

Politechnika Poznańska

Wydział Elektroniki i Telekomunikacji

Katedra Telekomunikacji Multimedialnej i Mikroelektroniki



Streszczenie rozprawy doktorskiej

**“Kodowanie informacji o ruchu dla kompresji  
sekwencji trójwymiarowych”**

Jacek Konieczny

Promotor: prof. dr hab. inż. Marek Domański

Poznań, 2012



# Spis treści

<b>Spis treści.....</b>	<b>3</b>
<b>1. Zakres i cel rozprawy.....</b>	<b>4</b>
1.1. Zakres rozprawy.....	4
1.2. Cele i teza rozprawy.....	8
1.3. Metodologia badań.....	9
1.4. Przegląd pracy.....	11
<b>2. Zasadnicze wyniki rozprawy.....</b>	<b>12</b>
2.1. Stan obecny.....	12
2.2. Oryginalne metody predykcji między-widokowej wykorzystujące informację o głębi.....	13
2.3. Autorskie techniki kodowania informacji o ruchu wykorzystujące zaproponowane metody predykcji między-widokowej.....	15
2.4. Wpływ zaproponowanych metod predykcji na efektywność kompresji kodeków wielowidokowych.....	19
<b>3. Podsumowanie pracy oraz wnioski .....</b>	<b>23</b>
3.1. Podsumowanie pracy.....	23
3.2. Oryginalne osiągnięcia zaprezentowane w rozprawie.....	26
3.3. Ogólne wnioski.....	27
<b>Literatura .....</b>	<b>30</b>
<b>Dokumentacja dorobku naukowego autora .....</b>	<b>32</b>

