

Tomasz Dzieścielewski
Tomasz Grajek
Jarosław Marek
Instytut Elektroniki i Telekomunikacji,
Politechnika Poznańska
[tomdz;tgrajak;jmarek]@multimedia.edu.pl

Eksperymentalna analiza częstotliwości wyboru trybów zaawansowanego kodowania wizyjnego.

Streszczenie: Mnogość oferowanych trybów predykcji stanowi o sile nowoczesnych hybrydowych koderów wizyjnych. Wraz ze wzrostem efektywności kompresji przynosi ona jednak znaczny przyrost komplikacji kodera, dlatego kluczowym zagadnieniem staje się opracowanie metod szybkiego wyboru trybów kodowania. W niniejszym artykule przedstawiono wyniki badań nad statystyką wyboru trybów w koderze H.264/AVC przeprowadzonych dla zbioru typowych sekwencji obrazów przy pełnej optymalizacji. Wyciągnięte zostały wnioski dotyczące możliwości skrócenia czasu kodowania przy minimalnym spadku efektywności kompresji.

1. Wprowadzenie

Ostatnie lata przyniosły znaczny rozwój znanych technik kompresji sekwencji wizyjnych. Postęp ten zaowocował powstaniem nowych ulepszonych algorytmów kodowania hybrydowego, takich jak VC1 [5] czy standard H.264/AVC [1,2,3].

W stosunku do stosowanego dziś w telewizji cyfrowej standardu MPEG-2, nowe techniki kompresji umożliwiają zakodowanie sekwencji o porównywalnej jakości przy dwukrotnie zredukowanej prędkości transmisji. Stosując koder H.264/AVC, bardzo dobrej jakości obraz telewizyjny uzyskuje się już przy strumieniu o prędkości 2 Mbit/s.

Tak znaczącą poprawę efektywności kodowania udało się osiągnąć dzięki równoczesnemu udoskonaleniu wszystkich elementów składowych kodera (pętla predykcji, koder entropijny), jednakże głównego źródła postępu należy upatrywać przede wszystkim w opracowaniu nowych zaawansowanych technik predykcji. We współczesnym koderze hybrydowym, makrobloki obrazu mogą być kodowane w jednym z wielu dostępnych trybów predykcji, co umożliwia precyzyjny dobór stosowanych narzędzi do lokalnego charakteru kodowanej sekwencji.

Mnogość dostępnych trybów kodowania powoduje jednakże ogromny przyrost komplikacji kodera. Dokonanie nawet suboptymalnego wyboru trybu wymaga analizy tak wielu wariantów kodowania, iż

niemożliwe staje się zastosowanie tak skonstruowanego kodera w aplikacjach czasu rzeczywistego.

Dla redukcji wymagań obliczeniowych konieczne jest więc opracowanie technik szybkiego wyboru trybów predykcji. Techniki te powinny umożliwić eliminację a priori jak największej liczby sprawdzanych trybów przy możliwie minimalnym spadku efektywności kompresji względem kodowania z użyciem pełnej optymalizacji.

Podstawą algorytmów szybkiego wyboru trybów predykcji jest obserwacja, iż w pewnych sytuacjach prawdopodobieństwo wyboru poszczególnych trybów jest znacznie mniejsze niż prawdopodobieństwo wyboru trybów pozostałych. Możliwe jest więc założenie, iż pominięcie tego podzbioru trybów w procesie optymalizacji wyboru nie powinno wpłynąć w istotny sposób na pogorszenie parametrów zakodowanego strumienia. Ponieważ analiza pojedynczego trybu zawiera zwykle tak kosztowne obliczeniowo operacje jak estymacja ruchu czy transformacja błędu predykcji, uzyskujemy w ten sposób wyraźne skrócenie czasu kodowania.

