

Eksperymentalna analiza efektywności trybów kodowania międzyobrazowego w zaawansowanym koderze wizyjnym

Streszczenie: Rozwój technik kompresji sekwencji wizyjnych wiąże się nierozdzielnie ze wzrostem złożoności obliczeniowej. Współczesne standardy zaawansowanej kompresji sekwencji wizyjnych wykorzystują wiele różnorodnych narzędzi w celu podniesienia efektywności kompresji. W praktyce ze względu na złożoność obliczeniową nie można użyć wszystkich dostępnych narzędzi. Celem pracy jest ocena zależności między zbiorem wybranych narzędzi a zapotrzebowaniem na moc obliczeniową i efektywnością kompresji.

1. WSTĘP

Gwałtowny rozwój techniki, z którym mamy do czynienia w ostatnim okresie, spowodował rosnące zapotrzebowanie na nowoczesne metody komunikacji. Jedną z ich form jest cyfrowa transmisja obrazów ruchomych. Powstałe w latach 90-tych ubiegłego wieku standardy kompresji sekwencji wizyjnych dokonały rewolucji w sposobie przekazywania tego typu treści. Patrząc z obecnej perspektywy można powiedzieć, że standardy takie jak H.261 (1990r.), MPEG-1 Video (1993r.) [4] oraz MPEG-2 Video (1994r.) [3] cechowała niewielka złożoność. Nieco bardziej zaawansowanymi były, bardzo do siebie podobne, standardy H.263 (1995r., 1997r.) [5] i MPEG-4 część 2 - Visual (1998r.) [2].

Postęp, który dokonał się w ostatnim czasie, w dziedzinie produkcji wydajnych układów scalonych sprawił, że możliwe stało się stworzenie standardu bardziej zaawansowanego, wykorzystującego wiele nowych technik. Powstała nowa generacja technik nazywanych *zaawansowanymi technikami kompresji* i usankcjonowana nowymi międzynarodowymi standardami Międzynarodowej Organizacji Standaryzacyjnej (ISO) i Międzynarodowej Unii Telekomunikacyjnej (ITU). W wyniku prac ekspertów ITU-T¹ VCEG² oraz ISO/IEC³ MPEG⁴, którzy ustanowili połączoną grupę JVT⁵ został stworzony standard ISO/IEC FDIS 14496-10: Information Technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.

* Autor jest stypendystą Fundacji na Rzecz Nauki Polskiej

¹ International Telecommunication Union-Telecommunication Standardization Sector

² Video Coding Experts Group

³ International Standards Organisation/International Electrotechnical Commission

⁴ Moving Picture Experts Group

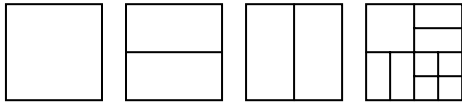
⁵ Joint Video Team

Standard H.264/AVC definiuje wiele nowych narzędzi kompresji, przez co wprowadza ogromny wzrost komplikacji koderów. Mnogość dostępnych sposobów kodowania powoduje, że dokonanie nawet suboptymalnego wyboru trybu kodowania jest zadaniem bardzo czasochłonnym. Niezwykle trudne jest skonstruowanie koderów o wysokiej efektywności kodowania, który spełniałby wymóg kompresji w czasie rzeczywistym. W rezultacie niezbędne staje się opracowanie algorytmów, które pozwoliłyby na znaczne uproszczenie procesu decyzyjnego w stosunku do kodowania opartego na pełnej analizie wszystkich dostępnych możliwości.

