

KOMPRESJA SEKWENCJI WIZYJNYCH Z MAŁYM OPÓŹNIENIEM KODOWANIA

Andrzej Popławski

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

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

STRESZCZENIE

Artykuł przedstawia wyniki badań eksperymentalnych programowego, skalowalnego kodeka sekwencji wizyjnych, opartego na trójwymiarowym kodowaniu falkowym. Badany jest wpływ ograniczenia opóźnienia kodowania na efektywność kompresji sekwencji wizyjnych. Poprzez zmianę liczby obrazów w grupie uzyskiwano zadane opóźnienie kodowania. Badania przeprowadzono dla grupy obrazów składającej się z 1, 2, 4, 8, 16, 32 ramek sekwencji.

1. WPROWADZENIE

Jedną z istotnych cech opisanych w literaturze trójwymiarowych falkowych koderów wizyjnych wykorzystujących filtrację z kompensacją ruchu jest skalowalność przestrzenna i czasowa (typu SNR) [3, 4, 15]. Funkcja skalowalności 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 (tzw. sieci heterogeniczne). Małe opóźnienie kodowania sekwencji wizyjnych odgrywa z kolei bardzo ważną rolę w wielu zastosowaniach gdzie wymagana jest interakcja, jak na przykład: wideotelefon, systemy wideokonferencyjne, systemy zdalnej obserwacji i nadzoru wizyjnego [11, 12, 13].

W ostatnich latach opracowano wiele standardów kompresji sekwencji wizyjnych. Zalecenie IUT-T H.261 [5] znajduje zastosowania w wideotelefonii oraz systemach wideokonferencyjnych, gwarantuje wartość opóźnienia kodowania na poziomie nie większym niż 150 ms. Standard IUT-T H.263 [6] stanowiący rozwinięcie H.261 opracowano na potrzeby kodowania cyfrowych sygnałów wizyjnych w systemach wideotelefonicznych, zalecenie H.264/AVC [7] umożliwia kompresję o dużej efektywności. Wymienione wyżej zalecenia wykorzystują transformację kosinusową w procesie przetwarzania obrazów, ponadto nie zapewniają skalowalności.

Ważną techniką pozwalającą na uzyskanie efektywnej kompresji sekwencji wizyjnych jest stosowanie filtracji obrazu w dziedzinie czasu wykorzystujące transformatę falkową [1, 2, 3,

8, 9, 10, 14]. Dotychczas zaproponowane techniki kodowania trójwymiarowego wprowadzają w procesie przetwarzania obrazów duże opóźnienia czasowe, wynikające z przetwarzania jednocześnie grupy obrazów składającej się zazwyczaj z 16 ramek.

Przeprowadzone badania eksperymentalne pokazały, że w przypadku testowych sekwencji wizyjnych o umiarkowanej dynamice ruchu, efektywność kompresji rośnie wraz ze wzrostem liczby obrazów w grupie. Natomiast w przypadku sekwencji o dużej dynamice ruchów wpływ rozmiaru grupy obrazów na efektywność kompresji jest niewielka, niezależnie od badanej rozdzielczości przestrzennej i przepływności strumienia danych.

2. OPIS BADANEGO KODEKA

Do badań wykorzystano kodek IMC-3DEZBC (*Invertible Motion Compensated 3-Dimensional Embedded Zero-Block coding of wavelet coefficients and Context modeling*) autorstwa S.-T. Hsiang-a oraz J.W. Woods-a, opisany w [4]. Jest to trójwymiarowy kodek falkowy wykorzystujący kompensację ruchu, ze skalowalnością typu SNR. Kodek ten charakteryzuje się następującymi właściwościami:

- czteropoziomową dekompozycją przestrzenną obrazu,
- trójwymiarowym kodowaniem subpasmowym z kompensacją ruchu,
- użyciem filtrów „9/7” do analizy/syntezy w dziedzinie filtracji przestrzennej,
- modułem estymacji ruchu z hierarchicznym dopasowaniem (zmienny rozmiar bloku),
- niezależnym kodowaniem entropowym każdej grupy obrazów z wykorzystaniem kontekstowego kodowania arytmetycznego.