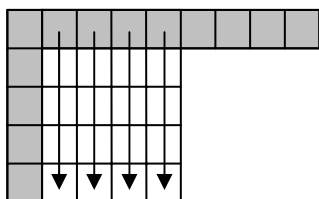
W niniejszym artykule przedstawione zostaną wyniki badań dotyczących rozkładu wyboru trybów predykcji w nowoczesnym hybrydowym koderze wizyjnym. Celem tych badań było sprawdzenie, w jaki sposób zmienia się statystyka wyborów poszczególnych trybów dokonanych przez dobrej jakości koder w warunkach pełnej optymalizacji w zależności od charakteru kodowanej sekwencji. Wyniki uzyskane w eksperymencie mają posłużyć za punkt wyjścia do dalszych badań nad metodami szybkiego wyboru trybów predykcji w koderach hybrydowych. W szczególności, umożliwiają one wskazanie, które z trybów wybierane są najrzadziej, a zatem powinny być rozpatrywane jako kandydaci do eliminacji.

Ze względu na zakończenie prac standaryzacyjnych oraz docelowe zastosowanie w nowoczesnych platformach telewizji cyfrowej, do badań wybrano koder H.264/AVC pracujący w profilu głównym.

2. Tryby predykcji kodera H.264/AVC

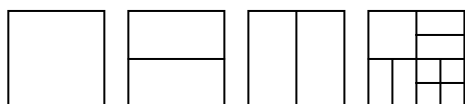
Podobnie jak w starszych technikach kodowania hybrydowego (MPEG-2, H.263 [6]), w standardzie H.264/AVC makroblok obrazu może być zakodowany z wykorzystaniem predykcji wewnątrzobrazowej (INTRA) lub międzyobrazowej (INTER) [1,2,3].

Nowością jest zastosowanie kierunkowej predykcji przestrzennej do kodowania makrobloków w trybie wewnątrzobrazowym (makrobloki typu I). Predykcja ta może odbywać się w blokach 4x4 lub 16x16 i polega na ekstrapolacji próbek sąsiadujących z aktualnie kodowanym blokiem zgodnie z zadanym kierunkiem predykcji (rys. 1). Transformacji oraz kwantowaniu poddawany jest następnie błąd predykcji.



Rys 1. Predykcja kierunkowa w blokach Intra 4x4 na przykładzie predykcji w kierunku pionowym

Kodowanie międzyobrazowe z kompensacją ruchu zostało usprawnione poprzez wprowadzenie możliwości podziału kodowanego makrobloku na niezależne partycje, z których każda posiada odrębne wektory ruchu. Dozwolone są podziały na po dwie partycje 16x8 lub 8x16 oraz cztery partycje 8x8 (rys. 2). Ponadto, każda z partycji 8x8 stanowi tzw. sub-makroblok, który może ulec dalszemu podziałowi na bloki 8x4, 4x8 oraz 4x4.



Rys 2. Schematy podziału makrobloków i sub-makrobloków w predykcji międzyobrazowej

W każdej z kodowanych partycji wykonywana jest medianowa predykcja wektorów ruchu, a rzeczywiste wektory ruchu są kodowane w strumieniu jako błędy tej predykcji. Dodatkowo, w obrazach typu B, makroblok lub sub-makroblok może zostać zakodowany w tzw. trybie Direct. Wówczas w strumieniu nie przesyła się żadnej informacji o ruchu, a potrzebne wektory wyznaczane są przy użyciu zaawansowanej predykcji czasowej lub przestrzennej.

Transformacji i kwantyzacji poddawany jest sygnał błędu otrzymany po predykcji z kompensacją ruchu.

Standard H.264/AVC przewiduje również możliwość zakodowania makrobloku jako SKIP. Makroblok taki jest pomijany w strumieniu i rekonstruowany w dekoderze przez skopiowanie odpowiedniego fragmentu obrazu odniesienia.

Wszystkie dostępne w H.264/AVC tryby predykcji zostały zebrane w tabelach 1 oraz 2.