# 1. Zakres i cel rozprawy

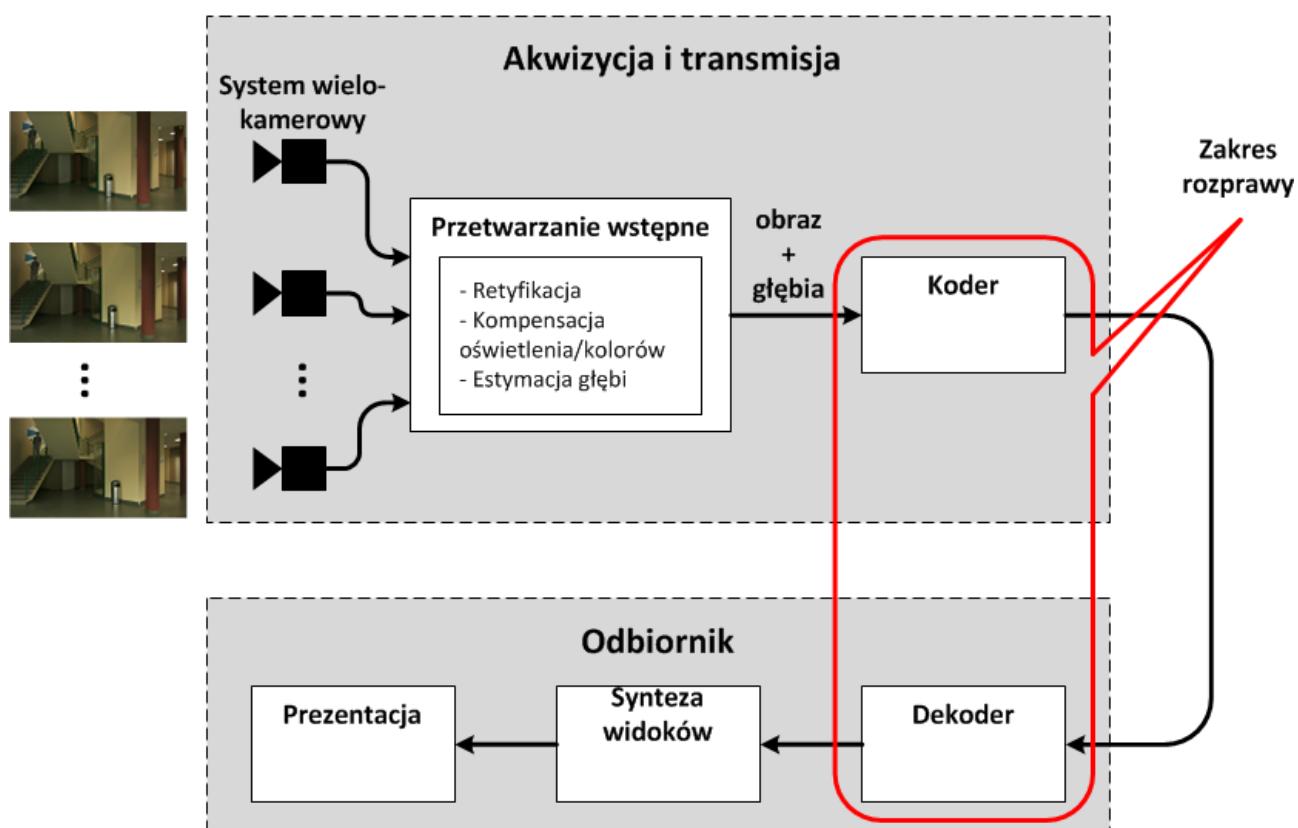
## 1.1. Zakres rozprawy

Rozprawa doktorska dotyczy zagadnienia **poprawy kompresji trójwymiarowych sekwencji wizyjnych**. Zagadnienie to ma fundamentalne znaczenie dla rozwoju usług wykorzystujących trójwymiarowe sekwencje wizyjne, takich jak telewizja trójwymiarowa nowej generacji (3D-TV) oraz telewizja swobodnego punktu widzenia (FTV). Wspomniane nowe typy systemów wizyjnych budzą ogromne zainteresowanie, wykraczając zdecydowanie poza możliwości oferowane przez obecne systemy stereoskopowe dostępne komercyjnie. Systemy FTV pozwalają użytkownikowi na zmianę punktu obserwacji sceny niezależnie od położenia fizycznych kamer użytych do jej akwizycji. Treść prezentowana dla widoków wirtualnych uzyskiwana jest w tym przypadku poprzez wykorzystanie algorytmów syntezy, z użyciem informacji o teksturze i geometrii sceny wizyjnej. Z drugiej strony, technologia trójwymiarowa nowej generacji obecna w systemach 3D-TV pozwala na doświadczanie bardziej realistycznego efektu głębi. W rezultacie możliwe staje się odtworzenie efektu paralaksy wynikającej z ruchu, lub też percepcja głębi stereoskopowej bez konieczności stosowania specjalnych okularów. Przewiduje się, iż wykorzystanie trójwymiarowych sekwencji wizyjnych nowej generacji zaowocuje licznymi korzyściami, zwłaszcza w obszarze nauki i rozrywki. Dlatego też, w przyszłości, trójwymiarowe sekwencje wizyjne oferowane będą w różnego typu aplikacjach, jak choćby telewizji rozsiewczej, transmisji strumieniowej w Internecie, czy usługach mobilnych. Takie systemy wizyjne przyszłości wymagać będą jednak transmisji ogromnych strumieni danych do odbiorcy. Oznacza to, iż możliwości istniejących technik transmisji danych mogą okazać się niewystarczające, zwłaszcza w przypadku usług mobilnych. W rezultacie istnieje silna motywacja by stworzyć nowe techniki kompresji trójwymiarowych sekwencji wizyjnych, oferujące większy stopień kompresji dzięki bardziej efektywnemu wykorzystaniu istniejącej korelacji międzyobrazowej w sekwencjach wizyjnych.

We wspomnianych systemach multimedialnych nowej generacji przetwarzaniu poddawane są wielowidokowe sekwencje wizyjne [Smo06, Kauf07], a więc sekwencje przedstawiające tą samą scenę wizyjną zarejestrowaną jednocześnie z wielu punktów obserwacji. Dodatkowo, w przypadku licznych aplikacji trójwymiarowych nowej generacji, obrazowi wielowidokowemu towarzyszy również informacja o strukturze geometrycznej sceny w postaci np. głębi stereoskopowej. Ten

szczególny rodzaj sekwencji wielowidokowych określany jest mianem trójwymiarowych sekwencji wizyjnych w formacie MVD (ang. *Multiview Video plus Depth*) [Smo07]. **Niniejsza rozprawa dotyczy właśnie zagadnień związanych z przetwarzaniem trójwymiarowych sekwencji wizyjnych formacie MVD.**

Typowy system przetwarzania wielowidokowych sekwencji wizyjnych składa się z części odpowiedzialnych za akwizycję obrazu, transmisję oraz prezentację (Rys. 1.1). Informacja o głębi zazwyczaj wykorzystywana jest w takim systemie do syntezy widoków wirtualnych z użyciem dedykowanych algorytmów projekcji [Kauf07]. W rezultacie, po stronie odbiorczej takiego systemu wymagana jest dostępność zarówno wybranych widoków przetwarzanej sekwencji wielowidokowej, parametrów opisujących system akwizycji, jak również informacji o głębi. Zagadnieniem kluczowym jest zatem dostarczenie powyższych danych do odbiornika w sposób możliwie wydajny, co realizowane jest w blokach kodera i dekodera. Dlatego **niniejsza rozprawa dotyczy części systemu wielowidokowego odpowiedzialnej za proces kodowania i dekodowania.**



Rys. 1.1. System wielowidokowy.

Dotąd podjęto wiele starań w celu opracowania bardziej efektywnych technik kompresji sekwencji wielowidokowych i trójwymiarowych. W sekwencjach naturalnych istnieje wysoki

stopień redundancji czasowej i przestrzennej w kodowanej treści. Wykorzystywane jest to podczas kodowania w celu zwiększenia stopnia kompresji, przy znikomym wpływie na subiektywną jakość obrazu zdekodowanego. Najbardziej wydajna i powszechnie stosowana klasa kodeków obrazu ruchomego, zwana kodekami hybrydowymi [Dom98, Ska93, Shi00, Ohm04, Dom10], wykorzystuje mechanizmy oparte o predykcję z kompensacją ruchu oraz kodowanie tranformatowe w celu uzyskania znacznego stopnia kompresji. Kompensacja ruchu wykonywana jest zazwyczaj w niewielkich prostokątnych blokach. Wektory ruchu obliczone w ten sposób dla każdego bloku w koderze umieszczane są w strumieniu bitowym i przesyłane do dekodera. W rezultacie, strumień bitowy wytwarzany przez typowy koder hybrydowy składa się z następujących trzech typów danych: wektorów ruchu, współczynników transformaty błędu predykcji oraz dane sterujące [Dom98, Ska93, Ric02]. Oznacza to, iż redukcja części strumienia reprezentującej wektory ruchu powinna przyczynić się do znacznego zwiększenia stopnia kompresji takiego koder.