Opóźnienie dekodowania wynosi $N-1$ obrazów (N – liczba obrazów w grupie), zaś całkowite opóźnienie (*end-to-end delay*) wynosi $2N-1$.

Badania przeprowadzono mierząc wartości współczynnika PSNR w zależności od parametru TN (*temporal number*). Parametr ten określa liczbę poziomów czasowej subpasmowej dekompozycji obrazu; na bazie tego parametru wyznaczana jest grupa obrazów $GOP = 2^{\text{temporal_number}}$. Wartość zerowa parametru oznacza, że wykonywane jest jedynie kodowanie wewnątrzbrazowe, tzn. grupa obrazów składa się z jednej ramki.

3. BADANE SEKWENCJE WIZYJNE

Badania przeprowadzono na kilku wybranych sekwencjach testowych, charakteryzujących się różną dynamiką scen, zarówno na pierwszym jak i na drugim planie (tło). Omawiane sekwencje są szeroko stosowane przez innych do badania efektywności koderów wizyjnych. Sekwencja City (rys. 1 a)) przedstawia widok miasta z poziomu lotu ptaka. Na pierwszym

planie widoczny jest szczyt jednego z budynków, drugi plan zaś przedstawia inne budynki. Dynamika sekwencji jest stosunkowo niewielka, ruch kamery jest dość płynny, brak jest w obrazie elementów nieruchomych.



Rys. 1. Obrazy sekwencji (od lewej): a) City, b) Crew, c) Harbour

Sekwencja Crew (rys. 1 b)) przedstawia grupę kilku osób ubranych w pomarańczowe kombinezony idące wolnym krokiem. Tło sekwencji porusza się w początkowej fazie bardzo nieznacznie, zaś pod koniec ruch kamery jest już znaczny. Związane jest to ze zmianą kierunku marszu osób w kolejnych obrazach. Dynamika sekwencji jest nieco większa niż w przypadku obrazu City. Sekwencja Harbour (rys. 1c)) przedstawia widok portu, w który przycumowane są żagłówki. Kamera zamocowana prawdopodobnie na statywie, wykonuje jedynie niewielkie ruchy na boki, które mogą być spowodowane małą sztywnością statywu. Większą część tła stanowi tafla wody, po której przepływają żagłówki, ogólna dynamika sekwencji jest stosunkowo niewielka.



Rys. 2. Obrazy sekwencji (od lewej): a) Ice, b) Soccer

Sekwencja Ice (rys. 2 a)) przedstawia widok lodowiska, na którym kilkanaście osób jeździ na łyżwach. Tło sekwencji jest nieruchome, z lodowiska korzystają osoby zarówno stojące, jak i szybko poruszające się po jego płycie. W sekwencji występują zarówno elementy nieruchome, jak i elementy szybko poruszające się. Sekwencja Soccer (rys 2. b)) przedstawia widok boiska, na którym kilka osób gra w piłkę nożną. Kamera podąża za piłką, momentami tło jest niemal nieruchome, czasami zaś ruch kamery jest bardzo dynamiczny. Ogólnie w sekwencji występują zarówno momenty o niewielkiej dynamice, jak i elementy o dużej dynamice ruchu. Sekwencja Flat jest specjalnie spreparowaną sekwencją, w której wszystkie

klatki są jednakowe, nie występuje w niej żaden ruch. Sekwencja powstała przez powielenie pierwszej klatki sekwencji City.

Wartości współczynnika PSNR dla jednej ramki obrazu wyliczone zostały ze wzoru (1).

$$PSNR = 20 \log_{10} \left(\frac{255}{RMSE} \right), \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}, \quad (1)$$

gdzie: N – liczba punktów obrazu,

x_i – oryginalny element obrazu,

y_i – odtworzony element obrazu.

