

Sławomir Maćkowiak
Marek Domański
Instytut Elektroniki i Telekomunikacji
Politechnika Poznańska, Poznań

Badania eksperymentalne kodeka MPEG-2 cyfrowego sygnału telewizyjnego

Praca przedstawia wyniki badań testowych programowej realizacji kodeka standardu MPEG-2. Eksperymenty prowadzone z wykorzystaniem sekwencji telewizyjnych "Susie" oraz "Mobile and calendar" pokazują zależność między przepływnością a jakością sygnału zrekonstruowanego mierzoną wartością PSNR dla składowej luminancji. Badano też wpływ innych parametrów kodera związanych ze strukturą grupy obrazów.

1. Wstęp

Opracowanie efektywnych technik kompresji cyfrowych sygnałów wizyjnych, przyjęcie odpowiednich standardów międzynarodowych oraz osiągnięcia w zakresie budowy szybkich układów kompresji przyczyniły się w ostatnich latach do dynamicznego rozwoju multimedialnych technik komunikacyjnych oraz cyfrowej telewizji. Telekonferencje, cyfrowa transmisja obrazów, wideotelefonie zbliżyły się do codzienności. Rozwój komunikacji multimedialnej oraz rozprzestrzenianie się telewizji cyfrowej stymuluje z kolei dalszy rozwój metod kompresji, a także odpowiednich standardów.

Przedstawiana praca dotyczy standardu ISO/IEC IS 13818 [1] znanego głównie pod nazwą MPEG-2. Ten zatwierdzony w 1994 roku standard dotyczy kompresji cyfrowych sygnałów audiowizualnych, a zwłaszcza kodowania sygnałów telewizji cyfrowej, zarówno o standardowej jakości, jak i o wysokiej rozdzielczości (HDTV). Oczywiście w trakcie procesu ustanawiania standardu, a także później, badano efektywność algorytmu MPEG-2, jednak w literaturze można znaleźć stosunkowo niewiele kompleksowych wyników dotyczących pracy kodeka w różnych trybach.

W tej pracy prezentowane są właśnie wyniki eksperymentalnych badań kodeka w różnych trybach pracy. Wykorzystanie implementacji programowej przygotowanej przez MPEG Simulation Group [5] pozwoliło testować pracę kodera przy różnych, łatwo zmienianych parametrach. Uzyskane wyniki są pewną wskazówką dla wyboru parametrów kodeka w realizacjach sprzętowych za pomocą wyspecjalizowanych układów scalonych. Część badań wykonano w ramach przygotowania pracy dyplomowej magisterskiej [4].

Przeprowadzone eksperymenty miały na celu praktyczne pokazanie jak w rzeczywistości zachowuje się koder i dekodek standardu MPEG-2. Obraz telewizyjny zgodny z zaleceniem ITU-R BT.601 poddano kompresji i badano jakość obrazów zdekodowanych.

Do badań użyto dwóch sekwencji złożonych ze 150 obrazów w formacie 4:2:0: „Susie” oraz „Mobile and calendar” różniących się między sobą charakterem. Sekwencja „Susie” (w standardzie NTSC), przedstawia kobietę rozmawiającą przez telefon. W okolicy 40-90 obrazu występuje gwałtowny ruch wynikający z odrzucenia przez aktorkę włosów do tyłu. Sekwencja „Mobile and calendar” (w standardzie PAL) zawiera bardzo złożony ruch. Na pierwszym planie kolejka poruszająca się po torach toczy przed sobą obracająca się piłkę.

Na drugim planie wiszący na ścianie kalendarz przesuwa się w dół. Dodatkowo występuje w sekwencji najazd kamery na plan. Na rys.1 pokazane są jedne z pierwszych obrazów z obu sekwencji.



Rys.1. Obrazy sekwencji testowych. Po prawej obraz „Susie”, po lewej obraz sekwencji „Mobile and calendar”

Pomiar obiektywny został przeprowadzony za pomocą współczynnika PSNR zdefiniowanego

$$PSNR = 10 \log_{10} \frac{N \cdot 255^2}{\sum_i (y_i - x_i)^2} \quad (1)$$

gdzie:

- x_i - wartość i -tej próbki obrazu zdekodowanego,
- y_i - wartość i -tej próbki obrazu oryginalnego,
- N - liczba próbek w obrazie.

W tej pracy współczynnik PSNR jest liczony osobno dla każdej składowej Y,U oraz V.

