

Skalowalne kodowanie cyfrowych sygnałów wizyjnych

Sławomir Maćkowiak

Zastosowanie technik kompresji cyfrowych sygnałów wizyjnych umożliwiło powstanie nowych usług multimedialnych, takich jak: telewizja cyfrowa, w tym telewizja wysokiej rozdzielczości (HDTV) i telewizja interaktywna, wideotelefony, wideokonferencje, elektroniczna poczta multimedialna, wiadomości multimedialne w telefonii bezprzewodowej, interaktywne gry komputerowe. Liczne potencjalne zastosowania spowodowały koncentrację badań w zakresie metod kompresji.

Na całym świecie prowadzi się badania mające na celu zwiększenie efektywności kompresji, jednak największe znaczenie zyskują prace mające na celu uzupełnianie algorytmów kompresji tak, by wykazywały nowe cechy funkcjonalne, np. skalowalność.

Przedstawiona rozprawa dotyczy kompresji cyfrowych sekwencji wizyjnych, a szczegóły rozwiązania stanowią propozycję usprawnienia standardu MPEG-2 powszechnie stosowanego w systemach telewizji cyfrowej.

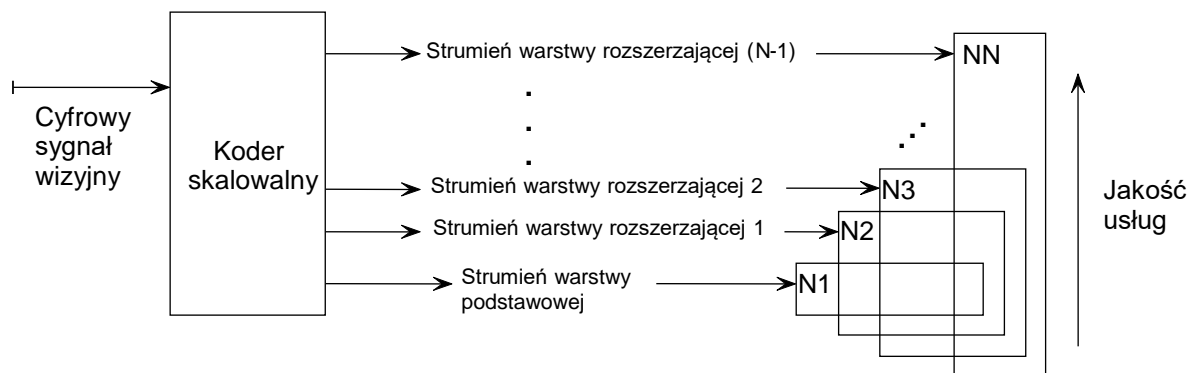
Kompresja skalowalna

Powstający w trakcie kompresji skalowanej strumień binarny kodujący obrazy jest podzielony na: **strumień warstwy podstawowej**, który można niezależnie dekodować i który reprezentuje sekwencje obrazów o małej rozdzielczości lub niskiej jakości oraz **strumień warstwy rozszerzającej** reprezentujące dodatkową informację potrzebną do odtworzenia obrazów o pełnej rozdzielczości i pełnej jakości. Wydzielenie w transmitowanym sygnale pewnej informacji podstawowej umożliwia zwiększenie odporności na błędy transmisji. Warstwa podstawowa może zostać zakodowana z większą nadmiarowością, podczas gdy informacje umożliwiające odtworzenie sekwencji o pełnej jakości będą odbierane prawidłowo tylko w przypadku braku zakłóceń w kanale. Warstwy rozszerzające mogą zostać także pominięte w czasie transmisji w warunkach dużego nasilenia ruchu w sieci.

Wiele z przedstawionych usług może być znacznie bardziej rozbudowanych i efektywniej świadczonych przy korzystaniu ze skalowalności, np. telewizja cyfrowa może docierać do szerszego kręgu odbiorców przy transmisji jednego sygnału zakodowanego przy użyciu koderów skalowalnych. Odbiorniki podłączone do systemów kablowych lub odbierające sygnał naziemnej telewizji cyfrowej o prędkości transmisji ok. 4-5Mb/s odbierają strumień warstwy podstawowej i rozszerzającej. Dekodując oba strumienie odbiorniki odbierają sygnał o pełnej jakości i pełnej rozdzielczości. Odbiorniki mobilne wykorzystujące systemy bezprzewodowe UMTS, założmy o przepustowości kanału 1Mb/s, odbierają tylko strumień warstwy podstawowej reprezentujący obrazy małej rozdzielczości. Jeden skalowalny sygnał wizyjny

dociera do szerokiego kręgu odbiorców, umożliwiając oglądanie ulubionych programów praktycznie w każdym dostępnym miejscu.