Tabela 1. Tryby predykcji makrobloku w koderze H.264/AVC

Tryb	Typ predykcji	Dostępny w obrazach typu
INTRA_4x4	wewnątrzobrazowa, 9 kierunków predykcji	I, P, B
INTRA_16x16	wewnątrzobrazowa, 4 kierunki predykcji	I, P, B
16x16	międzyobrazowa	P, B
16x8	międzyobrazowa	P, B
8x16	międzyobrazowa	P, B
8x8	międzyobrazowa	P, B
Direct 16x16	międzyobrazowa	B
SKIP	międzyobrazowa	P, B

Tabela 2. Tryby predykcji sub-makrobloku w koderze H.264/AVC

Tryb	Typ predykcji	Dostępny w obrazach typu
8x8	międzyobrazowa	P, B
8x4	międzyobrazowa	P, B
4x8	międzyobrazowa	P, B
4x4	międzyobrazowa	P, B
Direct 8x8	międzyobrazowa	B

3. Wyniki eksperymentalne

W przeprowadzonym eksperymencie zbadano prawdopodobieństwo wyboru poszczególnych trybów kodowania w zależności od treści kodowanego materiału oraz prędkości transmisji. Analizowane sekwencje zostały poddane kompresji z użyciem pełnego przeszukiwania, tj. wybór najlepszego trybu następował poprzez kolejne sprawdzenie każdego z nich.

W badaniu wykorzystano zestaw 11 sekwencji testowych:

- **football** – sekwencja sportowa o dużej dynamice ruchu; poziomy ruch kamery; zawiera rozległe obszary o bogatej, niejednorodnej teksturze (błotnista trawa),
- **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,
- **Stefan** – sekwencja sportowa; ruch kamery poziomy oraz zoom; dynamicznie poruszająca się postać; nieruchome tło,
- **fun** - bogata w detale oraz ruch obiektów (głównie obrotowy); nieruchoma kamera,

- **bus** – sekwencja statyczna; wolny poziomy ruch kamery; wolno poruszający się obiekt w centrum sceny (autobus),
- **flow** – sekwencja statyczna; powolny ruch poziomy kamery; brak poruszających się obiektów,
- **WWW** – fragment filmu DVD (*Wild Wild West*) – dużo ciemnych, jednolitych powierzchni; ujęcia ze statyczną kamerą; niewielki ruch postaci i obiektów,
- **SW** – fragment filmu DVD (*Star Wars: Episode I*) – statyczna kamera; kilka poruszających się obiektów; nieruchome tło,
- **Matrix** – fragment filmu DVD (*Matrix*) – duża dynamika ruchu kamery (zoom); tło o bogatej i nieregularnej teksturze; silnie zmienne oświetlenie sceny,
- **FF** – fragment filmu DVD (*Final Fantasy*) – animacja komputerowa; bardzo duża dynamika ruchu obiektów we wszystkich kierunkach; gęsty, półprzezroczysty dym.

Wszystkie sekwencje za wyjątkiem *fun* (64 ramki) składały się z ok. 200 ramek i posiadały standardowe rozdzielczości telewizyjne 704×576 lub 704×480.

Eksperyment polegał na zakodowaniu każdej z sekwencji przy pomocy kodera referencyjnego H.264/AVC [4] z dwiema różnymi prędkościami strumienia, a następnie zbadaniu liczby wystąpień poszczególnych trybów predykcji w strumieniach wyjściowych. Na podstawie tych wielkości określona została statystyka rozkładu wybieranych trybów dla trzech możliwych typów obrazów (I, P, B). Analizowany był procentowy udział makrobloków kodowanych w danym trybie w całkowitej liczbie makrobloków użytych do zakodowania obrazów określonego typu. Podobne dane zgromadzono dla sub-makrobloków powstałych w wyniku dokonania w makroblokach podziałów na bloki 8x8.