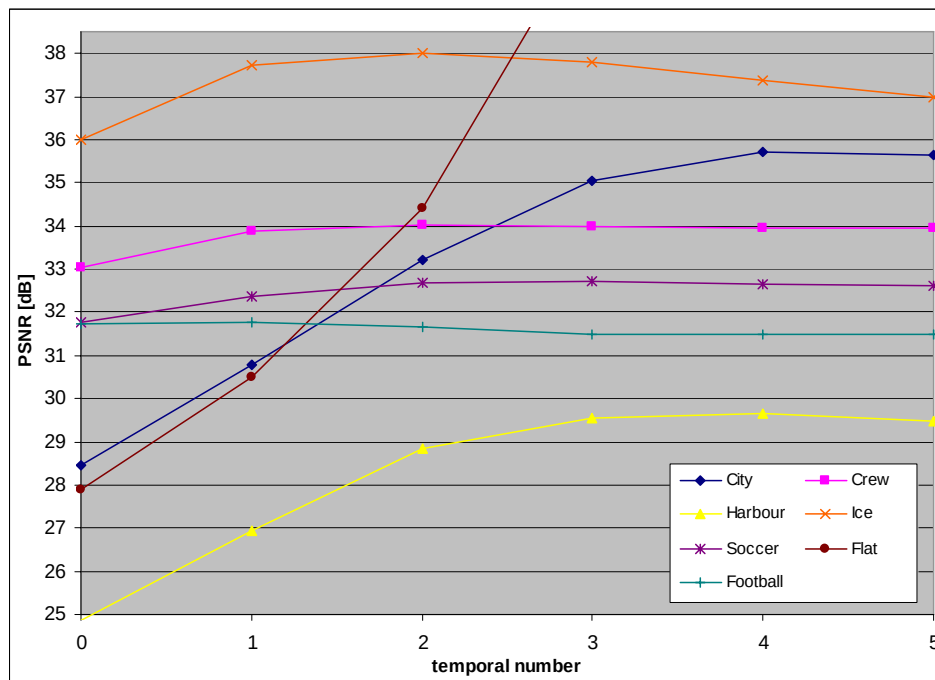
Średnia wartość PSNR wszystkich zdekodowanych obrazów, stanowi miarę jakości wyjściowej sekwencji wizyjnej.