Rozwiązania, które można znaleźć w literaturze zakładają przede wszystkim, iż możliwe są wszystkie tryby kodowania, jednakże niektóre z nich weryfikowane są wyłącznie po spełnieniu pewnych warunków, na poziomie makrobloku [8]. Takie podejście powoduje skrócenie czasu kodowania, który jednakże pozostaje silnie zależny od treści obrazu. W przypadku konstruowania koderów mającego pracować w czasie rzeczywistym, może to zaowocować tym, iż dla pewnych przypadków nakład obliczeniowy wzrośnie do nieakceptowanego poziomu. W poniższej pracy przedstawione zostaną wyniki badań dotyczących możliwości całkowitej eliminacji określonych trybów dla określonego scenariusza pracy koderów. Jako, że standard zaawansowanego kodowania sekwencji wizyjnych H.264/AVC udostępnia zestaw uniwersalnych narzędzi kodowania międzyobrazowego, przebadano pewien ich podzbiór w celu stwierdzenia, które z nich są rzeczywiście użyteczne w przypadku kodowania sekwencji o parametrach charakterystycznych dla telewizji cyfrowej standardowej rozdzielczości.

2. TRYBY PREDYKCJI MIĘDZYOBRAZOWEJ KODERA H.264/AVC

Jedną z najważniejszych cech odróżniającą standard H.264/AVC [1] od starszych technik kodowania hybrydowego (MPEG-2, H.263 [6]) jest usprawniona predykcja międzyobrazowa z kompensacją ruchu [5][7]. Usprawnienie to polega na wprowadzeniu możliwości podziału makrobloku na niezależne partycje, z których każda posiada odrębne wektory ruchu. W ten sposób estymacja ruchu powinna lepiej dopasowywać się do kształtu kodowanych obiektów. Dozwolone są podziały makrobloku na dwie partycje 16x8 lub 8x16 lub cztery partycje 8x8. Ponadto, każda z partycji 8x8 stanowi tzw. sub-makroblok i może ulec dalszemu podziałowi na bloki 8x4, 4x8 oraz 4x4 (rys. 1).



Rys. 1. Przykłady podziału makrobloków i sub-makrobloków w predykcji międzyobrazowej.

Dla każdej partycji w strumieniu przesyła się informacje o wektorach ruchu kodowanych różnicowo (tzn. przesyłana jest różnica między wyznaczonym wektorem a jego predykcją). Dodatkowo, w obrazach typu B, makroblok lub sub-makroblok może zostać zakodowany w tzw. trybie *Direct*. Oznacza to, że w strumieniu nie przesyła się żadnej informacji o ruchu, a potrzebne wektory ruchu wyliczane są na podstawie zaawansowanej predykcji przestrzennej lub czasowej. Przesyłany jest natomiast transformowany i poddany kwantyzacji sygnał błędu predykcji międzyobrazowej z kompensacją ruchu.

Istnieje również możliwość zakodowania makrobloku w trybie *Skip*. W takim przypadku dany makroblok jest pomijany w strumieniu, a jego rekonstrukcja w dekodерze odbywa się poprzez bezpośrednie skopowanie odpowiedniego fragmentu obrazu odniesienia.

Wszystkie tryby predykcji międzyobrazowej dostępne w standardzie H.264/AVC zostały przedstawione w tabeli 1.

Tabela 1. Tryby predykcji międzyobrazowej w standardzie H.264/AVC.

Tryb Predykcji	Dostępny w obrazach typu
SKIP	P, B
Direct 16x16	B
16x16	P, B
16x8	P, B
8x16	P, B
Direct 8x8	B
8x8	P, B
8x4	P, B
4x8	P, B
4x4	P, B

3. OPIS EKSPERYMENTU

W przeprowadzonym eksperymencie zbadano stopień poprawy efektywności kodowania w zależności od zbioru użytych w kodowaniu schematów podziału makrobloków. Wyniki te zestawiono z pomiarami średnich czasów kodowania uzyskanych w poszczególnych przypadkach. Uzyskane w ten sposób wykresy miary jakości zakodowanych sekwencji (PSNR składowej luminancji) w funkcji średniego czasu kodowania pozwoliły ustalić, dla których trybów przyrost jakości nie kompensuje wzrostu złożoności kodowania.