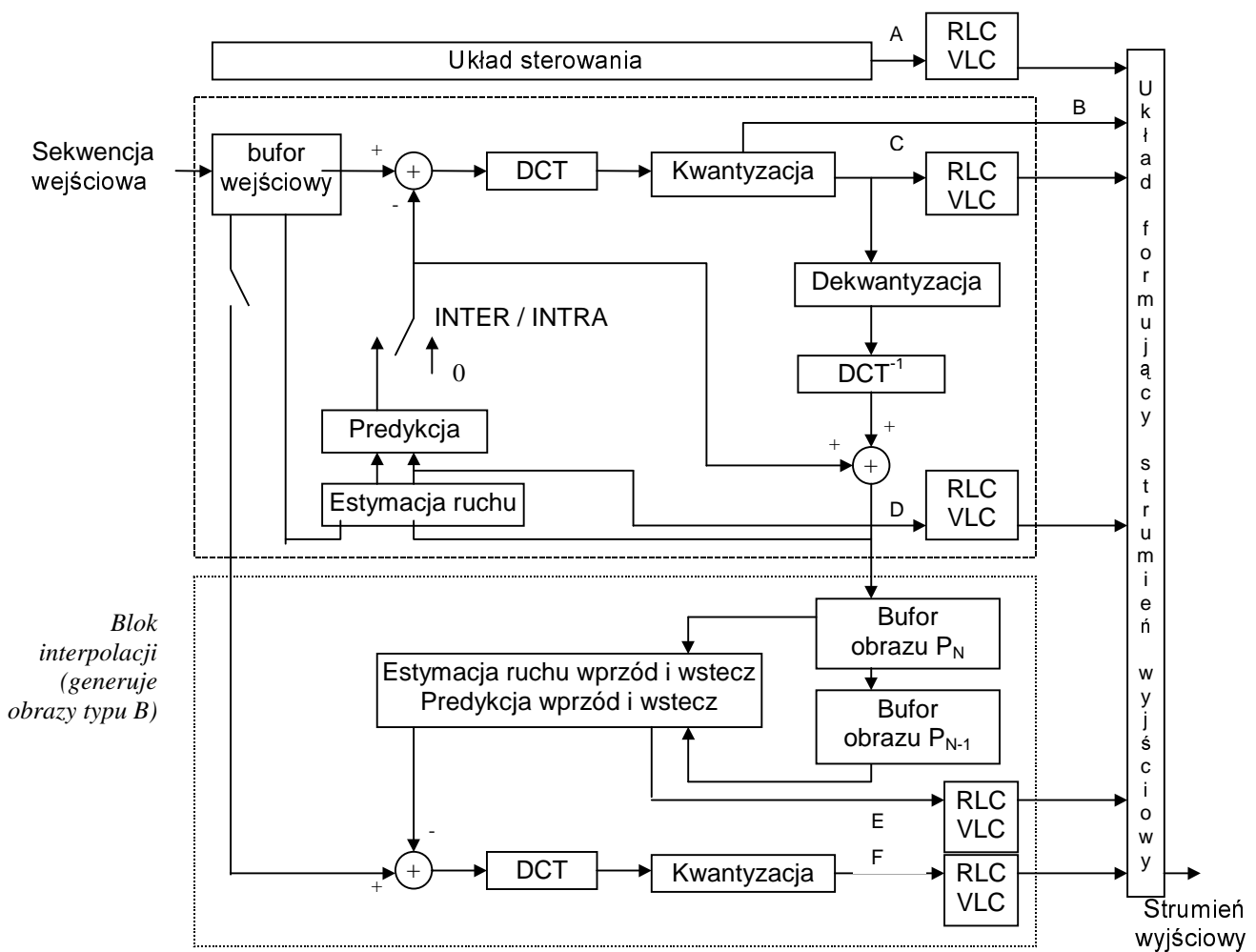
2. Standard MPEG-2

Standard ISO/IEC 13818-2 nie określa budowy kodera lecz jedynie generowany strumień bitów implikując w pewien sposób budowę kodera i dekodera. Koder blokowy MPEG-2 (rys.2.) wykorzystuje dwie podstawowe techniki. Redukcja informacji w czasie wykorzystuje korelację wzajemną pomiędzy kolejnymi obrazami i opiera się na predykcji oraz kompensacji ruchu. Natomiast redukcja informacji przestrzennej korzysta z przekształcenia kosinusowego (DCT), kwantyzacji, kodowania entropowego oraz kodowania długości ciągów.