Prędkości bitowe strumieni ustalono na:

- 2 Mbit/s - odpowiednia dla transmisji programów telewizyjnych o dobrej jakości oraz
- 1,2 Mbit/s - odpowiednia dla transmisji programów telewizyjnych o obniżonej jakości.

W doświadczeniu użyto wersji 10.1 oprogramowania referencyjnego H.264/AVC [4]. W celu otrzymania możliwie wiarygodnych wyników, kodowania dokonano przy pełnej optymalizacji wyboru (tzw. tryb optymalizacji R-D [3]). Wszystkie sekwencje kodowano zakładając typową dla zastosowań telewizyjnych strukturę grupy obrazów, tj.: IBBP BBP BBP BBP.

Wyniki eksperymentu przedstawiono na rys. 3-7.

4. Wnioski

Wyniki eksperymentu pokazują, iż testowane sekwencje można przyporządkować do jednego z dwóch zbiorów. Do zbioru pierwszego należą sekwencje: *football*, *Matrix* oraz *FF*, w zbiorze drugim znajdują się sekwencje pozostałe.

Cechą charakterystyczną sekwencji ze zbioru pierwszego jest bardzo bogata w detale, dynamiczna tekstura występująca w dominujących obszarach obrazu. W przypadku sekwencji *Matrix* i *FF* tworzą ją przede wszystkim skomplikowane efekty świetlne oraz dymy. W sekwencji *football* podobne zjawiska nie występują, jednakże porównywalny stopień złożoności i nieregularności wykazuje błotnisto-trawiaste tło.

Specyficzny charakter tych trzech sekwencji powoduje, iż zachowują się one dość nietypowo. W szczególności można zauważyć, iż odsetek makrobloków typu INTRA wybranych w obrazach typu P i B przewyższa znacznie wyniki uzyskane dla pozostałych sekwencji i w przypadku sekwencji *Matrix* przekracza wartość 90%. Należy więc przypuszczać, iż w sekwencjach o podobnych właściwościach algorytmy szybkiego wyboru trybów odpowiednie dla sekwencji typowych będą częściej podejmowały nieoptymalne decyzje.

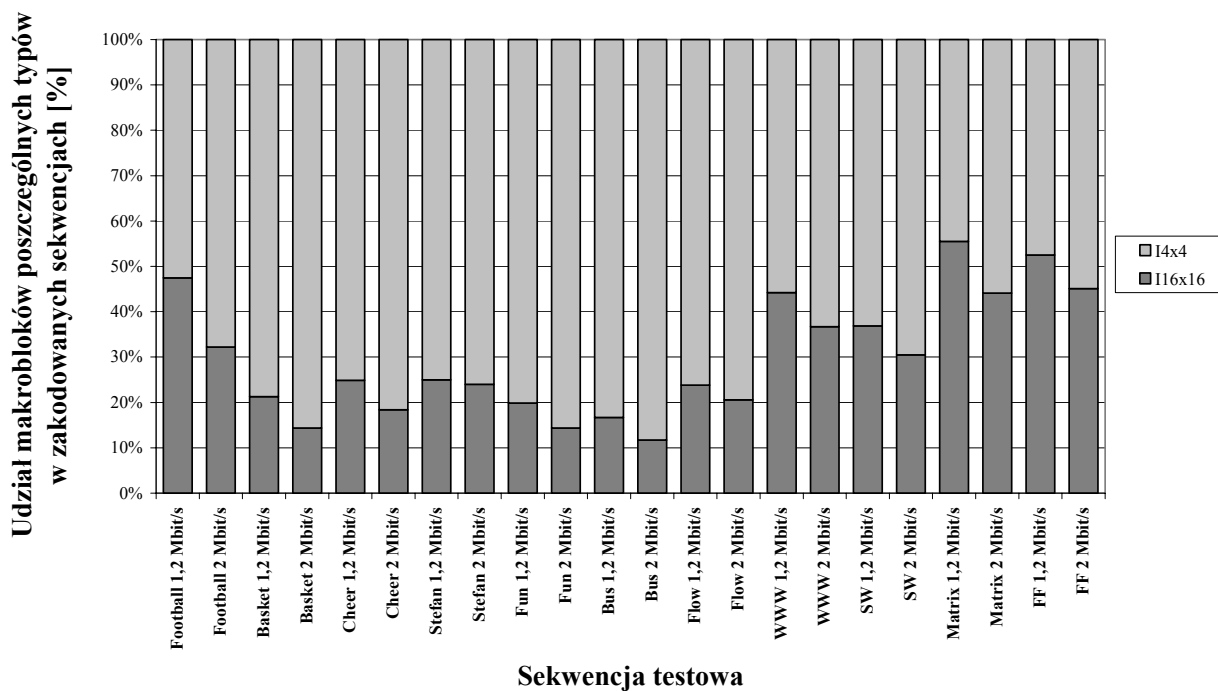
W pozostałych przypadkach można zaobserwować występowanie znacznego odsetku makrobloków typu SKIP. W obrazach typu P wynosi on średnio 17,5%, w obrazach typu B - 39,4%. Ponieważ optymalną decyzję o wysłaniu makrobloku tego typu można podjąć niezależnie od analizy pozostałych trybów, istnieje możliwość znacznej redukcji czasu kodowania poprzez analizę trybu SKIP w pierwszym kroku kodowania. Pozostałe tryby są wówczas sprawdzane tylko wówczas, gdy kodowany makroblok nie może zostać pominięty.