W badaniu wykorzystano zestaw 4 sekwencji testowych:

- **Basket** – sekwencja sportowa o dużej dynamice ruchu; bogata w detale; dominuje ruch kamery poziomy oraz zoom;

- **Cheer** – sekwencja sportowa; kamera nieruchoma; dynamicznie poruszające się postacie w centrum sceny oraz tło;
- **WWW** – fragment filmu DVD (*Wild Wild West*) – dużo ciemnych, jednolitych powierzchni; ujęcia ze statyczną kamerą; niewielki ruch postaci i obiektów;
- **Matrix** – fragment filmu DVD (*Matrix*) – duża dynamika ruchu kamery (zoom); silnie zmienne oświetlenie sceny.

Wszystkie sekwencje składały się ze 199 ramek i posiadały standardową rozdzielczość telewizyjną 704×576.

Eksperyment polegał na zakodowaniu każdej z sekwencji przy pomocy kodera referencyjnego H.264/AVC [9] z pięcioma różnymi prędkościami transmisji i różnymi zestawami dozwolonych podziałów makrobloków w predykcji międzyobrazowej. Przebadało następujące zestawy:

- 16x16;
- 16x16, 16x8, 8x16;
- 16x16, 16x8, 8x16, 8x8;
- 16x16, 16x8, 8x16, 8x8, 8x4, 4x8;
- 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4.

Ponadto w każdym z powyższych przypadków dozwolone było również kodowanie w trybach *Direct* oraz *Skip*, a także przy użyciu predykcji wewnątrzobrazowej. Dla każdego zestawu zmierzono średni czas kodowania obrazu oraz jakość zakodowanej sekwencji.

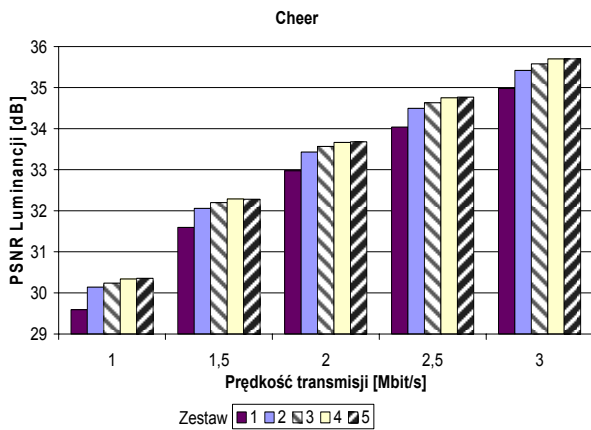
Do badań wybrano prędkości bitowe z zakresu 1 Mbit/s (odpowiednia dla transmisji programów telewizyjnych o obniżonej jakości) do 3 Mbit/s (odpowiednia dla transmisji programów telewizyjnych o bardzo dobrej jakości).

W doświadczeniu użyto wersji 10.2 oprogramowania referencyjnego H.264/AVC [9]. W celu otrzymania możliwie wiarygodnych wyników kodowania dokonano przy pełnej optymalizacji wyboru polegającej na zakodowaniu makrobloku z wykorzystaniem wszystkich dostępnych trybów i wybraniu najlepszego z nich (tzw. optymalizacja R-D). Sekwencje kodowano zakładając typową dla zastosowań telewizyjnych strukturę grupy obrazów, tj.: I BBP BBP BBP BBP. Testy przeprowadzono na platformie Windows XP z procesorem Pentium 4 3,6GHz i 1GB pamięci RAM.

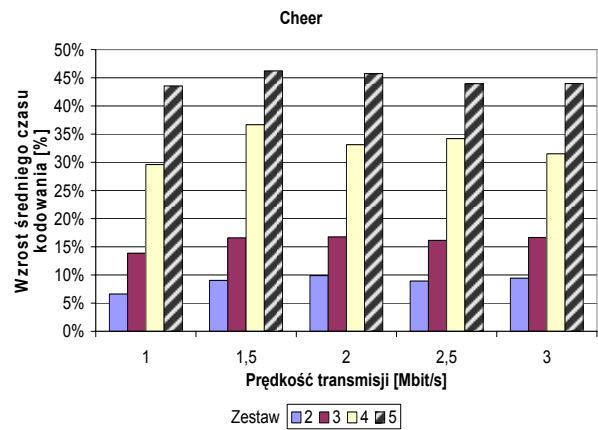
4. WYNIKI EKSPERYMENTU

Pierwszy z przeprowadzonych eksperymentów polegał na zbadaniu jak zezwalanie na różne sposoby podziału makrobloku wpływa na jakość kodowanej sekwencji. Wyniki dla sekwencji *Cheer* i *Matrix* przedstawiono odpowiednio na rysunkach 2 i 3. Dla pozostałych przebadanych sekwencji miały one podobny charakter. Oznaczenia użyte na wykresach należy rozumieć jako zezwolenie na następujące podziały makrobloku:

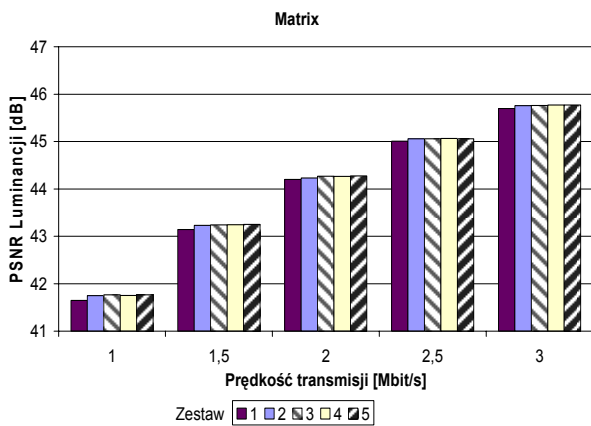
- 1 - 16x16;
- 2 - 16x16 + 16x8 + 8x16;
- 3 - 16x16 + 16x8 + 8x16 + 8x8;
- 4 - 16x16 + 16x8 + 8x16 + 8x8 + 8x4 + 4x8;
- 5 - 16x16 + 16x8 + 8x16 + 8x8 + 8x4 + 4x8 + 4x4.



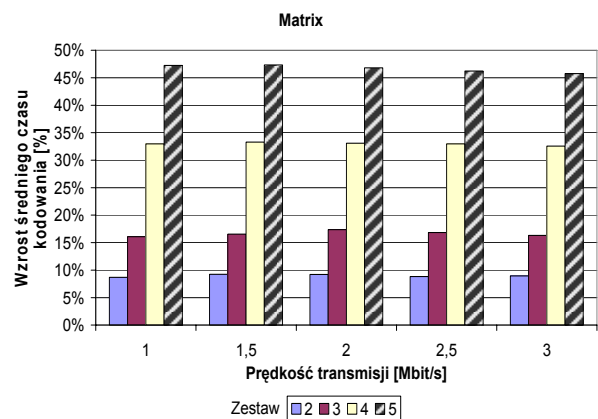
Rys. 2. Wzrost jakości kodowanej sekwencji dla danej przepływności bitowej w zależności od zestawu dozwolonych podziałów makrobloku w predykcji międzyobrazowej dla sekwencji Cheer.



Rys. 4. Wzrost średniego czasu kodowania obrazu w stosunku do trybu 16x16 dla danej przepływności bitowej w zależności od zestawu dozwolonych podziałów makrobloku w predykcji międzyobrazowej dla sekwencji Cheer.



Rys. 3. Wzrost jakości kodowanej sekwencji dla danej przepływności bitowej w zależności od zestawu dozwolonych podziałów makrobloku w predykcji międzyobrazowej dla sekwencji Matrix.



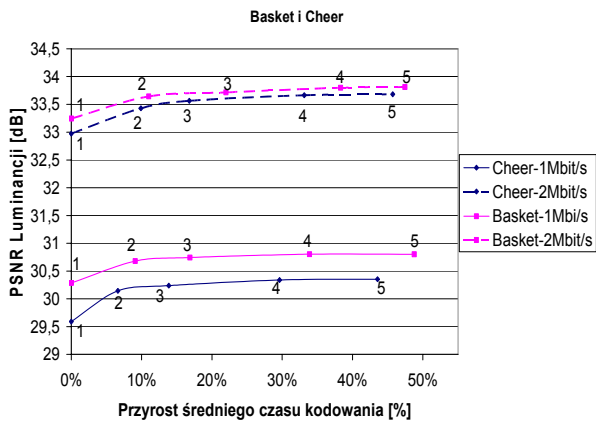
Rys. 5. Wzrost średniego czasu kodowania obrazu w stosunku do trybu 16x16 dla danej przepływności bitowej w zależności od zestawu dozwolonych podziałów makrobloku w predykcji międzyobrazowej dla sekwencji Matrix.