Stosuje się trzy różne sposoby kodowania a wraz z tym trzy różne typy obrazów. W kodowaniu wewnątrzobrazowym *intraframe* (obrazy typu I – *intra-coded picture*) pojedynczy obraz zostaje poddany podziałowi na bloki o rozmiarze 8x8, a następnie te bloki są poddawane kodowaniu transformatowemu. Obliczana jest dwuwymiarowa transformata kosinusowa (DCT) każdego bloku. Próbkę transformaty są następnie równomiernie kwantowane przez podzielenie

współczynników transformaty DCT przez rozmiar kroku kwantyzacji. Rozmiary kroków kwantyzacji dla poszczególnych próbek transformaty są definiowane przez tablicę kwantyzacji, którą przygotowuje się tak by uwzględnić właściwości percepcyjne wzroku ludzkiego. Skwantowane wartości współczynników są poddawane kodowaniu długości ciągów RLC - *run length coding* i kodowaniu entropowemu VLC - *variable length coding* [1][3].

Kolejnym sposobem jest kodowanie międzyobrazowe *interframe* predycyjne wprzód (obraz typu P *predictive-coded picture*). Jego istotą jest kodowanie błędu predykcji obliczonego jako różnica między aktualnym oryginalnym obrazem na wejściu kodera a wytworzonym w predyktorze obrazem przewidywanym na podstawie obrazów poprzednich. Ten ostatni tworzony jest dwuetapowo. Najpierw estymator wylicza wektory ruchu między dwoma kolejnymi obrazami i przesyła tą informację do predyktora. Następnie predyktor tworzy obraz na podstawie obrazu poprzedniego i otrzymanych wektorów ruchu. Do dekodera wysyłane są skwantowane wartości współczynników transformaty DCT oraz wektory ruchu.



Rys.2. Schemat kodera blokowego MPEG-2, A-nagłówki i dane sterujące, B-parametry kwantyzacji, C-współczynniki DCT w obrazie typu I lub P, D-wektory ruchu dla obrazów typu P, E-wektory ruchu dla obrazów typu B, F-współczynniki DCT w obrazie typu B

Najbardziej efektywną i zarazem najbardziej pracochłonną techniką kodowania jest kodowanie międzyobrazowe predycyjne dwukierunkowe (interpolacyjne - obraz typu B *bidirectionally predictive-coded picture*). Obrazy typu B są obrazami, które powstają na podstawie informacji zawartych w dwóch kolejnych obrazach P (lub I).

Dla celów kodowania według standardu MPEG-2 sekwencje dzielone są na grupy obrazów (GOP- *group of pictures*). Każda grupa zaczynając się obrazem typu I może być kodowana i dekodowana niezależnie od pozostałych. Pozostałe N obrazów grupy koduje się w trybie P lub B. Liczba N obrazów typu P i B w grupie jest parametrem procesu kompresji.

W standardzie przyjmuje się, że wewnątrz grupy obrazów GOP każdy kolejny M-ty obraz jest typu P, a pozostałe są typu B. Tak więc liczba N określająca łączną liczbę obrazów typu P i B w grupie jest wielokrotnością wspomnianej właśnie liczby M określającej odległość kolejnych obrazów typu P.

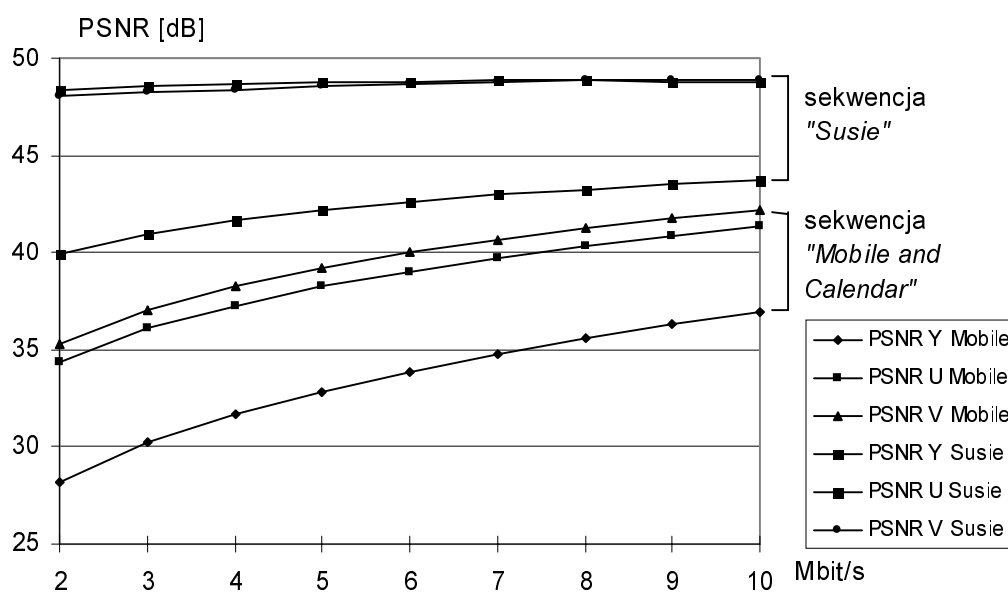
Badania zostały przeprowadzone na koderze nie działającym w czasie rzeczywistym (off-line). Prędkość nawet szybkich stacji roboczych jest na razie zbyt mała dla implementacji kodeka MPEG-2 w czasie rzeczywistym.