W przypadku wielowidokowych sekwencji wizyjnych występuje także dodatkowa redundancja przestrzenna, wynikająca z wysokiej korelacji pomiędzy treścią zarejestrowaną dla poszczególnych widoków [Feck05, Merk07, Su06]. Wspomniana korelacja między-widokowa może zostać skutecznie wykorzystana do ograniczenia ilości danych przesyłanych do odbiornika w systemie wielowidokowym. Prostą metodą wykorzystującą podobieństwo obrazów pochodzących z sąsiednich widoków jest użycie ich jako dodatkowych obrazów odniesienia w predykcji między-ramkowej [ISO11, Vet11]. Jednak w systemach trójwymiarowych nowej generacji możliwe staje się zastosowanie dużo bardziej zaawansowanych metod. Obecność informacji o głębi, opisującej geometrię sceny, umożliwia wykorzystanie wyrafinowanych metod predykcji opartych o projekcję w przestrzeni trójwymiarowej. W rezultacie, pozwala to na zwiększenie stopnia kompresji uzyskiwanego przez koder wielowidokowy [Mart06, Shim07]. Podążając tym tokiem rozumowania, **niniejsza rozprawa przedstawia wyniki badań mających na celu zwiększenie stopnia kompresji trójwymiarowych sekwencji wizyjnych poprzez wykorzystanie dostępnej informacji o głębi w celu efektywnej między-widokowej predykcji informacji o ruchu.**

Perspektywa rosnącego zapotrzebowania na technikę kompresji wielowidokowych i trójwymiarowych sekwencji wizyjnych zmotywowała komisję MPEG (ang. *Moving Picture Experts Group*) działająca z ramienia Międzynarodowej Organizacji Standaryzacyjnej ISO (ang. *International Organization for Standardization*) do rozpoczęcia w 2004 roku prac nad nową normą kodowania o nazwie Multiview Video Coding (MVC) [Pfis04]. Rezultatem powyższych prac jest nowa norma kodowania sekwencji wielowidokowych, oparta o technologię kodowania AVC (ang. *Advanced Video Coding*), ustanowiona przez ITU-T oraz ISO/IEC w 2009 roku [ISO11, Ric10].

Podstawowym podejściem przyjętym w normie MVC jest zastosowanie predykcji między-widokowej z kompensacją rozbieżności, opartej o mechanizmy podobne do predykcji z kompensacją ruchu wykorzystywanej w normie AVC. Ten prosty pomysł pozwolił na uzyskanie znacznych zysków kodowania w porównaniu z niezależnym kodowaniem poszczególnych widoków sekwencji wielowidokowej (ang. *simulcast*). Niemniej jednak, stopień kompresji wynikający z zastosowania MVC nie spełnia wymagań aplikacji trójwymiarowych nowej generacji. Co więcej, MVC nie opisuje żadnych dedykowanych metod kompresji sekwencji trójwymiarowych w formacie MVD.

W rezultacie, komisja MPEG rozpoczęła w 2010 roku prace nad stworzeniem nowych technik kodowania sekwencji trójwymiarowych w formacie MVD, których zadaniem jest sprostanie wymaganiom przyszłych trójwymiarowych systemów wizyjnych [MP11b]. Autor tejże rozprawy aktywnie włączył się w prace komisji MPEG dotyczące opracowania nowych, bardziej wydajnych metod kodowania informacji o ruchu w sekwencjach trójwymiarowych. Wyniki prac autora zostały opublikowane w dokumentach zgłoszonych do komisji MPEG, a niektóre z rozwiązań zaproponowanych w niniejszej rozprawie włączone w rezultaty prac komisji MPEG. W szczególności, algorytmy opisane w niniejszej rozprawie zostały zastosowane w technice kodowania sekwencji trójwymiarowych zgłoszonej przez Politechnikę Poznańską w odpowiedzi na wezwanie komisji MPEG do zgłaszania propozycji technologii będących podstawą przyszłego standardu kodowania. Propozycja Politechniki Poznańskiej uzyskała znakomite rezultaty, plasując się w ścisłej czołówce wśród technik zgłoszonych przez ośrodki badawcze z całego świata.

Algorytmy między-widokowej predykcji informacji o ruchu stanowią ważne zagadnienie w obszarze badań dotyczących poprawy efektywności kompresji wielowidokowych sekwencji wizyjnych. Pomimo prowadzonych intensywnych prac badawczych, nadal nie opracowano ostatecznego rozwiązania dla efektywnej reprezentacji informacji o ruchu w sekwencjach trójwymiarowych, które wykorzystywałoby dostępność informacji opisującej położenie obiektów w scenie. Dlatego też, **w niniejszej rozprawie poruszono problem zwiększenia stopnia kompresji trójwymiarowych sekwencji wizyjnych poprzez wykorzystanie nowych, zaproponowanych przez autora metod między-widokowej predykcji informacji o ruchu w oparciu o dostępną informację o głębi.**

## 1.2. Cele i teza rozprawy

**Celem rozprawy** jest opracowanie nowych technik kompresji trójwymiarowych sekwencji wizyjnych w formacie MVD (ang. *Multiview Video plus Depth*). Zaproponowane zostaną nowe metody predykcji między-widokowej w sekwencjach wielowidokowych, pozwalające na zredukowanie prędkości bitowej wymaganej do transmisji takich sekwencji poprzez zapewnienie bardziej efektywnej reprezentacji informacji o ruchu oraz wykorzystanie dostępnej informacji o głębi. Przebadane zostaną nowe algorytmy i narzędzia umożliwiające poprawę efektywności kompresji przy niewielkim wzroście złożoności oraz wymagań stawianych wielowidokowy kodekom wizyjnym. Zaproponowane techniki zostaną zbadane eksperymentalnie w celu oceny ich wpływu na efektywność kompresji istniejących oraz przyszłych wielowidokowych kodeków wizyjnych.