Analizując powyższe wykresy można zauważyć, iż niezależnie od zadanej prędkości transmisji zezwalanie na podziały makrobloku na coraz mniejsze części ma niewielki wpływ na poprawę obiektywnej miary jakości kodowanej sekwencji. Stopień poprawy uzależniony jest jednakże od kodowanej treści. Dla sekwencji obrazów bogatych w szczegóły (*Basket*, *Cheer*) zestawy do czwartego włącznie dają poprawę, choć największy zysk osiągnąć jest już dla zestawu drugiego. Gdy natomiast treść jest uboga w szczegóły (*WWW*, *Matrix*) to podobnie jak poprzednio widoczny zysk jest osiągnięty dla zestawu drugiego, ale kolejne zestawy nie wnoszą już praktycznie żadnej poprawy.

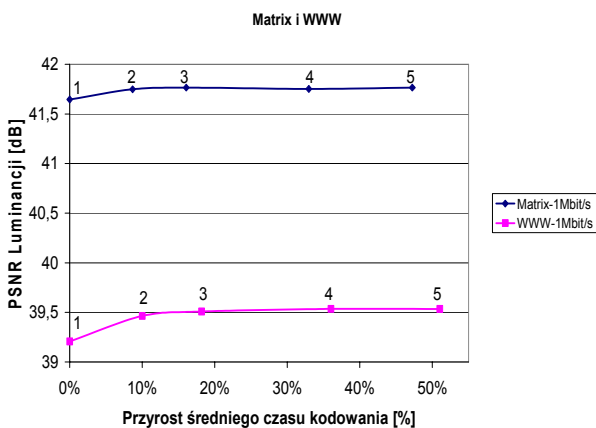
W następnym kroku zbadano jak zezwalanie na różne sposoby podziału makrobloku wpływa na złożoność obliczeniową procesu kodowania wyrażoną poprzez średni czas kodowania obrazu. Na rysunkach 4 i 5 zaprezentowane zostały wykresy przedstawiające przyrost średniego czasu kodowania obrazu w stosunku do sytuacji, gdy podział makrobloku na mniejsze części jest niedozwolony.

Jak widać na powyższych wykresach, niezależnie od kodowanej sekwencji wizyjnej, zezwalanie na analizę coraz większego podzbioru trybów kodowania makrobloku powoduje znaczny wzrost średniego czasu kodowania obrazu. Szczególnie interesującym jest fakt, że przyrost ten staje się coraz szybszy wraz z podziałem makrobloku na coraz mniejsze partycje. Tendencja ta jest niezależna od ustalonej prędkości transmisji. Należy również zauważyć, iż zezwolenie na wszystkie definiowane przez standard H.264/AVC podziały makrobloku powoduje wzrost czasu kodowania o niemal 50% w stosunku do przypadku, gdy podział nie jest dozwolony.

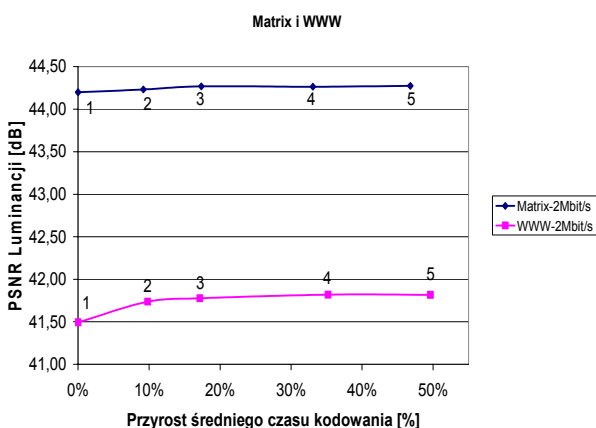
Rysunki 6, 7 i 8 przedstawione poniżej ilustrują krzywe zależności jakości kodowanej sekwencji od przyrostu czasu kodowania. Ich analiza pozwala na ocenę, czy dodanie kolejnych trybów podziału makrobloku znajduje uzasadnienie w świetle wzrostu zapotrzebowania kodera na moc obliczeniową.



