Łukasz Błaszak
Sławomir Maćkowiak
Politechnika Poznańska, Instytut Elektroniki i Telekomunikacji
ul. Piotrowo 3A 60-965 Poznań
[lblaszak, smack]@et.put.poznan.pl

Wrocław, 25-27 czerwca 2003

KODER WIZYJNY AVC ZE SKALOWALNOŚCIĄ PRZESTRZENNO-CZASOWĄ

1. WSTĘP

***Streszczenie*: Atykuł prezentuje skalowalne rozszerzenie kodera AVC. Założeniem jest wprowadzenie możliwie mało modyfikacji do semantyki i składni strumienia binarnego oraz uniknięcie stosowania technologii, która nie jest obecna w istniejącej strukturze kodera AVC. W koderze wykorzystano połączenie skalowalności przestrzennej i skalowalności czasowej. Koder składa się z dwóch koderów cząstkowych, które kodują sekwencję wizyjną i wytwarzają dwa strumienie binarne odpowiadające dwóm różnym poziomom rozdzielczości przestrzennej i czasowej. Każdy z koderów cząstkowych ma swoją własną pętlę predykcji z niezależną estymacją ruchu. Proponowany system stosuje interpolację adaptacyjną. Zależne tryby interpolacji są uważnie włączone w hierarchię trybów kodera AVC. W ten sposób otrzymujemy kody, które odpowiadają prawdopodobieństwu występowania tych kodów.**

2. WPROWADZENIE

Ostatnio, komitet JVT przygotował wersję 1 nowego standardu kodowania sygnału wizyjnego nazwanego AVC [1]. Ten standard jest także nazywany H.264. Standard ten określa udoskonalone narzędzia hybrydowego kodowania sygnału wizyjnego. Główne cechy poprawiające efektywność kodowania to: elastyczny rozmiar prostokątnych bloków w predykcji z kompensacją ruchu, zaawansowana predykcja wewnątrzobrazowa, elastyczny wybór trybów predykcji, pamięć wielu ramek w predyktorze z kompensacją ruchu. Kodek AVC wykazuje znaczącą poprawę efektywności w porównaniu do kodeków H.263 i MPEG-4.

Kodek AVC w wersji 1 nie obsługuje skalowalności, która jest aktualnie uważana jako funkcjonalność ważna dla wielu zastosowań np. systemów bezprzewodowych o zmiennej i zanikającej przepustowości kanału transmisyjnego, transmisji sygnału wizyjnego w heterogenicznych sieciach komunikacyjnych, nierównomiernej odporności na błędy transmisji itd. [3,4].

Celem tego artykułu jest opisanie skalowalnego rozszerzenia kodera AVC. Założeniem jest wprowadzenie możliwie mało modyfikacji do semantyki i składni strumienia binarnego oraz uniknięcie stosowania technologii, która nie jest obecna w istniejącej strukturze kodera AVC. Takie założenie będzie ograniczało koszty implementacyjne skalowalności.

Aktualnie, głównie dwa rozwiązania są rozważane jako kandydaci do przyszłych standardów skalowanych koderów wizyjnych [5], tj.:

- zmodyfikowany hybrydowy koder wizyjny z predykcją i kompensacją ruchu oraz transformacjami opartymi o bloki,
- koder falkowy ze szczególnym naciskiem na trójwymiarowy wizyjny koder falkowy z zespołem filtrów z kompensacją ruchu.

Zakłada się także różne kombinacje obu powyżej wspomnianych podejść.

W przedstawianej pracy, pierwsze podejście jest rozważane, ponieważ nie wymaga ono wprowadzenia żadnych głębszych modyfikacji struktury kodeka AVC. W kontekście kodera H.26L (wcześniejsza wersja kodera AVC), podobne podejście było wykorzystane w sposób opisany w pracy [2]. Pomimo tego podejście z [2] zastosowano w różnych strukturach kodera ze wspólną estymacją ruchu, której wynikiem była gorsza kompensacja ruchu.