3. Pomiar współczynnika PSNR w zależności od przepływności bitowej kanału

Przed badaniem zostały ustawione wszystkie zmienne sterujące koderem. Podczas kolejnych pomiarów zmieniana jest tylko wartość przepływności binarnej. Parametry ustawiono następująco:

- liczba obrazów grupie GOP - N=15,
- odległość między obrazem I a P, M=3,
- precyzja wyznaczania współczynników DCT 10 bitów,
- tablice kwantyzacji takie same dla luminancji i chrominancji,
- skanowanie zig-zag,
- przeplot,
- parametry wynikające z użytej sekwencji: rozmiar, format chrominancji, format sygnału wizyjnego, macierze koloru.

Wyznaczono średnią wartość PSNR dla całej sekwencji złożonej ze 150 obrazów.

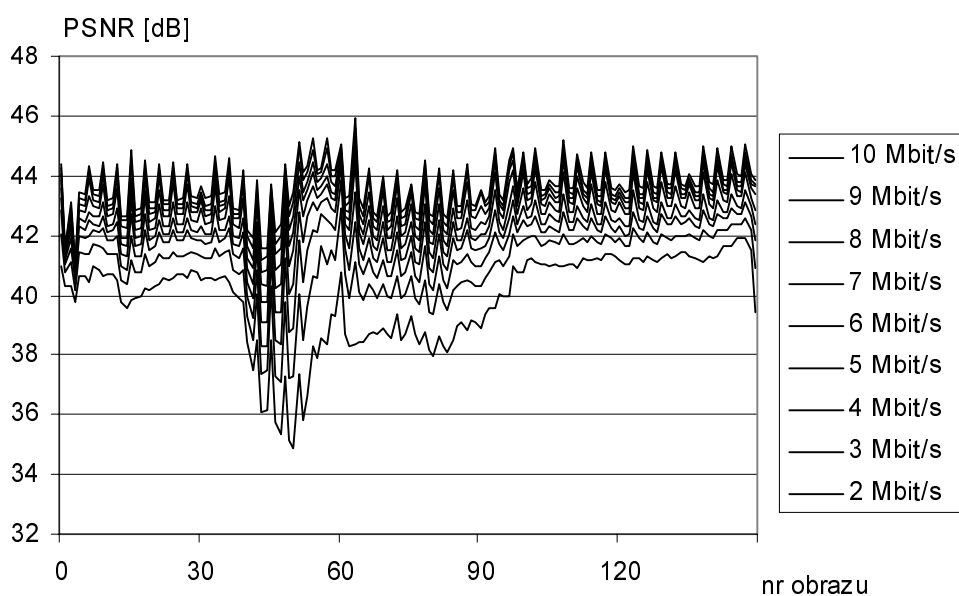


Rys.3. Wykres uśrednionego współczynnika PSNR dla sekwencji "Susie" i "Mobile and calendar" w zależności od przepływności

Oczywiście wyniki pomiarów wskazują, że niezależnie od sekwencji zmniejszenie przepływności powoduje zmniejszenie współczynnika PSNR (rys.3). Wizualnie objawia się to pogorszeniem jakości koloru oraz uwidocznieniem efektów blokowych (widać podział obrazu na bloki). Zmniejszając przepływność binarną kanału powodujemy przyznanie odpowiednio mniejszej

liczby bitów na dane obrazu. To powoduje, że musimy przesłać dane reprezentujące współczynniki transformaty DCT z odpowiednio mniejszą dokładnością. Składnia strumienia wizyjnego ma pewną strukturę wymagającą odpowiedniej liczby bitów. Dlatego też struktura strumienia (kody nagłówkowe, sterujące) bez danych obrazu implikuje minimalną przepływność binarną kanału wymaganą do transmisji strumienia.

W eksperymentach użyto identycznych tablic kwantowania zarówno dla składowych chrominancji, jak i dla luminancji. Widmo sygnału chrominancji ma znacznie węższe pasmo, co uzasadnia decymację chrominancji przyjętą w formacie danych cyfrowych (format 4:2:0). Jednak nawet zastosowanie decymacji pozwala kwantować składową chrominancji nieco zgrubniej niż luminancji. Wyniki eksperymentalne pokazują, że zastosowanie identycznych tablic kwantowania zarówno dla składowych chrominancji, jak i dla luminancji powoduje, że wartości PSNR dla luminancji są o około 6dB niższe niż dla chrominancji.