Konieczność dopasowania procesu kompresji cyfrowych sygnałów wizyjnych do warunków narzuconych przez heterogeniczne sieci telekomunikacyjne (rys.1), a także wymaganie zwiększenia odporności na błędy transmisji stymulują zainteresowanie skalowalnym kodowaniem sekwencji wizyjnych.



Rys.1. System wielowarstwowego skalowalnego kodowania sygnału wizyjnego w sieci heterogenicznej zawierającej podsieci N_1, \dots, N_N charakteryzujące się różnymi parametrami

W kodowaniu skalowalnym całkowita prędkość transmisji B_s jest sumą prędkości transmisji wszystkich warstw

$$B_s = B_b + B_{e_1} + \dots + B_{e_{N-1}},$$

gdzie: B_b - prędkość transmisji warstwy podstawowej,

$B_{e_1}, \dots, B_{e_{N-1}}$ - prędkość transmisji warstw rozszerzających 1, ..., N-1.

W praktyce, prędkość transmisji B_s jest większa niż prędkość B_{SL} wymagana przy transmisji danych podczas kodowania nieskalowalnego, tj. $B_s > B_{SL}$. Różnica $B_s - B_{SL}$ reprezentuje koszt wprowadzenia skalowalności i wyraża zwiększenie prędkości transmisji związane ze skalowalnością.

Teza i metody

Istniejące standardy (MPEG-2 i MPEG-4) określają bardzo nieefektywne kodery skalowalne. Dlatego odnotowuje się duże zainteresowanie poszukiwaniami efektywnych algorytmów skalowalnej kompresji sygnałów wizyjnych.

Celem pracy jest zaproponowanie modyfikacji istniejących algorytmów kompresji cyfrowych sygnałów wizyjnych w celu uzyskania systemów skalowalnego kodowania charakteryzujących się możliwie niewielkim zwiększeniem prędkości transmisji związanej ze skalowalnością. Jako technikę odniesienia użyto międzyobrazowe kodowanie predykcyjne z kompensacją ruchu, wykorzystujące blokowe kodowanie transformatowe. Założono wysoki

poziom zgodności ze standardami MPEG - koder skalowalny będzie wykorzystywał bloki funkcjonalne obecne w koderze standardu MPEG-2.

Główna teza dysertacji jest następująca: możliwa jest poprawa efektywności skalowalnego kodowania sygnału wizyjnego przy użyciu skalowalności przestrzenno-czasowej. Możliwe jest osiągnięcie dobrej efektywności, konkurencyjnej w stosunku do metod ujętych w standardzie MPEG-2, jak i do metod obecnie poddawanych standaryzacji.

Szczegółowym celem rozprawy jest ulepszenie algorytmu kodowania zdefiniowanego przez standard MPEG-2.

Skalowalność przestrzenno-czasowa

Większość rozwiązań znanych z literatury oparta jest o jeden typ skalowalności. Jednakże uniwersalna technika kodowania skalowalnego powinna łączyć różne typy. Jeden typ skalowalności nie może obsłużyć szerokiego zakresu przepływności w sieciach oraz szerokiego wyboru terminali o różnych parametrach. Istnieje także ograniczenie w ilości warstw.

W rozprawie rozważa się skalowalność przestrzenno-czasową, opisującą funkcjonalność systemu kompresji sygnału wizyjnego, w którym warstwa podstawowa (warstwa małej rozdzielczości) odpowiada obrazom o zredukowanej rozdzielczości przestrzennej i czasowej. Warstwa rozszerzająca jest używana do transmisji informacji potrzebnej do odtworzenia obrazu o pełnej rozdzielczości przestrzennej i czasowej.

Wielopętlowe kodery ze skalowalnością przestrzenno-czasową

W pracy włożono wielki wysiłek w wyszukanie właściwego algorytmu kodowania i odpowiadających mu struktur kodera. W tym celu przebadano wiele wariantów i struktur kodera. Dla każdego stworzono oddzielny model weryfikacyjny i przetestowano go z telewizyjnymi sekwencjami testowymi.