4. WYNIKI BADAŃ

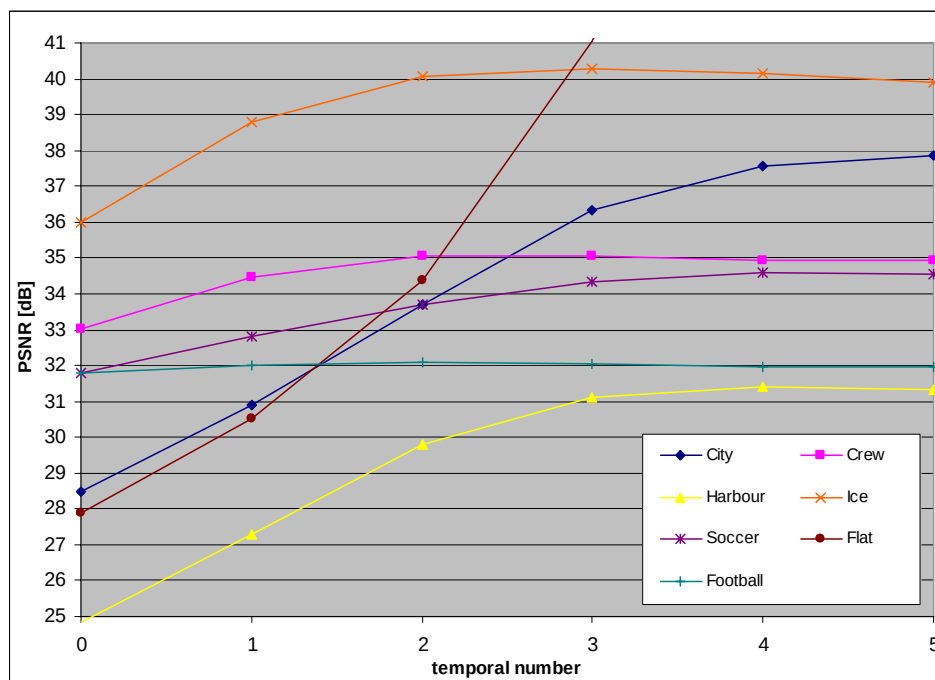
Rozmiar grupy obrazów wpływa bezpośrednio na wielkość opóźnienia pojawiającego się w czasie kodowania / dekodowania sekwencji. W klasycznych koderach hybrydowych [6, 7] większy rozmiar *GOP* wpływa korzystnie na podniesienie wartości współczynnika PSNR dekodowanych sekwencji wizyjnych, z drugiej jednak strony powoduje powstawanie tym większych opóźnień, im większa jest grupa obrazów. Pomiary miały dać odpowiedź na pytanie, o ile zwiększa się efektywność kodowania przy wzrastającej grupie obrazów. Tabela 1 przedstawia wartości przebadanych przepływności dla omówionych powyżej testowych sekwencji wizyjnych. Badania przeprowadzono dla trzech rozdzielczości przestrzennych (176x144, 352x288, 704x576) oraz dwóch rozdzielczości czasowych (15Hz i 30Hz).

Tab. 1. Wartości badanych przepływności dla zadanych rozdzielczości przestrzennych i czasowych testowych sekwencji wizyjnych

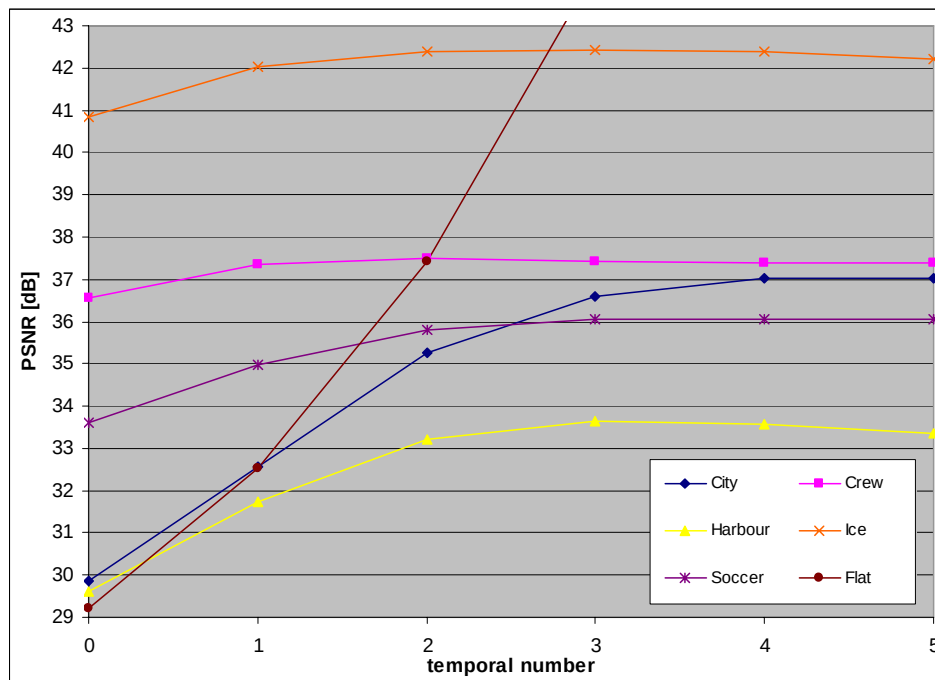
	QCIF, 15Hz (176x144)	CIF, 15Hz (352x288)	CIF, 30Hz (352x288)	4CIF, 30Hz (704x576)
C i t y	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
C r e w	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
H a r b o u r	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
I c e	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
S o c c e r	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
F l a t	128, 192	256, 384, 512	512, 768, 1024	1536, 3072
F o o t b a l l	128, 192	256, 384, 512	512, 768, 1024	X