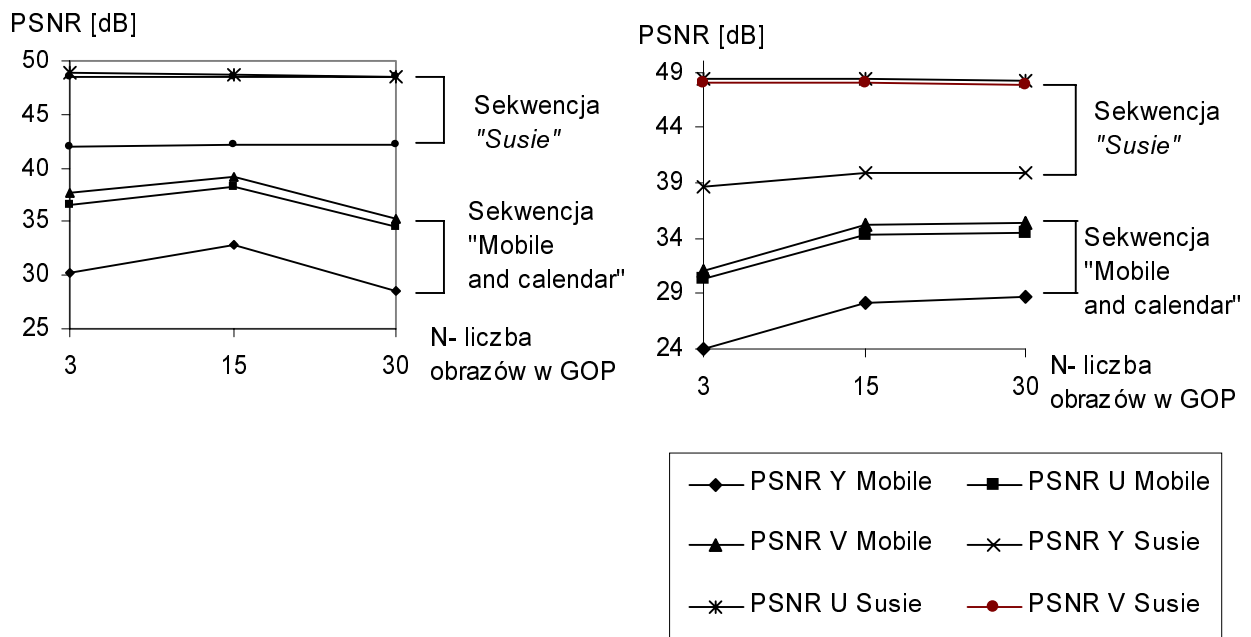


Rys.4. Zmienność w czasie współczynnika PSNR składowej luminancji sekwencji "Susie"

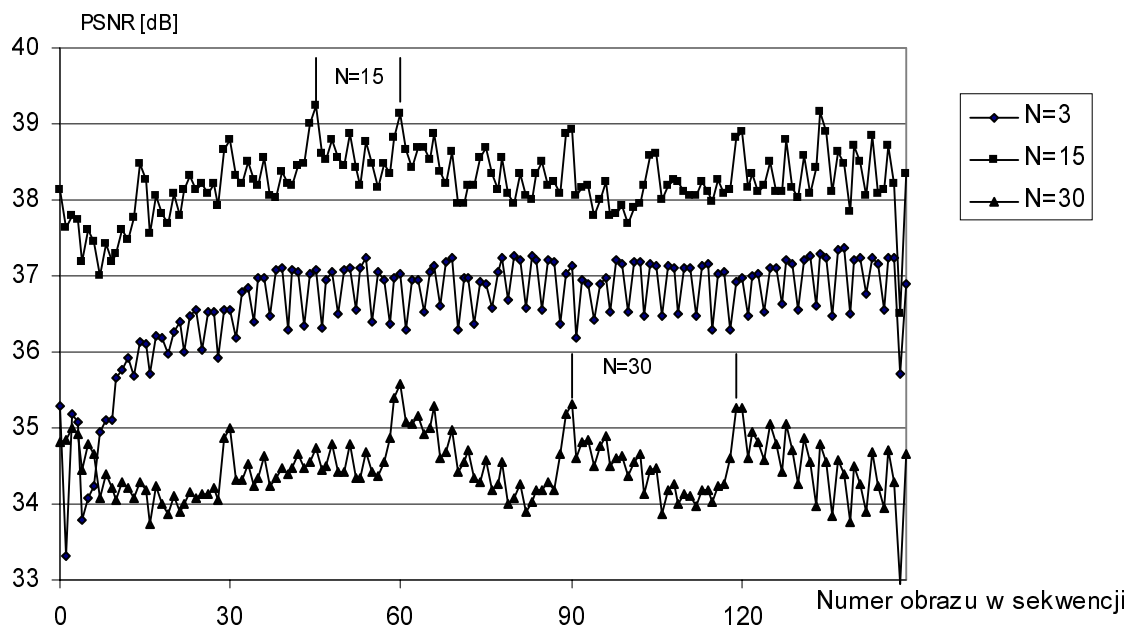
Rysunek 4 przedstawia wartości współczynnika PSNR składowej luminancji Y w funkcji czasu (dla kolejnych obrazów) w zależności od przepływności binarnej kanału. W okolicach 40-90 obrazu następuje gwałtowny ruch wynikający z odrzucenia przez aktorkę włosów do tyłu. Wykres znakomicie przedstawia nagły spadek wartości współczynnika dla danej przepływności wynikający z ruchu obiektu w sekwencji testowej. Spowodowane to jest koniecznością przesłania dodatkowej informacji związanej z ruchem obiektu.