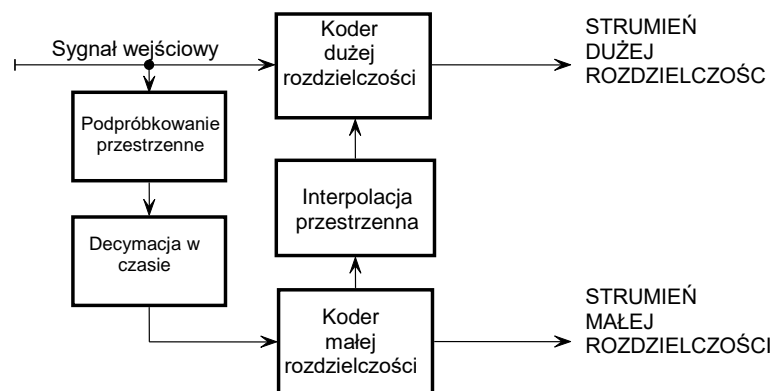
W trakcie prac nad efektywną strukturą kodera skalowalnego autor przebadął system z trójwymiarowym zestawem filtrów użytym do analizy przestrzenno-czasowej. System ten był bardzo złożony, a jego bloki nie były zgodne ze standardami MPEG. Dlatego autor przedstawił tylko ogólną strukturę tego kodera.

W celu zwiększenia zgodności z technologią ujętą w standardzie MPEG-2, autor rozwinął koncepcję dwuwymiarowego kodera subpasmowego opartą na podziale danych obrazów typu B między warstwy z kompensacją ruchu. W tym rozwiązaniu każdy obraz dzielony jest na cztery subpasma. Subpasmo LL odpowiadające niskim częstotliwościom przestrzennym jest umieszczane w warstwie podstawowej, podczas gdy trzy pozostałe subpasma (LH, HL i HH) są łącznie transmitowane w warstwie rozszerzającej.

Takie rozwiązanie umożliwiło poprawę efektywności kodowania, czasami jednak prowadziło do wytworzenia warstwy podstawowej o zbyt dużej prędkości transmisji w stosunku do warstwy rozszerzającej. W celu uniknięcia wspomnianych wad i trudności, autor opracował system ze skalowalnością przestrzenno-czasową niewykorzystujący rozkładu subpasmowgo. Zaproponowany koder skalowalny oparty jest o blokowe kodowanie transformatowe oraz niezależną estymację i kompensację ruchu w warstwie podstawowej i rozszerzającej.

Koder złożony jest z dwóch koderów cząstkowych (rys. 2):

- niezależnie działającego kodera warstwy podstawowej, w pełni zgodnego ze standardem MPEG-2;
- kodera warstwy rozszerzającej, który wykorzystując informacje z warstwy podstawowej dokonuje kodowania obrazów o pełnej rozdzielczości przestrzennej i czasowej (zmodyfikowany koder standardu MPEG-2).

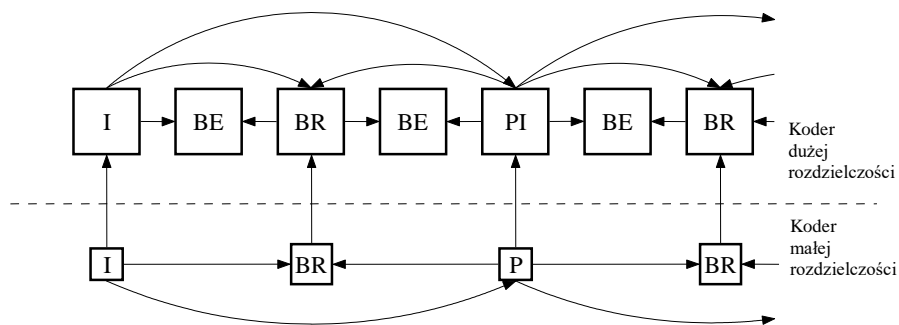


Rys. 2. Podstawowy schemat blokowy kodera hybrydowego z funkcjonalnością skalowalności przestrzenno-czasowej