W rozprawie przyjęto następujące **założenia**:

- jako podstawową technologię kompresji zastosowano hybrydowe kodeki wizyjne uznawane za reprezentatywne dla obecnego stanu wiedzy,
- informacja o głębi opisująca scenę wizyjną jest dostępna w dekodерze,
- zaproponowane techniki powinny zapewniać możliwie wysoki stopień kompatybilności z istniejącymi technikami kompresji uznawanymi za reprezentatywne dla obecnego stanu wiedzy.

Główne **tezy rozprawy** są następujące:

- Możliwe jest zwiększenie efektywności reprezentacji informacji o ruchu w kodowaniu trójwymiarowych sekwencji wizyjnych w formacie MVD poprzez wykorzystanie korelacji pomiędzy polami ruchu w sąsiednich widokach oraz dostępnej informacji o ruchu.
- Możliwe jest opracowanie technik reprezentacji informacji o ruchu, konkurencyjnych względem metod opisanych w literaturze i rozwijanych równoległe do prac autora.



### 1.3. Metodologia badań

Celem rozprawy jest zbadanie możliwości zwiększenia efektywności kompresji istniejących technik kompresji sekwencji wielowidokowych poprzez stworzenie nowych metod między-widokowej predykcji informacji o ruchu, wykorzystujących zaawansowane modelowanie sceny wizyjnej. W rezultacie, punktem wyjścia dla przeprowadzonych badań była analiza istniejących technik predykcji informacji o ruchu w koderach wielowidokowych. Szczególną uwagę poświęcono najbardziej zaawansowanym rozwiązaniom zaproponowanym podczas prac komisji MPEG nad normą MVC [ISO11, Yang09, Koo06]. Wspomniane techniki zostały przebadane eksperymentalnie pod kątem efektywności kodowania informacji o ruchu.

Wyniki powyższych eksperymentów pozwoliły na sformułowanie problemu predykcji oraz reprezentacji informacji o ruchu przy kodowaniu trójwymiarowych sekwencji wizyjnych wzbogaconych o informację o głębi. Zaproponowano nowe techniki między-widokowej predykcji informacji o ruchu w warunkach dostępności informacji o głębi. Ze względu na wysoki stopień złożoności problemu, proces tworzenia nowych technik predykcji został zdekomponowany na kilka etapów. Tak wypracowane rozwiązania zostały następnie zaimplementowane w oprogramowaniu modelowym i przebadane eksperymentalnie w celu sprawdzenia ich przydatności w opracowywanych algorytmach kodowania wielowidokowych sekwencji wizyjnych. Wnioski wyciągnięte z tychże eksperymentów zostały wykorzystane do usprawnienia zaproponowanych metod. Powyższy proces został powtórzony w kolejnych iteracjach.

W przeprowadzonych eksperymentach, jako kodeków odniesienia użyto następujących technik kodowania wielowidokowych sekwencji wizyjnych:

- MVC (Multiview Video Coding), opracowany przez komisję MPEG jako *aneks H* normy kodowania sekwencji wizyjnych AVC (ISO/IEC MPEG-4 część 10, ITU-T H.264) [ISO11, Chen09]. Kodek ten posłużył jako baza dla implementacji algorytmów zaproponowanych przez autora niniejszej rozprawy.
- JMVM (Joint Multiview Model), opracowany przez komisję MPEG podczas prac nad normą kodowania MVC [MP08, Pan08]. Kodek ten zawiera szereg dodatkowych narzędzi kodowania zaprojektowanych specjalnie dla potrzeb kodowania wielowidokowych sekwencji wizyjnych, m. in. narzędzie do między-widokowej predykcji informacji o ruchu o nazwie Motion Skip.
- MV-HEVC, wielowidokowy kodek wizyjny oparty o technikę kodowania nowej generacji HEVC (ang. *High Efficiency Video Coding*) [MP11, Dom12], opracowany oryginalnie przez

Politechnikę Poznańską [Dom11]. W kodeku tym wykorzystano wersję oprogramowania modelu testowego HEVC o nazwie HM 3.0 [MP11a] do stworzenia kodeka wielowidokowego działającego zgodnie ze schematem kompresji znanym z kodeka MVC. Tak zaprojektowany kodek oferuje mechanizmy między-widokowej predykcji znane z MVC, używając jednak rdzenia HEVC jako podstawowej techniki kodowania.

Techniki kompresji opisane powyżej zostały wykorzystane w badaniach zaprezentowanych w niniejszej rozprawie, ponieważ stanowią odzwierciedlenie najbardziej współczesnych trendów kodowania sekwencji wielowidokowych stosowanych na świecie oraz ze względu na fakt, iż autor rozprawy posiadał nieograniczony dostęp do ich kodu źródłowego, co umożliwiło z kolei ich modyfikację i rozbudowę o algorytmy zaproponowane przez autora.

Podczas badań eksperymentalnych przeanalizowano efektywność kodowania informacji o ruchu oraz całkowitą efektywność kompresji kodeka wizyjnego: istniejące techniki kodowania informacji o ruchu zostały porównane z oryginalnymi rozwiązaniami zaproponowanymi przez autora niniejszej rozprawy. W tym celu zbadano korelację pomiędzy polami ruchu będącymi wynikiem estymacji ruchu w koderze oraz predykcji między-widokowej, a także zmierzono prędkość bitową i jakość wynikowych strumieni wizyjnych. Jako miarę zniekształcenia użyto obiektywnej miary jakości PSNR oraz miary Bjontegaarda, stosowanych powszechnie w badaniach prowadzonych m. in. w komisji MPEG oraz VCEG (ang. *Video Coding Experts Group*) [Bjo01, Pat07]. Dla potrzeb określenia złożoności zaproponowanych metod, zmierzono czas wykonywania obliczeń przez badane kodeki wizyjne. Jest to metoda powszechnie wykorzystywana w badaniach prowadzonych w ramach prac komisji MPEG i JCT-VC (ang. *Joint Collaborative Team-Video Coding*) [Sul10, Sue10].

