ: Prentice Hall, 1993.
- [23] Vetterli M., Kovačević J.: Wavelets and Subband Coding, Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [24] Woods J. W.: Subband Image Coding. – Kluwer Academic Publisher, 1991.
- [25] Wu D., Hou Y., Zhang Y.: Scalable video coding and transport over broad-band wireless networks, Proc. of the IEEE, vol. 89, 6-20, 2001.
- [26] Zhan Y., Picard M., Pesquet-Popescu B., Heijmans H., Long temporal filters in lifting schemes for scalable video coding, doc. m8680, Klagenfurt MPEG meeting, July 2002.