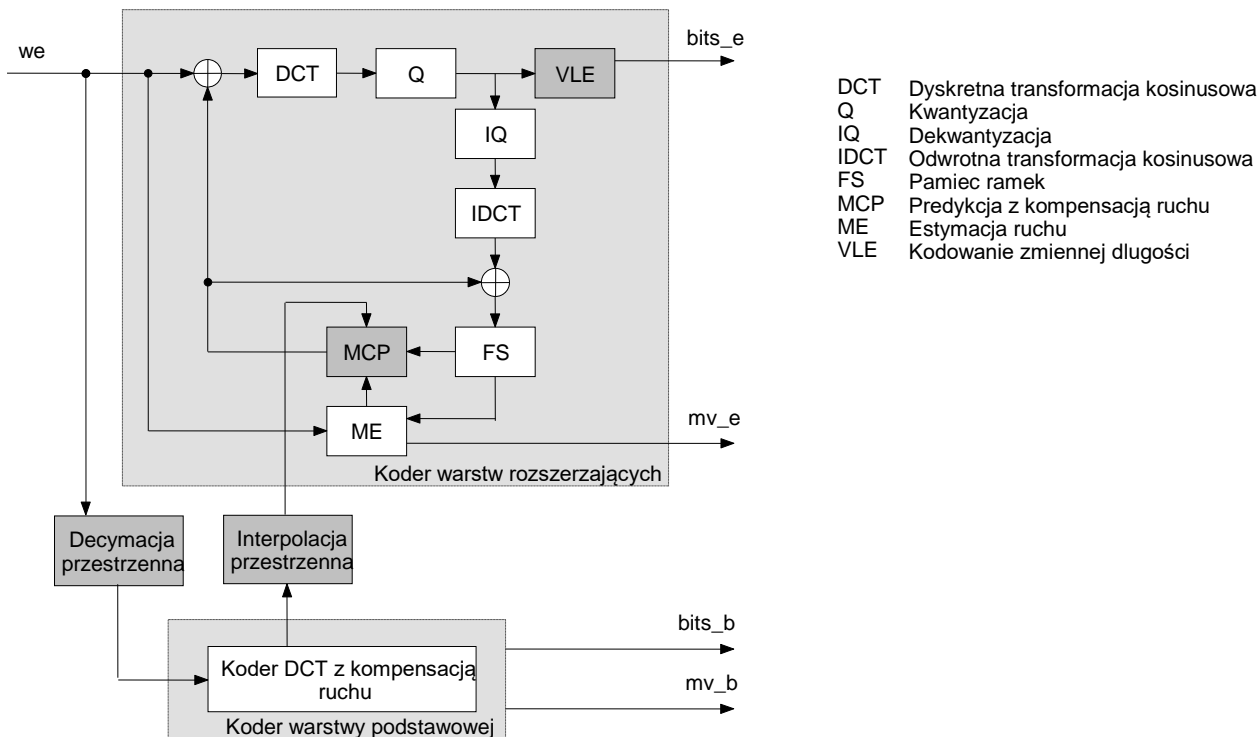
W proponowanym rozwiązaniu warstwa podstawowa odpowiada strumieniowi obrazów o zredukowanej rozdzielczości zarówno czasowej, jak i przestrzennej. Natomiast warstwa rozszerzająca jest używana do transmisji informacji potrzebnych do odtworzenia sekwencji wizyjnej o pełnej rozdzielczości czasowej i przestrzennej.

Redukcja rozdzielczości czasowej jest osiągnięta poprzez podział strumienia obrazów typu B między warstwę podstawową a warstwę rozszerzającą. W tym celu zaproponowano dwa rodzaje obrazów typu B:

- obrazy BE kodowane w warstwie rozszerzającej,
- obrazy BR kodowane w obu warstwach (rys.3).



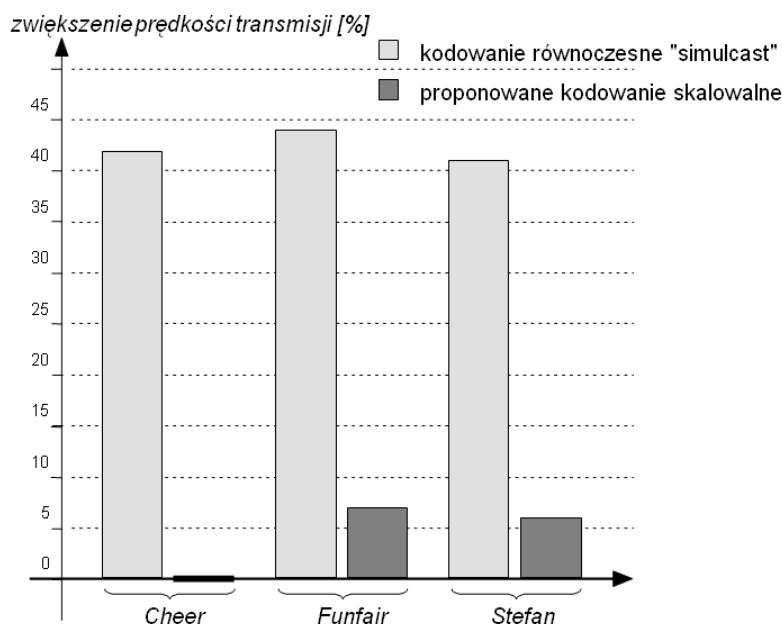
Rys. 3. Rozmieszczenie obrazów w obu warstwach