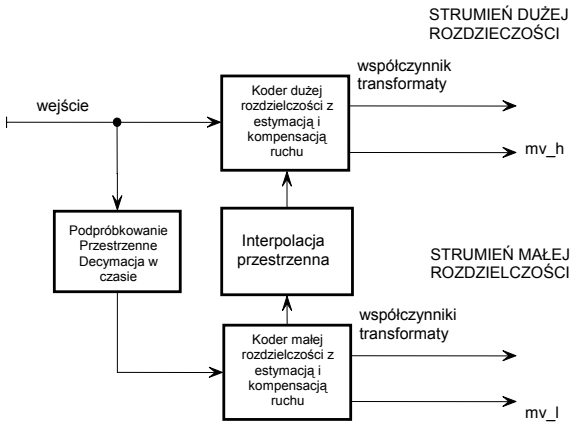
3. STRUKTURA KODERA

Artykuł dotyczy koderów, które wykorzystują skalowalność przestrzenną i czasową. W takim przypadku, warstwa podstawowa reprezentuje sekwencje wizyjną ze zredukowaną rozdzielczością przestrzenną i czasową. Takie połączenie jest bardzo praktyczne, ponieważ dopuszcza na wejściu telewizyjnym warstwę podstawową o rozsądnej rozdzielczości przestrzennej, tj. CIF/SIF.

Proponowany koder skalowalny zawiera struktury wcześniej zaproponowane dla kodeków MPEG-2 oraz H.263 [6-8]. Koder składa się z dwóch koderów cząstkowych z kompensacją ruchu (Rys. 1), które wytwarzają dwa strumienie binarne odpowiadające dwóm różnym wartościom rozdzielczości przestrzennej i czasowej. Każdy koder cząstkowy ma swoją własną pętlę predykcji z niezależną estymacją ruchu.

Koder cząstkowy małej rozdzielczości jest zaimplementowany jako standardowy koder AVC z kompensacją ruchu, który wytwarza strumień binarny w pełni zgodny ze składnią standardu AVC. Koder dużej rozdzielczości jest zmodyfikowanym koderem AVC, który jest zdolny do wykorzystania interpolowanych

makrobloków ze zdekodowanego strumienia warstwy podstawowej. Interpolowane makrobloki są używane jako makrobloki odniesienia dla predykcji ilekroć pozwalają obniżyć koszt. Inne dodatkowe makrobloki odniesienia są tworzone przez uśrednienie odniesienia z predykcji w czasie i makrobloku interpolowanego.



Rys. 1. Ogólna struktura rozważanego kodera skalowalnego. *mv_l* i *mv_h* oznaczają wektory ruchu, odpowiednio z warstwy małej i dużej rozdzielczości.

4. DEKOMPOZYCJA PRZESTRZENNO-CZASOWA

Dobre wykonanie przestrzenno-czasowej decymacji i interpolacji jest krytyczne dla efektywności całego kodera.

Decymacja przestrzenna zawiera dolnoprzepustową filtrację przestrzenną, która chroni przed zjawiskiem aliasingu w sekwencji warstwy podstawowej małej rozdzielczości. Wybór filtru jest kompromisem pomiędzy wysokim rzędem filtru i krótką jego odpowiedzią czasową. Wyniki eksperymentalnych porównań dowodzą ważności odpowiedniego doboru schematu decymacji i interpolacji.

Rozważany system stosuje adaptacyjną względem krawędzi bi-kwadratową interpolację opisaną w pracy [9]. Technika ta jest stosowana zarówno do składowych luminancji i chrominancji.

Technika interpolacji adaptacyjnej względem krawędzi [9] jest rozszerzeniem standardowej nie adaptacyjnej bi-kwadratowej separowalnej interpolacji, która może być opisana w sposób poniższy. Dwuwymiarowa interpolacja jest wykonywana w dwóch krokach: poziomym i pionowym. Niech $f(x)$ jest wartością interpolowaną. Najbliższe dostępne wartości są umiejscowione we współrzędnych x_k (na lewo) i x_{k+1} (na prawo). Niech

$$s = x - x_k \text{ , } 1 - s = x_{k+1} - x \text{ , gdzie } 0 \le s \le 1.$$

Stąd

$$f(x) = f(x_{k-1})(-s^3 + 2\,s^2 - s)/2 + f(x_k)(3s^3 - 5s^2 + 2)/2 + f(x_{k+1})(-3s^3 + 4s^2 + s)/2 + f(x_{k+2})(s^3 - s^2)/2,$$

gdzie x_{k-1}, x_k, x_{k+1} i x_{k+2} są pozycjami czterech sąsiednich znanych punktów.