4. Pomiar uśrednionego współczynnika PSNR w funkcji liczby ramek w GOP

Pierwsze badanie dotyczące struktury grupy obrazów dotyczy wpływu liczby obrazów zawartych w grupie na wartość współczynnika PSNR. Założono zmianę liczby obrazów w grupie pozostawiając jej wewnętrzną strukturę stałą. Przyjęto $M=3$ czyli 2 obrazy B między kolejnymi obrazami I lub P. Pozostałe parametry pracy kodera pozostały niezmiennione w stosunku do badania poprzedniego.



Rys.5. Wartości uśrednione współczynnika PSNR w funkcji liczby obrazów w grupie. Po lewej stronie dla przepływności 5Mbit/s, po prawej dla przepływności 2Mbit/s



Rys.6. Wartości współczynnika PSNR składowej chrominancji U sekwencji „Mobile and calendar” w zależności od liczby ramek w GOP (dla przepływności 5Mbit/s)

Najmniejszą wartość współczynnika PSNR osiągnięto przy 3 obrazach w grupie. Tak częste przesyłanie obrazów typu I nie ma sensu z punktu widzenia efektywności kodowania. Obrazy I dostarczają punktów dostępu w zakodowanej sekwencji odkąd może nastąpić rozpoczęcie dekodowania. Z drugiej zaś strony przesyłanie zbyt rzadkie obrazów typu I (przypadek z 30 obrazami w GOP) powoduje, że jakość obrazu spada. Podobnie stwierdzono w [5], że najlepszą jakość obrazu i efektywność kodowania uzyskuje się przy 15 obrazach w GOP dla sekwencji NTSC

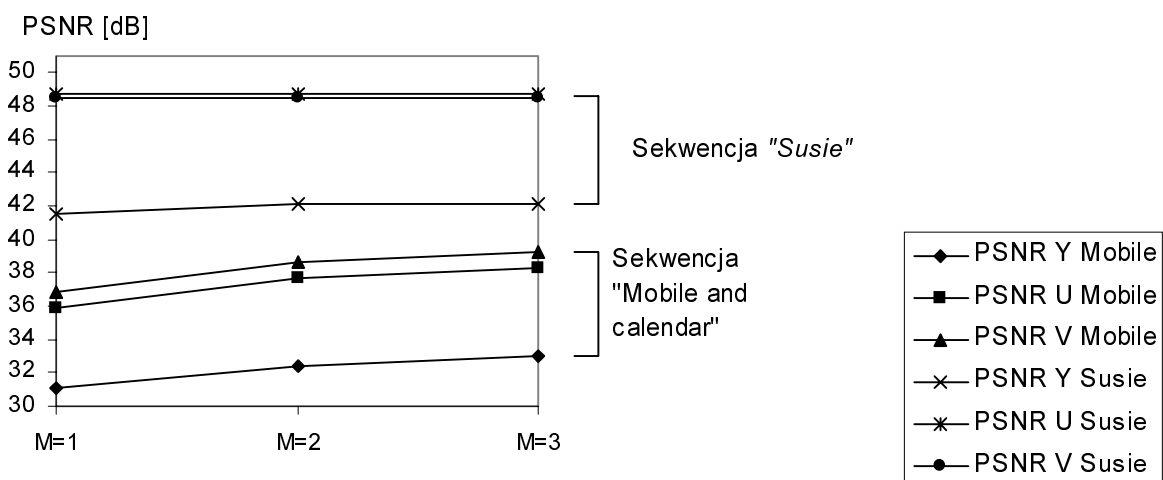
i 12 obrazach dla PAL. Na rysunku 6 zaznaczono dwie przykładowe grupy obrazów dla $N=30$ i $N=15$.