We wszystkich eksperymentalnych użyto wielowidokowych i trójwymiarowych sekwencji wizyjnych stanowiący zbiór testowy wykorzystywany w pracach standaryzacyjnych komisji MPEG, zróżnicowany pod względem występujących w nim typów ruchu i tekstur. Sekwencje te zostały wyselekcjonowane przez komisję MPEG w celu przeprowadzenia badań nad nowymi narzędziami i technikami kodowania sekwencji wielowidokowych oraz trójwymiarowych [MP11b]. Zakresy prędkości bitowych skompresowanych sekwencji wizyjnych wykorzystywane w eksperymentach również zostały dobrane tak, aby spełnić wymagania przyjęte przez komisję MPEG i VCEG dla celów porównawczych kodeków wizyjnych [Tan08, Bos11].

## 1.4. Przegląd pracy

Zagadnienia zaprezentowane w pracy ujęto w następujących rozdziałach.

W **rozdziale 2** przedstawiono podstawowe informacje dotyczące wielowidokowych systemów wizyjnych. Wyjaśniono zasady działania hybrydowych kodeków wizyjnych wraz z zagadnieniem estymacji ruchu oraz jego reprezentacji. Wprowadzono również miary korelacji pomiędzy polami ruchu, a także omówiono techniki kodowania wielowidokowych sekwencji wizyjnych, ze szczególnym naciskiem na najbardziej zaawansowane metody predykcji informacji o ruchu znane z literatury.

**Rozdział 3** zawiera opis dwóch autorskich technik predykcji między-widokowej przeznaczonych dla wielowidokowych kodeków wizyjnych. Rozważane są w nim różne warianty zaproponowanych metod, z uwzględnieniem zagadnień efektywności i złożoności.

W **rozdziale 4** umotywowano wykorzystanie między-widokowej predykcji informacji o ruchu w wielowidokowych kodekach wizyjnych. Przebadano korelację dla pól ruchu będących rezultatem predykcji między-widokowej oraz przeanalizowano eksperymentalnie udział poszczególnych elementów składni w strumieniu wytwarzanym przez wielowidokowy hybrydowy kodek wizyjny.

W **rozdziale 5** rozważane są sposoby wykorzystania oryginalnych metod między-widokowej predykcji informacji o ruchu zaproponowanych przez autora w istniejących i przyszłych wielowidokowych kodekach wizyjnych. Zaprezentowano różne warianty kodeków wielowidokowych opracowanych przez autora, oparte o najnowsze i najbardziej zaawansowane rozwiązania będące stanem obecnej wiedzy: kodeki wizyjne AVC oraz HEVC.

W **rozdziale 6** zaprezentowano wyniki eksperymentalne uzyskane dla różnych wariantów kodeków wizyjnych opisanych w rozdziale 5. Na podstawie uzyskanych rezultatów, rozważane są zarówno aspekty efektywności, jak i złożoności zaproponowanych implementacji.

**Rozdział 7** zawiera rozważania dotyczące możliwości wykorzystania metod między-widokowej predykcji informacji o ruchu zaproponowanych przez autora w przyszłych trójwymiarowych kodekach wizyjnych. Zaprezentowano również najważniejsze wyniki współpracy z komisją MPEG.

W **rozdziale 8** przedstawiono wnioski z przeprowadzonych badań oraz podsumowano główne osiągnięcia badawcze pracy.

## 2. Zasadnicze wyniki rozprawy

### 2.1. Stan obecny

Celem rozprawy było opracowanie nowych metod predykcji między-widokowej pozwalających na zredukowanie prędkości bitowej wymaganej do transmisji trójwymiarowych sekwencji wizyjnych w formacie MVD (ang. *Multiview Video plus Depth*), poprzez zapewnienie bardziej efektywnej reprezentacji informacji o ruchu oraz wykorzystanie dostępnej informacji o głębi.

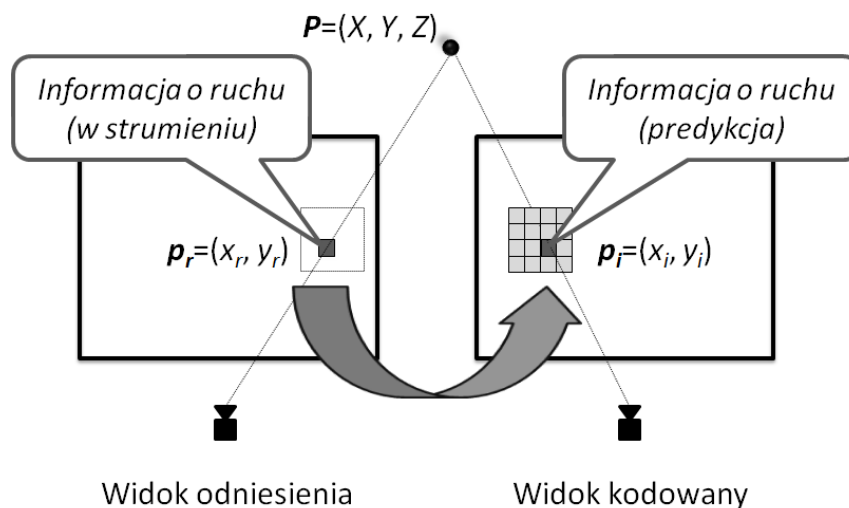
W opracowanych dotychczas technikach predykcji między-widokowej stosowano względnie proste metody określania zależności pomiędzy różnymi widokami sekwencji wielowidokowej [Guo06, Koo06, Chen07, Koo07, Lin07, Song07, Yang09]. Metody te oparte są na globalnym wektorze rozbieżności lub rozbieżności estymowanej dla niewielkich, prostokątnych bloków obrazu, jak ma to miejsce np. w narzędziu kodowania *Motion Skip* [Koo06, Yang09]. W wielu przypadkach, takie podejście może prowadzić do znacznych błędów, zwłaszcza jeśli jednakowa wartość rozbieżności stosowana jest do reprezentacji rozbieżności dla obszarów sceny wizyjnej o różniącej się znacznie odległości od kamery. Podobnie, zastosowanie lokalnej wartości rozbieżności przypisanej do bloków obrazu może powodować niedokładności. Wynika to z faktu, iż podział obrazu na prostokątne bloki nie gwarantuje spójności takiego podziału z krawędziami obiektów występujących w scenie wizyjnej.