Na przedstawionych wykresach zwraca uwagę również fakt, iż tryby predykcji INTRA oraz INTER 16x16 (w tym SKIP oraz Direct) stanowią zdecydowaną większość wybieranych typów makrobloków (średnio 68,7% w obrazach typu P i 82,3% w obrazach typu B). Sugeruje to, iż poprawnie skonstruowany algorytm szybkiego wyboru trybów powinien niemal zawsze uwzględniać te tryby przy podejmowaniu decyzji, pozostałe tryby zaś należy analizować wyłącznie w uzasadnionych przypadkach.

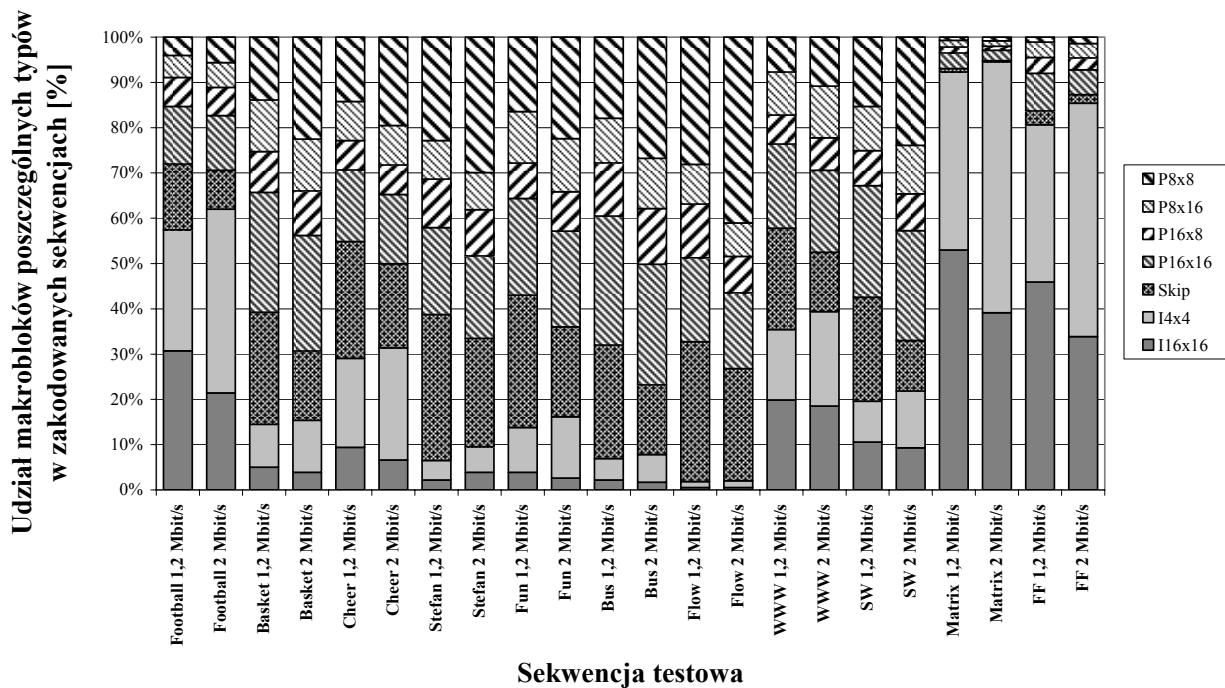
Interpretując wyniki badań dla sub-makrobloków, można stwierdzić, iż wybierane tryby 4x4 stanowią średnio mniej niż 1% powierzchni obrazu (0,6% w obrazach typu P, 0,2% w obrazach typu B), dlatego nawet ich całkowite pominięcie nie powinno wpłynąć znacząco na pogorszenie efektywności kompresji. Co więcej, w przypadku obrazów typu B, tezę tę można rozszerzyć także na tryby 8x4 i 4x8 (łącznie średnio 1,4% powierzchni obrazu).