Rys. 3. Wartości współczynnika PSNR dla rozdzielczości CIF, 15Hz, przepływność 384 kbps



Rys. 4. Wartości współczynnika PSNR dla rozdzielczości CIF, 30Hz, przepływność 768 kbps



Rys. 5. Wartości współczynnika PSNR dla rozdzielczości 4CIF, 30Hz, przepływność 3072 kbps

Wnioski

Na podstawie analizy wykresów przedstawionych na rys. 3, rys. 4, rys. 5 można wyciągnąć następujące wnioski:

- największą czułość na zmianę parametru TN (*temporal number*) wykazuje sekwencja Flat oraz City, co oznacza, że moduł kompensacji ruchu kodera pracuje efektywnie; warto zwrócić uwagę na dużą płynność ruchu kamery dla sekwencji City,
- sekwencja Ice najwyższą efektywność kompresji uzyskuje dla TN=2 lub TN=3 (w zależności od rozdzielczości), dla wyższych wartości TN następuje spadek wartości współczynnika PSNR,
- wartości PSNR sekwencji Crew dla parametru TN nie mniejszego niż 2 są do siebie zbliżone i zmieniają się w stopniu niezauważalnym; subiektywnie jakość obrazu jest podobna, zakłócenia w obrazie mają jednak nieco inny charakter dla różnych wartości TN,
- efektywność kompresji dla sekwencji Harbour rośnie przy zmianie parametru TN w przedziale od 0 do 3, dla TN=5 współczynnik PSNR nieznacznie spada,
- kształt krzywej dla sekwencji Soccer jest stosunkowo „płaski” dla niskich rozdzielczości, większą czułość na zmianę TN uzyskano przy większych

rozmiarach przestrzennych i dla wyższych wartości przepływności strumienia danych,

- sekwencja Football wykazuje bardzo małą wrażliwość na parametr TN; przyczyną tego stanu rzeczy może być duża dynamika sekwencji oraz bardzo duża zmienność ruchów obiektów stanowiących pierwszy plan, w konsekwencji błąd predykcji jest znaczny pomimo tego, że koder generuje dużą liczbę wektorów ruchu,
- kształt i nachylenie krzywych dotyczących konkretnych sekwencji testowych, niezależnie od badanej rozdzielczości przestrzennej i czasowej jest zbliżony,
- jakość (mierzona współczynnikiem PSNR) sekwencji wzrasta dla małych wartości TN, przy większych wartościach TN wartość współczynnika PSNR w większości przypadków zmienia się w niewielkim zakresie,
- najlepiej z punktu widzenia efektywności kompresji, najlepiej reagują na zmianę grupy obrazów sekwencje Flat, City i Harbour, nieco gorzej Soccer, najgorzej zaś Football, Crew oraz Ice,
- kształt krzywej dla sekwencji Football jest prawie „płaski”, czego przyczyną może być duża dynamika tej sekwencji, częste zmiany kierunku ruchu przez obiekty znajdujące się na pierwszym planie sekwencji,
- im większa dynamika sekwencji i mniejsza płynność ruchów, tym mniejszy jest wpływ wielkości grupy obrazów na wartość współczynnika PSNR,
- im większa jest niejednorodność ruchów pomiędzy kolejnymi obrazami, tym mniejsza jest czułość na parametr TN (w punktu widzenia współczynnika PSNR),
- z powyższego wynika, że kompensacja ruchu jest stosunkowo niedoskonała (duży błąd predykcji), skoro dla $TN=0$ i $TN=5$ wyniki są porównywalne w przypadku sekwencji Football, Crew, Soccer.