5. Pomiar uśrednionego współczynnika PSNR w zależności od liczby obrazów B w strukturze GOP

To badanie ma na celu sprawdzenie jak na wartość współczynnika PSNR wpływa liczba obrazów typu B. Już w poprzednim badaniu można było zaobserwować zmniejszenie wartości PSNR przy ograniczeniu liczby obrazów B. W tym badaniu długość GOP została zwiększona do $N=18$ ze względu na wymóg krotności M w stosunku do N . Badanie zostało przeprowadzone dla $M=1,2$ i 3 . Wygląda to następująco:

- dla $M=1$ I-P-I-P,
- dla $M=2$ I-B-P,
- dla $M=3$ I-B-B-P,

Badanie zostało przeprowadzone dla przepływności binarnej równej 5Mbit/s. Pozostałe parametry są ustawione jak w poprzednich badaniach.



Rys.7. Uśrednione wartości współczynnika PSNR dla sekwencji „Susie” i „Mobile and calendar” dla różnej liczby obrazów typu B w sekwencji

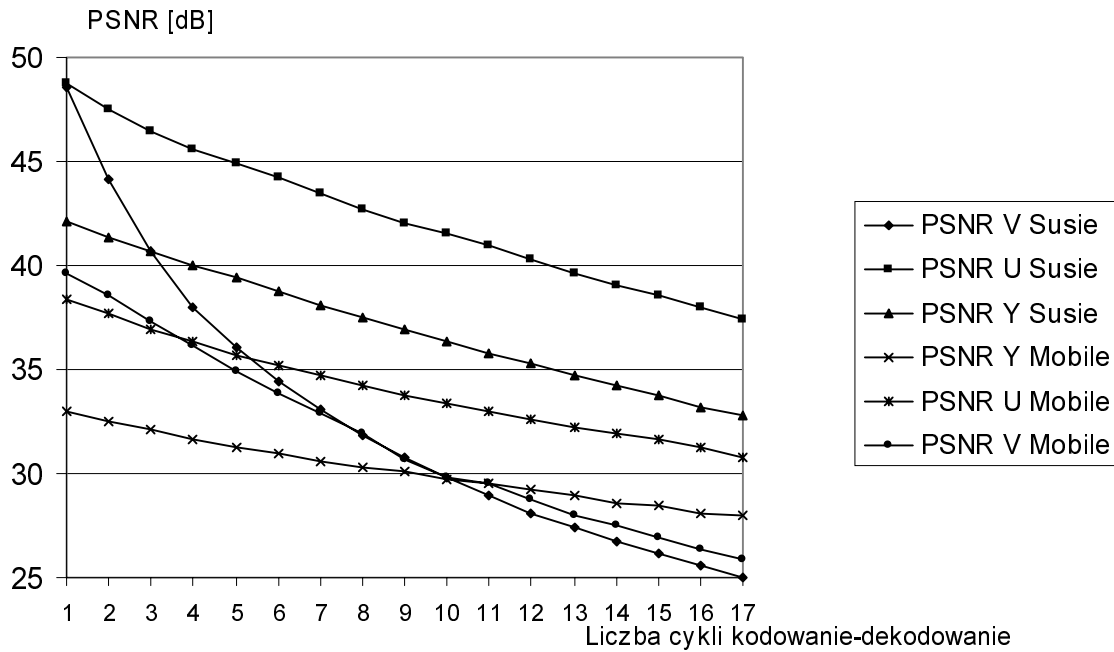
Największa wartość współczynnika PSNR osiągnięta jest przy 2 obrazach typu B pomiędzy kolejnymi P. Oczywistym faktem jest, że niewykorzystanie obrazów typu B (przypadek $M=1$) powoduje znaczny spadek jakości obrazu a tym samym współczynnika PSNR. Obrazy typu B umożliwiają osiągnięcie największego stosunku kompresji. Im większa część obrazów jest typu B, tym efektywniejsza powinna być kompresja. Jednak zbyt duża liczba obrazów B jest związana ze znaczną odległością między kolejnymi obrazami P, co oznacza spadek dokładności predykcji. Te dwa przeciwstawne mechanizmy implikują istnienie optymalnej liczby obrazów B pomiędzy kolejnymi obrazami P.