W niniejszym artykule, na przykładzie kodera H.264/AVC, przedstawiono statystyki wyboru trybów predykcji makrobloków w nowoczesnym koderze wizyjnym uzyskane dla zbioru sekwencji testowych o rozdzielczościach telewizyjnych oraz zróżnicowanym charakterze treści. Uzyskane wyniki posłużyły do sformułowania ogólnych wniosków i uwag dotyczących możliwości zwiększenia wydajności czasowej zaawansowanych metod kompresji. Zgromadzone w eksperymencie dane stanowią będą punkt odniesienia w dalszych pracach nad metodami sterowania koderami wizyjnymi.

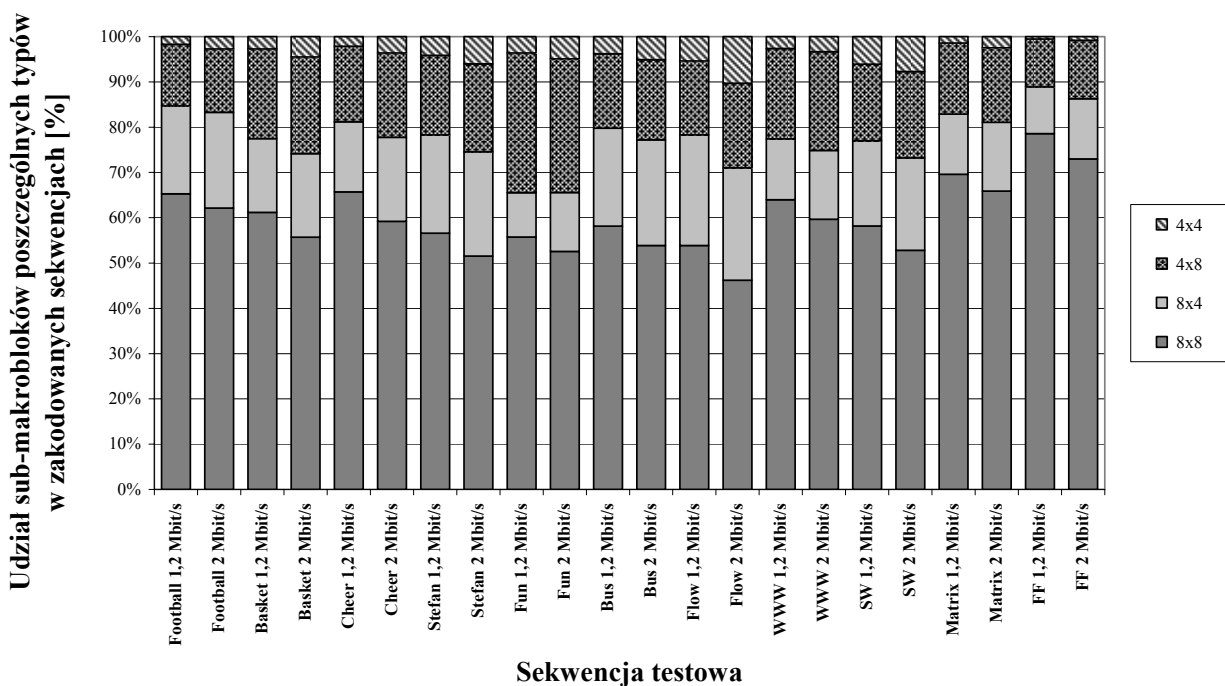
- [1] ISO/IEC FDIS 14496-10: Information Technology – Coding of audio-visual objects – Part 10: Advanced Video Coding
- [2] I. E. G. Richardson – *H.264 and MPEG-4 video compression*, Wiley 2003
- [3] *IEEE Transactions on Circuits and Systems for Video Technology – Special Issue on the H.264/AVC Video Coding Standard* - volume 13, number 7, July 2003
- [4] H.264/AVC software coordination site – <http://bs.hhi.de/~suehring/tml>
- [5] SMPTE VC-1 Site - <http://www.smppte-vc1.org/>
- [6] M. Domański – *Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych*. Wyd. 2, WPP, Poznań 2000



Rys 3. Procentowy udział typów makrobloków w obrazach typu I w zależności od treści sekwencji wizyjnej i prędkości transmisji