W schemacie z adaptacją względem krawędzi, zmodyfikowana wartość *s'* używana jest zamiast *s*.

$$s' = s - kAs(s - 1),$$

gdzie *k* jest dodatnim parametrem, który steruje intensywnością wygięcia i A jest funkcją asymetrii danych w sąsiedztwie *x*:

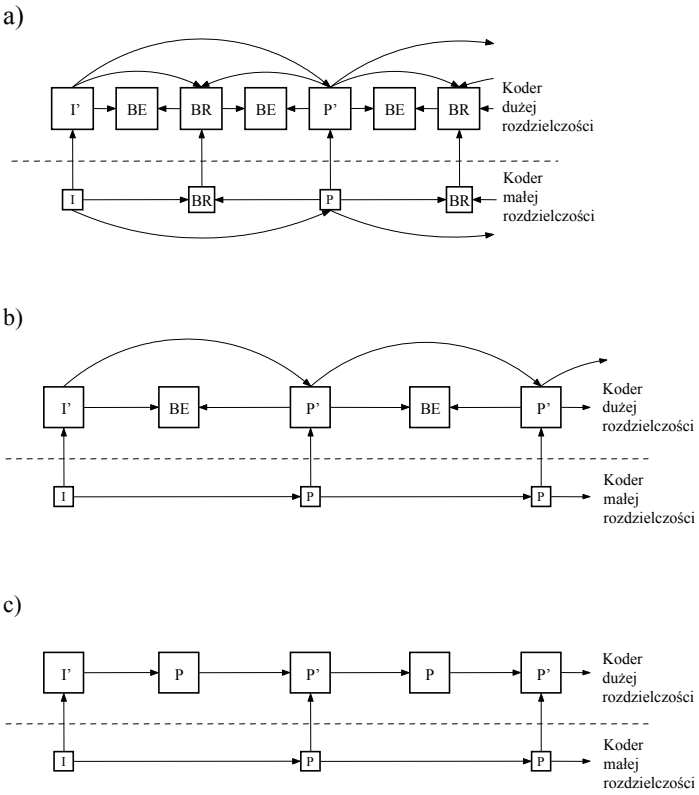
$$A = (|f(x_{k+1}) - f(x_{k-1})| - |f(x_{k+2}) - f(x_k)|)/(L - 1),$$

gdzie $L = 256$ dla 8-bitowej reprezentacji próbki. W eksperymentach $k = 3.05$.

W najprostszym przypadku, zmniejszenie rozdzielczości w czasie jest wykonywane przez opuszczanie ramek. W szczególności pomijanie ramek B tworzy bardzo efektywny i wydajny schemat dekompozycji czasowej (Rys. 2).

Dlatego mogą istnieć dwa typy obrazów B:

- obrazy BE, które istnieją tylko w warstwie rozszerzającej i
- obrazy BR, które istnieją zarówno w warstwie podstawowej jak i rozszerzającej.



Rys. 2. Przykładowe struktury sekwencji wizyjnych małej i dużej rozdzielczości z dekompozycją czasową o współczynniku 2.

5. OBRAZY ODNIESIENIA

W warstwie rozszerzającej, w schemacie kodowania wykorzystywane są dwa dodatkowe obrazy odniesienia:

- obraz interpolowany ze zdekodowanego aktualnie obrazu warstwy podstawowej małej rozdzielczości,
- obraz uśrednieniony z obrazu interpolowanego i ostatniego obrazu odniesienia w czasie.

Można użyć także większej liczby obrazów odniesienia, które są otrzymane jako kombinacja obrazu interpolowanego i różnych odniesień w czasie. Jednakże, ta możliwość nie została wykorzystana w przeprowadzonych eksperymentach i w kolejnych eksperymentach.

Ponadto dla ostatniego, powyżej wspomnianego obrazu odniesienia, można wykonać niezależną estymację ruchu. Estymacja ta określa optymalne wektory ruchu, które skutkują minimalnym błędem predykcji dla odniesienia będącego średnią przestrzennego i czasowego obrazu odniesienia. Ta opcja była użyta w eksperymentach, które zostały dalej przedstawione.

Zastosowanie dodatkowych obrazów odniesienia nie wymaga wprowadzenia modyfikacji składni strumienia binarnego, lecz jedynie małej modyfikacji semantyki zmiennych obrazu odniesienia.

6. WYBOR TRYBU PREDYKCJI

Wyrafinowane techniki predykcji wewnątrzobrazowej i międzyobrazowej przyczyniają się do znacznej poprawy efektywności kodera AVC w porównaniu do jego poprzedników (H.263 i MPEG-2). Koder cząstkowy warstwy rozszerzającej stosuje dodatkowe tryby predykcji, które wykorzystują interpolowany, aktualny obraz z warstwy podstawowej jako odniesienie. Inne tryby predykcji wykorzystują jako odniesienie średnie z predykcji w czasie i interpolacji przestrzennej.