6. Wpływ wielokrotnego kodowania na jakość obrazu

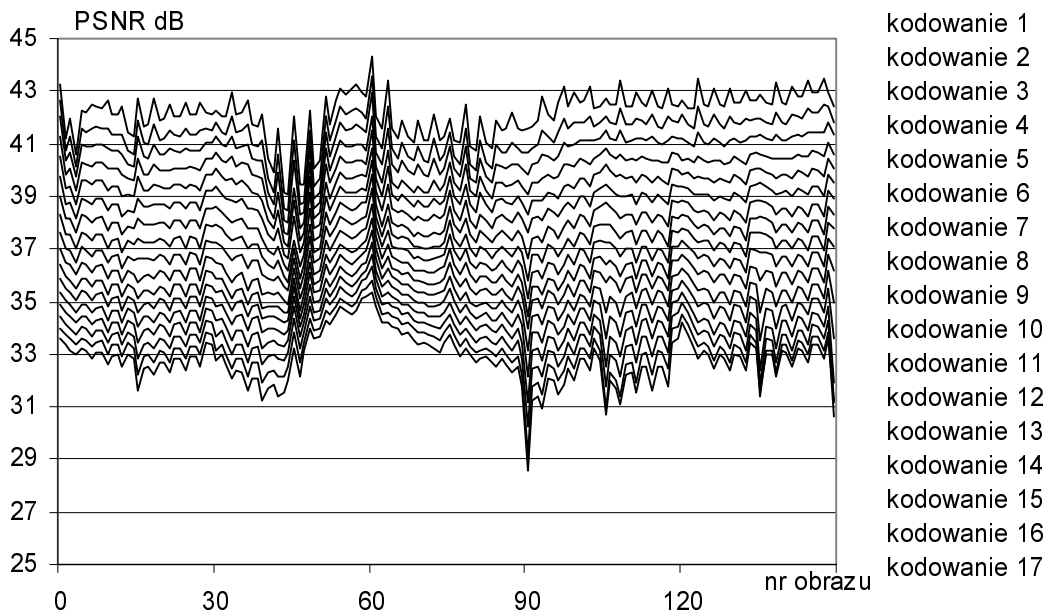
W przeprowadzonych eksperymentach dokonywano wielu cykli kodowania i dekodowania sekwencji, a następnie badano jakość zdekodowanych obrazów.

Badania zostały przeprowadzone przy następujących założeniach:

- liczba obrazów w grupie GOP: $N=15$,
- odległość między obrazem I a P, $M=3$,
- kolejność I B B P B B P
- przepływność binarna kanału 5Mbit/s.



Rys.8. Wartość uśredniona współczynnika PSNR sekwencji "Susie" i "Mobile and calendar" przy wielokrotnym kodowaniu



Rys.9. Wartość współczynnika PSNR składowej luminancji sekwencji „Susie” przy wielokrotnym kodowaniu

Nie istnieje tzw. strumień korzeniowy dla którego wartość współczynnika PSNR jest stała niezależna od dalszego kodowania i dekodowania. Wartość PSNR spada przy każdym kodowaniu, najbardziej przy początkowych kodowaniach zmniejszając spadek przy dalszych. Sugerowałoby to, że można dojść do strumienia korzeniowego. Niestety badania pokazały, że spadek współczynnika

przy dalszym kodowaniu nadal trwa. Składnia strumienia wymaga pewnej liczby bitów przeznaczonych na dane sterujące i informujące dekodera o parametrach sekwencji. Można by strumień w takim rozumieniu traktować jako korzeniowy. Podczas kodowania obraz poprzednio zdekodowany traktowany jest jak obraz źródłowy. Sekwencja nadal posiada tę samą informację o ruchu zmniejsza się tylko liczba szczegółów w obrazie wymaganych do przesłania.

7. Podsumowanie

W pracy przedstawiono wpływ doboru parametrów procesu kodowania na uzyskiwaną efektywność kompresji wyrażoną zależnością między stopniem kompresji a jakością zdekodowanej sekwencji. Należy podkreślić, że zależności te są bardzo silne i w dodatku znacznie modyfikowane przez charakter badanej sekwencji telewizyjnej.

Literatura

1. International Telecommunication Union „Transmission of non-telephone signals – Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Video” ITU-T Recommendation H.262, ITU-T 7/95
2. Chad Fogg „MPEG-2 Frequently Asked Questions List”, MPEG Software Simulation Group Draft 3.4, June 18, 1994
3. Haskell B.G., Puri A. Netraval A.N., "Digital video: an introduction to MPEG-2", Chapman & Hall, 1997
4. Maćkowiak Sławomir, „Kodek telewizji cyfrowej standardu MPEG-2” Praca dyplomowa-magisterska, Politechnika Poznańska, Instytut Elektroniki i Telekomunikacji, Poznań 1997
5. Strona WWW [html://www.mpeg.org/MSSG/tm5](http://www.mpeg.org/MSSG/tm5)