5. PODSUMOWANIE

Badania miały na celu zbadanie wpływu rozmiaru grupy obrazów na efektywność kompresji. Początkowe założenie, mówiące o tym, że wraz ze wzrostem rozmiaru grupy jakość obrazu powinna również rosnąć, uzyskało potwierdzenie jedynie dla niektórych z badanych sekwencji. Sekwencje mało dynamiczne, o płynnych i jednostajnych ruchach, dla których wektory ruchów mogły być dokładnie wyznaczone (co powodowało powstawanie niewielkich błędów predykcji), dobrze reagowały na zmiany rozmiarów grupy obrazów. Natomiast sekwencje dynamiczne, z elementami często zmieniającymi swoje położenie w różnych kierunkach okazały się niską czułością na zmianę rozmiaru grupy obrazów z punktu widzenia wartości współczynnika PSNR.

LITERATURA

- [1] S. J. Choi, J. W. Woods: *Motion-compensated 3-D subband coding of video*, IEEE Transactions on Image Processing, vol. 8, pp. 155-167, 1999
- [2] Shih-T. Hsiang, J. W. Woods: *Invertible three-dimensional analysis/synthesis system for video coding with half-pixel-accurate motion compensation*, Proc. SPIE Conf. Visual Communications and Image Processing, vol. 3653, pp. 537-546, San Jose, CA, 1999
- [3] Shih-T. Hsiang, J. W. Woods: *Embedded image coding using zeroblocks of subband/wavelet coefficients and context modelling*, IEEE International Symposium on Circuits and Systems, vol. 3, pp. 662-665, 2000
- [4] Shih-T. Hsiang, J. W. Woods: *Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank*, Signal Processing: Image Communications, vol. 16, pp. 705-724, 2001
- [5] ITU-T Rec. H.261, Video codec for audiovisual services at p×64 kbit/s
- [6] ITU-T Rec. H.263, Video coding for narrow telecommunication channels at < 64 kbit/s
- [7] ISO/IEC/SC29/WG11/MPEG/N5555, ISO/IEC 14496-10 AVC / ITU-T Rec. H.264, Text Text of ISO/IEC FDIS 14496-10: Information Technology – Coding of audio-visual objects – Part 10: Advanced Video Coding, Pattaya, March 2002
- [8] J. R. Ohm: *Temporal domain sub-band video coding with motion compensation*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 3, San Francisco, CA, pp. 229-232, 1992
- [9] J. R. Ohm: *Three-dimensional motion-compensated subband coding*, Proceedings International Symposium on Video Communications and Fiber Optic Services, SPIE vol. 1977, pp. 188-197, 1993
- [10] J. R. Ohm: *Three-dimensional subband coding with motion compensation*, IEEE Transactions Image Processing, Vol. 3, pp. 559-571, 1994
- [11] J. R. Ohm: *Complexity and delay analysis of MCTF interframe wavelet structures*, doc. m8520, Klagenfurt MPEG meeting, July 2002
- [12] G. Pau, B. Pesquet-Popescu, M. van der Schaar, J. Viéron: *Delay-performance trade-offs in motion-compensated scalable subband video compression*, doc. m11084, Redmond MPEG meeting, July 2004
- [13] D. Santa-Cruz: *Verification of results for SVC on CE3: Low Delay*, doc. m11294, Palma MPEG meeting, October 2004
- [14] J. W. Woods, P. Chen, S. T. Hsiang: *Exploration experimental results and software*, doc. M8524, Klagenfurt MPEG meeting, July 2002
- [15] J. Xu, et al: *Three-dimensional embedded subband coding with optimized truncation (3D ESCOT)*, Applied and Computational Harmonic Analysis vol.10, pp.290-315, 2001