Kolejnym ograniczeniem znanych dotąd metod jest fakt, iż dokonują one predykcji informacji o ruchu dla prostokątnych bloków obrazu kodowanego na podstawie korespondujących z nimi prostokątnych bloków obrazu w widoku odniesienia. Jednakże, mając na uwadze treść obrazu kodowanego i obrazu odniesienia, zawartość obu bloków może się w pewnym stopniu różnić. Ponadto, prostokątny blok punktów obrazu w widoku kodowanym nie odpowiada zazwyczaj prostokątnemu blokowi w widoku odniesienia i na odwrót. W rezultacie, przewidywane w ten sposób pole ruchu jest mniej dokładne, co przekłada się na efektywność kompresji wielowidokowej sekwencji wizyjnej z użyciem takiego predyktora.

## 2.2. Oryginalne metody predykcji między-widokowej wykorzystujące informację o głębi

Najważniejszym osiągnięciem zaprezentowanym w rozprawie jest opracowanie dwóch autorskich metod predykcji między-widokowej z wykorzystaniem informacji o głębi: **Forward Projection Depth-Based Prediction (FPDBP)** oraz **Inverse Projection Depth-Based Prediction (IPDBP)**, które mogą być zastosowane w kompresji trójwymiarowych sekwencji wizyjnych i usuwają niedostatki istniejących metod.

W autorskich metodach predykcji między-widokowej zaproponowanych w niniejszej rozprawie wykorzystywana jest informacja o głębi, umożliwiająca modelowanie sceny w przestrzeni trójwymiarowej. Takie podejście pozwala na dokładne odwzorowanie położenia każdego punktu obrazu pomiędzy widokiem kodowanym i widokiem odniesienia, a w efekcie, lepszą predykcję informacji o ruchu z widoku odniesienia.



Rys. 2.1. Między-widokowa predykcja informacji o ruchu.

Metoda predykcji między-widokowej FPDBP oparta jest na rzutowaniu w przód (ang. *forward mapping*), znanym z grafiki komputerowej i wykorzystuje projekcję współrzędnych punktów obrazu w przestrzeni trójwymiarowej na podstawie informacji o głębi z widoku kodowanego w celu odwzorowania położenia punktów obrazu pomiędzy widokiem kodowanym i widokiem odniesienia. Metoda składa się z następujących kroków:

- *Między-widokowa projekcja współrzędnych punktów w przestrzeni trójwymiarowej:* współrzędne punktu w widoku kodowanym  $p_i = (x_i, y_i)$  rzutowane są do przestrzeni

trójwymiarowej - punkt  $P = (X, Y, Z)$  z wykorzystaniem informacji o głębi punktu  $p_i$ , a następnie z powrotem do widoku odniesienia – punkt  $p_r = (x_r, y_r)$ . W rezultacie znalezione zostaje odwzorowanie pomiędzy punktem w widoku kodowanym  $p_i$  i punktem w widoku odniesienia  $p_r$  (rys. 2.1).

- *Wykrywanie przysłoneń*: polega na sprawdzeniu, czy punkt  $p_i$  jest widoczny w widoku odniesienia na podstawie wybranego *algorytmu wykrywania przysłoneń*. W przypadku, gdy wykryto przysłonecie, punkt  $p_r$ , odpowiadający punktowi  $p_i$ , wyszukiwany jest w kolejnym dostępnym widoku odniesienia.
- *Uzupełnianie obszarów nieprzypisanych*: dla każdego punktu  $p_i$  w widoku kodowanym, dla którego nie udało się znaleźć odwzorowania w żadnym z dostępnych widoków odniesienia, wykonywana jest procedura dziedziczenia odwzorowania na widok odniesienia na podstawie informacji z sąsiadujących z nim punktów obrazu. Dziedziczenie to odbywa się zgodnie z wybranym *algorytmem uzupełniania obszarów nieprzypisanych*. W sytuacji, gdy uzupełnianie obszarów nieprzypisanych nie jest możliwe lub odwzorowany punkt  $p_r$  w widoku odniesienia nie posiada żadnej informacji o ruchu, którą można by wykorzystać do predykcji, jako źródło informacji o ruchu dla punktu  $p_i$  wykorzystany zostaje dostępny w kodeku predyktor wewnątrz-widokowy (np. dla kodeka AVC jest to predyktor medianowy).