Rys. 4. Schemat blokowy dwuwarstwowego kodera skalowalnego (dla kodowania międzyobrazowego)

Zaproponowany koder (rys. 4) stosuje niezależną kompensację ruchu w każdej z warstw. Niezależna estymacja i kompensacja ruchu w warstwach skutkuje wyższą efektywnością kodowania. Autor przeprowadził wiele eksperymentów dowodzących tej tezy. Niezależna kompensacja ruchu w warstwach umożliwia uzyskanie optymalnych wektorów ruchu dla makrobloków, które mogą być użyte do bardzo efektywnej predykcji z kompensacją ruchu.

Wyniki przedstawione w pracy dowodzą dobrej efektywności proponowanego kodera. Przy tej samej jakości, zwiększenie prędkości transmisji związanej ze skalowalnością wynosi jedynie od 5 do 15 proc (Rys.5).



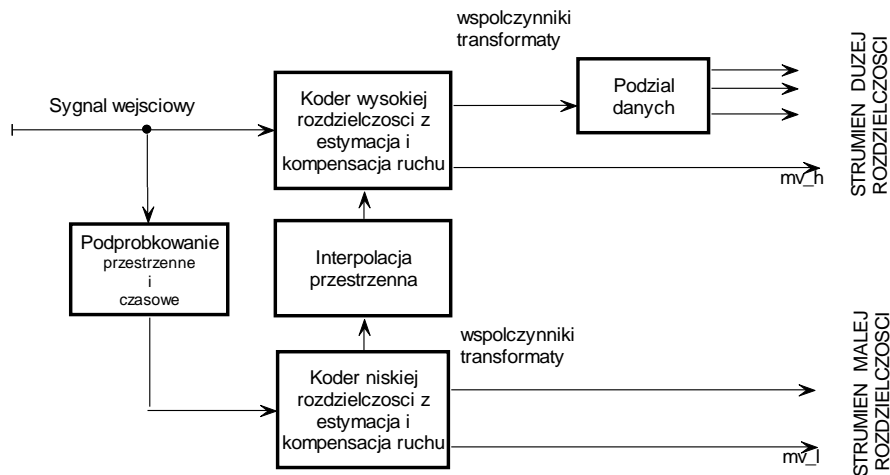
Rys. 5. Zwiększenie prędkości transmisji przy porównaniu kodowania równoczesnego i proponowanej metody kodowania skalowalnego. Zwiększenie prędkości transmisji jest mierzone jako procent prędkości transmisji dla kodowania nieskalowalnego, przypadek dla prędkości transmisji 5 Mb/s

Skalowalność drobnoziarnista z kompensacją ruchu

Klasyczne rozwiązania dla dwu- i trójwarstwowej skalowalności są mało efektywne lub nie dość elastyczne do wspierania aplikacji strumieniowych. Zaproponowane rozwiązania przedstawione powyżej mogą być rozszerzone do przypadku wielowarstwowego.

Standard MPEG-4 określa skalowalność drobnoziarnistą (*FGS – Fine Granularity Scalability*), które jest narzędziem precyzyjnego dopasowania prędkości transmisji do przepustowości kanału. Niestety, kodery MPEG-4 FGS wykazują znaczne zwiększenie prędkości transmisji w porównaniu do odpowiedniego kodowania nieskalowalnego.

Autor pokazał, że możliwe jest otrzymanie skalowalności drobnoziarnistej w schemacie opartym o zmodyfikowaną wersję klasycznego kodera MPEG-2. Skalowalność drobnoziarnista może być osiągnięta poprzez transmitowanie pierwszych kilku niezerowych współczynników DCT w każdym bloku.