Te tryby interpolacji są uważnie włączone w hierarchię trybów kodera AVC. W ten sposób otrzymujemy kody, które odpowiadają prawdopodobieństwu występowania tych kodów. Odpowiednia hierachia trybów jest pokazana w Tabeli 1.

Wybór odpowiedniego trybu predykcji o najniższym koszcie predykcji odgrywa kluczową rolę. Schemat kodowania upraszcza się do przypadku kodowania równoczesnego, gdy nie stosuje się interpolowanych makrobloków odniesienia w warstwie rozszerzającej. Innym przypadkiem jest sytuacja, gdy w warstwie rozszerzającej nie używa się predykcji w czasie, lecz jedynie interpolowane obrazy z warstwy podstawowej do predykcji makrobloków warstwy rozszerzającej (jak w MPEG-4 FGS). Druga sytuacja jest mało prawdopodobna ze względu na wysoką efektywność predykcji w czasie kodera AVC. Dobra dokładność schematu decymacja-interpolacja skutkuje rozsądnym prawdopodobieństwem, że interpolowany blok odniesienia z warstwy podstawowej wygeneruje mniejszy błąd predykcji aniżeli odniesienie z predykcji w czasie wewnątrz warstwy rozszerzającej.

Tabela 1. Hierarchia trybów predykcji

Typ obrazu	Tryby predykcji
Intra (I)	1. Interpolacja przestrzenna z warstwy podstawowej (rozmiar bloków 16×16). 2. Wszystkie standardowe tryby predykcji wewnątrzobrazowej.
Inter (P)	1. Predykcja (wprzód) z najbliższego obrazu odniesienia. 2. Interpolacja przestrzenna z warstwy podstawowej (rozmiary bloków 16×16 - 4×4). 3. Uśrednienie dwóch powyższych (1, 2). 4. Tryby predykcji czasowej z innych obrazów odniesienia wg kolejności określonej w specyfikacji kodera AVC. 5. Wszystkie standardowe tryby wewnątrzobrazowe.
Inter (B)	1. Predykcja (wprzód, wstecz i dwukierunkowa) z najbliższego obrazu odniesienia. 2. Interpolacja przestrzenna z warstwy podstawowej (rozmiary bloków 16×16 - 4×4). 3. Uśrednienie dwóch powyższych (1, 2). 4. Tryby predykcji czasowej z innych obrazów odniesienia wg kolejności określonej w specyfikacji kodera AVC. 5. Wszystkie standardowe tryby wewnątrzobrazowe.

7. WYNIKI EKSPERYMENTALNE

Model weryfikacyjny kodera i dekodera skalowalnego został zaimplementowany w odniesieniu do oprogramowania kodera JVT w wersji 2.1.

W celu przetestowania efektywności kodera AVC, wykonano szereg eksperymentów z sekwencjami o rozdzielczości (352×288).

Koder charakteryzował się następującymi cechami:

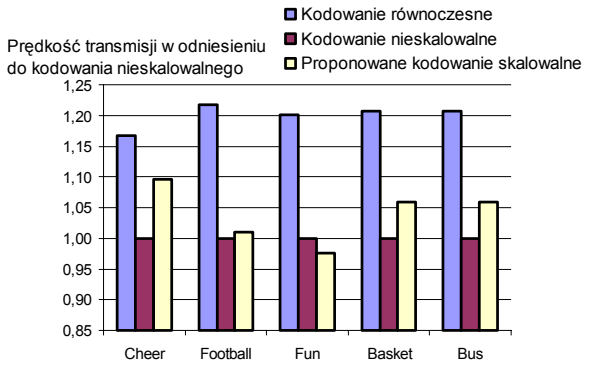
- kodowanie współczynników w trybie CABAC,
- estymacja ruchu z dokładnością do ¼ punktu w obu warstwach,
- włączone wszystkie tryby predykcji.

Eksperymenty zostały wykonane dla różnych ustawień wartości współczynników kwantyzacji. Wartości te były zdefiniowane niezależnie dla obrazów typu I (QP_I), obrazów typu P (QP_P) oraz obrazów typu B (QP_B). W przeprowadzonych eksperymentach zastosowano równe wartości QP_I, QP_P i QP_B dla warstwy podstawowej i dla warstwy rozszerzającej.