W wyniku wykonania powyższych kroków, dla każdego punktu kodowanego obrazu określone zostaje źródło informacji o ruchu, która może zostać wykorzystana do predykcji z kompensacją ruchu. Istotny element zaproponowanego podejścia stanowi fakt, iż w procesie kodowania każdy punkt kodowanego bloku dziedziczy informację o ruchu z przypisanego mu źródła w sposób niezależny. Oznacza to, że predykcja z kompensacją ruchu wykonywana podczas kodowania może być inna dla każdego punktu takiego bloku, niezależnie od rozmiaru i podziału bloku na partycje.

Należy również podkreślić, że jedyną informacją dziedziczną z widoku odniesienia jest informacja oryginalnie zakodowana w strumieniu bitowym tego widoku i żadne dodatkowe obliczenia lub modyfikacje procesu kodowania w widoku odniesienia nie są wymagane. W rezultacie, dodatkowe operacje wpływające na złożoność obliczeniową procesu kodowania i dekodowania związane są z projekcją w przestrzeni trójwymiarowej, wykrywaniem przysłoneń oraz uzupełnianiem obszarów nieprzypisanych.

Druga z autorskich metod predykcji między-widokowej – IPDBP, oparta jest na znanym z grafiki komputerowej rzutowaniu wstecz (ang. *inverse mapping*) i, podobnie jak omówiona wcześniej metoda FPDBP, służy do określenia odwzorowania położenia punktów obrazu pomiędzy widokiem kodowanym oraz widokiem odniesienia. Jednak w przeciwieństwie do metody FPDBP,

projekcja współrzędnych punktów obrazu w przestrzeni trójwymiarowej wykonywana jest wyłącznie na podstawie informacji o głębi z widoku odniesienia. Jest to niezwykle istotna cecha, umożliwiającą wykorzystanie autorskiej metody predykcji w sytuacji, gdy koder lub dekodek z różnych względów nie mogą skorzystać z informacji o głębi widoku kodowanego.

Metoda predykcji IPDBP składa się z kroków omówionych wcześniej dla metody FPDBP. Najważniejszą różnicą względem metody FPDBP jest jednak odwrócony kierunek projekcji: w metodzie IPDBP to współrzędne punktu w widoku odniesienia  $p_r$  rzutowane są do przestrzeni trójwymiarowej na podstawie dostępnej informacji o głębi punktu  $p_r$ , a następnie z powrotem do widoku kodowanego, w skutek czego określone zostaje odwzorowanie punktu  $p_r$  na punkt obrazu  $p_i$  w widoku kodowanym. Kolejną różnicą jest brak konieczności przeprowadzania procedury wykrywania przysłonieć w odrębnym kroku. W przypadku metody IPDBP informacja o tym, które z punktów obrazu kodowanego nie są widoczne w widoku odniesienia uzyskiwana jest podczas projekcji między-widokowej – punkty w widoku kodowanym, dla których nie odnaleziono odwzorowania są automatycznie klasyfikowane jako przysłonięte. Skutkuje to mniejszą złożonością obliczeniową, lecz procedura ta wymaga dodatkowego bufora pamięci przechowującego współrzędne wszystkich punktów odwzorowanych. Dla porównania, w przypadku niektórych algorytmów wykrywania obszarów przysłoniętych stosowanych w metodzie predykcji FPDBP bufor taki nie jest potrzebny.

Autor przeanalizował szereg wariantów zaproponowanych metod predykcji. W rezultacie, w pracy przedstawiono również oryginalny algorytm wykrywania przysłonieć *o-mask*, adaptację algorytmu *z-test* jako szybkiego algorytmu wykrywania przysłonieć o niskiej złożoności oraz trzy autorskie metody uzupełniania obszarów nieprzypisanych o niskiej złożoności: *FILLmax*, *FILLmin* oraz *FILLsim*.

## **2.3. Autorskie techniki kodowania informacji o ruchu wykorzystujące zaproponowane metody predykcji między-widokowej**

### **2.3.1. Sposób wykorzystania autorskich metod predykcji między-widokowej**

W rozprawie zaproponowano **autorskie techniki kodowania informacji o ruchu pozwalają na efektywne wykorzystanie nowych metod predykcji między-widokowej FPDBP oraz IPDBP**. Specjalnie opracowane techniki umożliwiają zastosowanie oryginalnych, stworzonych przez autora metod predykcji w istniejących i przyszłych wielowidokowych kodekach wizyjnych w celu zredukowania prędkości bitowej wymaganej do transmisji trójwymiarowych sekwencji wizyjnych w formacie MVD. **Zaprezentowano różne warianty kodeków wielowidokowych opracowanych przez autora, oparte o najnowsze i najbardziej zaawansowane rozwiązania będące stanem obecnej wiedzy: kodeki wizyjne AVC oraz HEVC.**

Zadaniem autorskich metod predykcji między-widokowej FPDBP oraz IPDBP jest znalezienie odwzorowania pomiędzy współrzędnymi każdego punktu obrazu kodowanego oraz odpowiadającymi im punktami obrazu odniesienia. Dzięki temu, dla każdego widoku zależnego sekwencji wielowidokowej, możliwa jest predykcja informacji o ruchu z dostępnych, zakodowanych już widoków odniesienia. Autorskie techniki kodowania informacji o ruchu zaproponowane w rozprawie wykorzystują przewidywaną w ten sposób informację o ruchu do zareprezentowania informacji o ruchu kodowanego bloku obrazu. W przypadku, gdy podczas kodowania, w procesie optymalizacji uwzględniającym prędkość bitową i zniekształcenie, koder wybierze informację o ruchu przewidzianą między-widokowo według jednej z zaproponowanych metod predykcji, wektory ruchu oraz indeksy obrazów odniesienia, wykorzystywane w procesie predykcji z kompensacją ruchu kodowanego bloku obrazu, zostają bezpośrednio odziedziczone z widoku odniesienia, niezależnie dla każdego punktu kodowanego bloku obrazu (rys. 2.1). W ten sposób informacja o ruchu z widoku odniesienia zostaje ponownie wykorzystana, zapobiegając jednocześnie konieczności jej retransmisji w strumieniu widoku kodowanego. Takie podejście skutkuje zwiększeniem efektywności kompresji kodeka wielowidokowego w przypadku, gdy korelacja pomiędzy polami ruchu widoku kodowanego i widoku odniesienia jest duża.