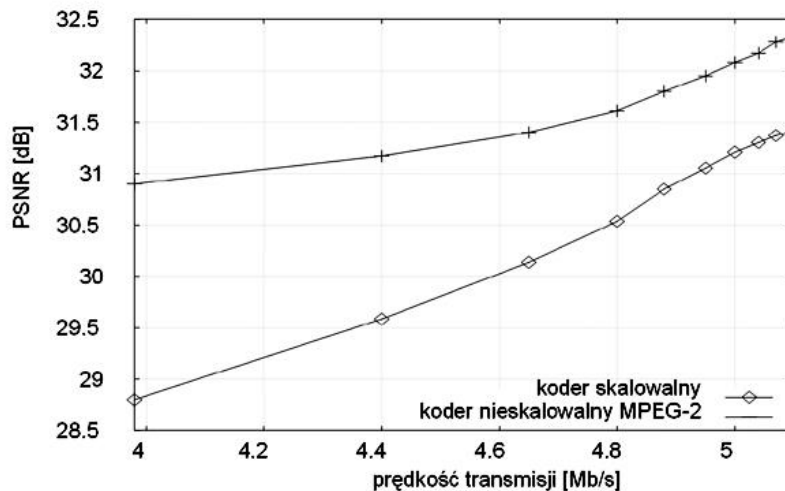


Rys. 6. Struktura kodera wielowarstwowego: (mv_l i mv_h – odpowiednio: wektory ruchu generowane w warstwie małej i dużej rozdzielczości)

Konstrukcja kodera (Rys. 6) wielowarstwowego jest oparta na schemacie kodera dwuwarstwowego.

Dane, które są wytwarzane przez oba kodery, mogą być przypisane do kolejnych warstw. Podstawowy segment danych zawiera wszystkie nagłówki i wektory ruchu warstwy rozszerzającej mv_h . Następne segmenty zawierają współczynniki DCT, które są zakodowane jako pary (run, level). Każda warstwa wymaga powtórzenia nagłówek warstw makrobloków w celu zapewnienia im synchronizacji w przypadku wystąpienia błędów.

W wyniku takiej organizacji strumienia danych można kontrolować strumień każdej warstwy.



Rys. 7. Skalowalność drobnoziarnista w kodowaniu skalowalnym dwuwarstwowym (dolna krzywa) w porównaniu do kodowania nieskalowalnego (górną krzywą). Sekwencja testowa *Funfair*, prędkość transmisji około 5 Mb/s, prędkość transmisji warstwy podstawowej około 1.66 Mb/s, długość GOP=12

Podsumowanie

Główne osiągnięcia pracy to:

- opracowanie struktur koderów skalowalnych z niezależnymi pętlami z niezależną kompensacją i estymacją ruchu;
- propozycja koderów ze skalowalnością drobnoziarnistą;
- szczegółowe, eksperymentalne przebadanie cech struktur zaproponowanych koderów;
- porównanie własności różnych zaproponowanych modyfikacji struktur koderów skalowalnych.

Inne oryginalne osiągnięcia dysertacji to:

- rozwinięcie struktury koderów skalowalnych z trójwymiarowym zespołem filtrów do analizy przestrzenno-czasowej;
- dwuwymiarowy subpasemowy system skalowalny, w którym część obrazów typu B jest przypisana do warstwy rozszerzającej;
- analiza wyników eksperymentalnych kodowania skalowalnego z dwoma różnymi typami macierzy kwantyzacji i zmodyfikowanym sposobem skanowania współczynników DCT;
- zmodyfikowana technika kodowania sygnału ΔLL za pomocą płaszczyzn bitowych;
- analiza struktury koderów subpasemowych wykorzystujących korelacje pomiędzy subpasemami;
- prosty empiryczny model globalnego sterowania parametrami koderów;
- analiza strumienia w trybie wewnątrzobrazowym dla kodowania nieskalowalnego i skalowalnego;
- modele weryfikacyjne proponowanych koderów skalowalnych.

Dzięki zastosowaniu w proponowanych koderach dekompozycji przestrzenno-czasowej osiągnięto dobrą efektywność kodowania, co zostało dowiedzione przez eksperymentalne porównanie z koderami standardu MPEG-2. Pokazano, że możliwe jest osiągnięcie efektywnego systemu kodowania ze skalowalnością przestrzenno-czasową przy akceptowalnym zwiększeniu prędkości transmisji związanym z wprowadzeniem skalowalności.

Artykuł jest streszczeniem rozprawy doktorskiej dr inż. Sławomira Maćkowiaka

*Autor jest pracownikiem Instytutu Elektroniki i Telekomunikacji Politechniki
Poznańskiej*