Rys. 6. Wzrost jakości kodowanej sekwencji a przyrost średniego czasu kodowania obrazu dla sekwencji Basket i Cheer dla przepływności 1 i 2 Mbit/s. Cyfry od 1 do 5 oznaczają kolejne zestawy dozwolonych trybów.



Rys. 7. Wzrost jakości kodowanej sekwencji a przyrost średniego czasu kodowania obrazu dla sekwencji Matrix i WWW dla przepływności 1 Mbit/s. Cyfry od 1 do 5 oznaczają kolejne zestawy dozwolonych trybów.



Rys. 8. Wzrost jakości kodowanej sekwencji a przyrost średniego czasu kodowania obrazu dla sekwencji Matrix i WWW dla przepływności 2 Mbit/s. Cyfry od 1 do 5 oznaczają kolejne zestawy dozwolonych trybów.

Analiza powyższych rysunków pozwala na wyciągnięcie bardzo interesujących wniosków. Dodawanie kolejnych trybów podziału makrobloku, powodując

znaczny wzrost czasu kodowania, nie oferuje w zamian istotnych zmian w jakości kodowanej sekwencji. O ile można znaleźć uzasadnienie dla dodania trybów podziału 16x8 i 8x16, gdzie czas kodowania wzrósł o maksymalnie 13%, to użycie pozostałych trybów wydaje się być bezzasadne. Uzyskana poprawa jakości (<0,25dB w stosunku do użycia trybów 16x8 i 8x16), nawet gdy dozwolone są wszystkie tryby podziału definiowane przez standard H.264/AVC, nie uzasadnia wzrostu zapotrzebowania koderów na moc obliczeniową.

5. WNIOSKI

W niniejszej pracy przebadano różne podzbiory trybów kodowania międzyobrazowego standardu H.264/AVC w celu określenia ich przydatności do kompresji sekwencji wizyjnych o parametrach charakterystycznych dla telewizji cyfrowej standardowej rozdzielczości. Dla zbadanych sekwencji testowych jedynie dodanie podziałów 16x8 i 8x16 zaowocowało wzrostem jakości, który można uznać za satysfakcjonujący. Dodanie pozostałych trybów kodowania znacznie podnosiło zapotrzebowanie koderów na moc obliczeniową, natomiast wносиło bardzo niewiele do jakości kompresowanych sekwencji. Badania wykazały, że można uzyskać znaczne skrócenie czasu kodowania przy minimalnej stracie jakości <0.25dB dla składowej luminancji.

LITERATURA

- [1] ISO/IEC FDIS 14496-10: Information Technology – Coding of audio-visual objects – Part 10: Advanced Video Coding.
- [2] ISO/IEC FDIS 14496-2: Information Technology – Coding of audio-visual objects – Part 2: Visual.
- [3] ISO/IEC 13818-2: Information technology – Generic coding of moving pictures and associated audio information -- Part 2: Video.
- [4] ISO/IEC 11172-2: Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 2: Video.
- [5] I. E. G. Richardson – *H.264 and MPEG-4 video compression*, Wiley 2003.
- [6] M. Domański – *Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych*. Wyd. 2, WPP, Poznań 2000.
- [7] *IEEE Transactions on Circuits and Systems for Video Technology – Special Issue on the H.264/AVC Video Coding Standard* - volume 13, number 7, str. 560- 576, July 2003.
- [8] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG - JVT-N046 - *Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods*, Hong Kong, Jan, 2005.
- [9] H.264/AVC software coordination site – <http://bs.hhi.de/~suehring/tml>.