Wyniki eksperymentalne dla parametrów kwantyzacji QP_I=20, QP_P=21, QP_B=22 zostały przedstawione w Tabeli 2. Porównanie efektywności w stosunku do kodowania nieskalowalnego AVC przedstawione jest na rysunku 3

Tabela. 2 Wyniki eksperymentalne dla sekcji CIF dla QP_I=20, QP_P=21, QP_B=22

	Basket		Bus		Cheer		Football		Fun	
	PSNR [dB]	Prędkość transmisji [kbps]	PSNR [dB]	Prędkość transmisji [kbps]	PSNR [dB]	Prędkość transmisji [kbps]	PSNR [dB]	Prędkość transmisji [kbps]	PSNR [dB]	Prędkość transmisji [kbps]
Warstwa podstawowa	29,06	156,33	29,06	156,33	31,03	77,85	29,65	261,20	33,12	109,05
Warstwa rozszerzająca	30,28	645,36	30,28	645,36	31,52	434,19	30,76	955,72	34,17	420,95
Cały kodek skalowalny	30,28	801,69	30,28	801,69	31,52	512,04	30,76	1216,92	34,17	530,00
Kodowanie równoczesne	30,33	913,55	30,33	913,55	31,63	544,96	30,89	1466,69	34,39	651,76
Kodek nieskalowalny	30,33	757,22	30,33	757,22	31,63	467,11	30,89	1205,49	34,39	542,71



Rys. 3. Porównanie wymaganych prędkości transmisji dla kodowania skalowalnego i równoczesnego w odniesieniu do kodowania nieskalowalnego AVC (pojedyncza warstwa) sygnału o rozdzielczości CIF.

8. PODSUMOWANIE

Dla systemu dwuwarstwowego ze skalowalnością przestrzenno-czasową, wymagane zwiększenie prędkości transmisji związane ze skalowalnością przestrzenno-czasową jest pomiędzy -1% i 15% w zależności od zawartości sekwencji i prędkości transmisji (Rys.3).

Dla większości przypadków, koder skalowalny charakteryzuje się lepszą efektywnością niż dla przypadku kodowania równoczesnego (simulcast). Zazwyczaj efektywność kodowania skalowanego była znacząco większa niż dla przypadku kodowania równoczesnego.

W przypadku kodera skalowalnego, prędkość transmisji dla warstwy podstawowej wynosiła od około 15% do 22% całkowitej prędkości transmisji wymaganej dla kodera skalowanego dla dwóch warstw.

W artykule zaproponowano rozszerzenie skalowalne struktury kodera AVC. Głównymi cechami prezentowanego rozwiązania są:

- połączenie skalowalności przestrzenno-czasowej,
- niezależna estymacja ruchu dla każdej pętli kompensacji ruchu, tj. dla każdej warstwy,
- adaptacyjna decymacja i interpolacja.

Te powyższe cechy są także przyczyną dobrej efektywności całego kodera.

SPIS LITERATURY

[1] ISO/IEC/SC29/WG11/MPEG02/N4920, ISO/IEC 14496-10 AVC / ITU-T Rec. H.264, Text of Final Committee Draft of Joint Video Specification, Klagenfurt, July 2002.

[2] Y. He, R. Yan, F. Wu, S. Li, H.26L-based fine granularity scalable video coding, ISO/IEC JTC1/SC29/ WG11 MPEG02/M7788, Dec. 2001.

[3] D. Wu, Y. Hou, Y. Zhang, "Scalable video coding and transport over broad-band wireless networks," Proc. of the IEEE, vol. 89, str. 6-20, January 2001.

[4] M. van der Schaar, C.J. Tsai, T. Ebrahimi, Report of ad hoc group on scalable video coding, ISO/IEC JTC1/SC29/ WG11 MPEG02/M9076, Dec. 2002.

[5] J.-R.Ohm, M. van der Schaar, Scalable Video Coding, Tutorial material, Int. Conf. Image Processing ICIP 2001, 2001.

[6] M. Domański, S. Maćkowiak, "On improving MPEG spatial scalability", in Proc. Int. Conf. Image Proc., vol. 2, str. 848-851, 2000.

[7] Ł. Błaszak, M. Domański, A. Łuczak, S. Maćkowiak, Spatio-temporal scalability in DCT-based hybrid video coders, ISO/IEC JTC1/SC29/ WG11 MPEG02/M8672, July 2002.

[8] S. Maćkowiak, "Scalable Coding of Digital Video", Doctoral dissertation, Poznań University of Technology, Poznań 2002.

[9] G. Ramponi, Warped distance for space-variant linear image interpolation, IEEE Transactions on Image Processing, vol. 8, str. 629-639, May 1999.