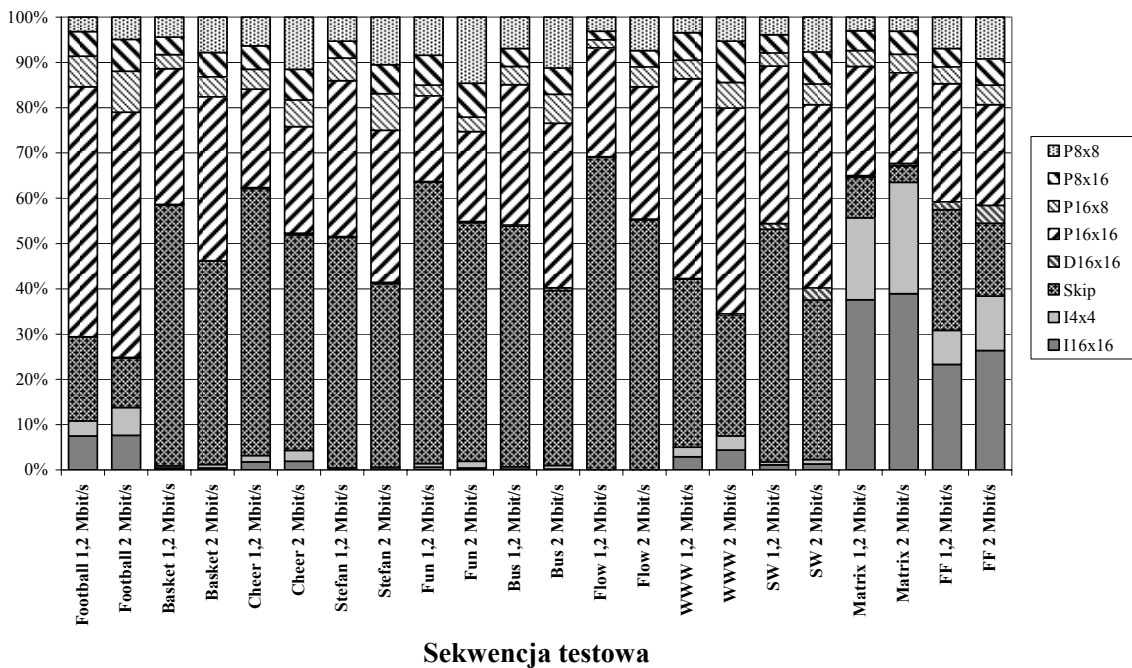


Rys 4. Procentowy udział typów makrobloków w obrazach typu P w zależności od treści sekwencji wizyjnej i prędkości transmisji



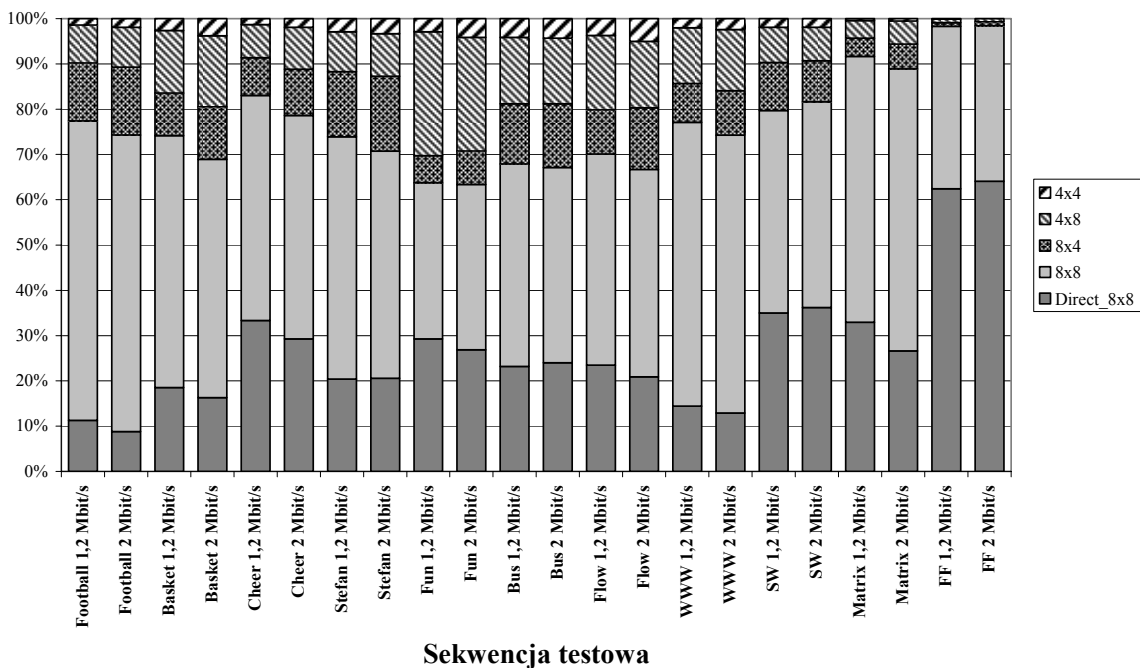
Rys 5. Procentowy udział typów sub-makrobloków w obrazach typu P w zależności od treści sekwencji wizyjnej i prędkości transmisji

Udział makrobloków poszczególnych typów w zakodowanych sekwencjach [%]



Rys 6. Procentowy udział typów makrobloków w obrazach typu B w zależności od treści sekwencji wizyjnej i prędkości transmisji

Udział sub-makrobloków poszczególnych typów w zakodowanych sekwencjach [%]



Rys 7. Procentowy udział typów sub-makrobloków w obrazach typu B w zależności od treści sekwencji wizyjnej i prędkości transmisji