Wybór informacji o ruchu przewidzianej między-widokowo według jednej z zaproponowanych w rozprawie metod predykcji dla kodowanego bloku obrazu musi zostać zasygnalizowana dekodderowi, by umożliwić poprawne zdekodowanie takiego bloku. Wybór właściwej strategii sygnalizacji jest zagadnieniem krytycznym z punktu widzenia maksymalizacji zysków kodowania. W niniejszej rozprawie **przedstawiono i przeanalizowano opracowane przez autora różne strategie sygnalizacji użycia między-widokowych technik kodowania informacji o ruchu, a także omówiono szczegóły związanych z ich implementacją niezbędnych modyfikacji składni najnowszych i najbardziej zaawansowanych wielowidokowych kodeków wizyjnych będących stanem obecnej wiedzy: kodeków MVC i MV-HEVC.**



## 2.3.2. Autorskie techniki kodowania informacji o ruchu w kodeku MVC

Na podstawie przeprowadzonych badań stwierdzono, iż zaproponowane algorytmy predykcji między-widokowej z wykorzystaniem informacji o głębi umożliwiają bardzo dokładną predykcję informacji o ruchu. Dlatego też techniki kodowania informacji o ruchu opracowane przez autora zostały zaadoptowane w kodeku MVC jako rozszerzenie standardowych trybów *Skip* oraz *Direct*, stanowiących najbardziej efektywny sposób kodowania informacji o ruchu w tymże kodeku. W rezultacie, koszt wyboru zaproponowanych metod predykcji informacji o ruchu wykorzystujących informację o głębi jest relatywnie niski, co umożliwia wykorzystanie potencjału zaproponowanych metod oraz maksymalizuje zyski kodowania wynikające z wprowadzonych modyfikacji kodeka wielowidokowego. W szczególności, **w rozprawie zaproponowano szereg wariantów wykorzystania zaproponowanych metod predykcji między-widokowej, tzw. trybów kodowania IV: Inter-View Direct (IVD) oraz Inter-View Skip (IVS)**. Wraz ze standardowymi trybami kodowania makrobloków, w tak zmodyfikowanym kodeku MVC dostępne są bowiem cztery efektywne sposoby reprezentacji informacji o ruchu – tryby kodowania: *Skip*, *Direct*, *IVS* oraz *IVD*. W takiej sytuacji istnieje pewna liczba możliwych strategii efektywnego kodowania informacji o ruchu, różniących się możliwością użycia i kosztem sygnalizacji wspomnianych trybów kodowania. **W rozprawie zaproponowano i przebadano osiem najbardziej obiecujących wariantów**, przy czym implementacja każdego z nich wiązała się z zaprojektowaniem niezbędnych modyfikacji składni kodeka wielowidokowego. Wnioski z przeprowadzonych badań pozwoliły udzielić odpowiedzi na pytanie, czy możliwość zastosowania wszystkich czterech trybów kodowania jest niezbędna do osiągnięcia jak największego stopnia kompresji oraz który z zaproponowanych wariantów składni pozwala na maksymalizację zysków kodowania osiąganych przez kodek wielowidokowy.

Istotnym problemem związanym z praktycznym zastosowaniem autorskich metod predykcji w kodeku MVC rozwiązany w niniejszej rozprawie jest kwestia sposobu sygnalizacji w strumieniu współczynników opisujących zakres głębi sceny wizyjnej [ISO10], niezbędnych w procesie projekcji między-widokowej z wykorzystaniem informacji o głębi. W normie AVC opisującej kodek wielowidokowy MVC zdefiniowano sposób sygnalizacji zewnętrznych i wewnętrznych parametrów systemu akwizycji z użyciem jednego z typów tzw. komunikatów SEI (ang. *Supplemental Enhancement Information*) o nazwie Multiview Acquisition Info SEI [Vet07]. Norma nie określa natomiast sposobu sygnalizacji współczynników opisujących zakres głębi sceny

wizyjnej. W rozprawie zaproponowano autorskie rozwiązanie polegające na rozszerzeniu istniejącego typu komunikatu SEI o nazwie Multiview Scene Info [Yea07] o dodatkowe pola opisujące brakujące współczynniki. Dodatkowym walorem takiego rozwiązania jest spójność z dotychczasowym podejściem opisanym w normie AVC.

**Wszelkich niezbędnych modyfikacji wymaganych do zaimplementowania autorskich metod kodowania informacji o ruchu dokonano w oprogramowaniu modelowym MVC, opracowanego przez komisję MPEG w ramach prac nad aneksem H normy kodowania sekwencji wizyjnych AVC (ISO/IEC MPEG-4 część 10, ITU-T H.264) [ISO11, Chen09].**

### **2.3.3. Autorskie techniki kodowania informacji o ruchu w kodeku MV-HEVC**

Ze względu na istotne różnice strukturalne pomiędzy kodekami MV- HEVC oraz MVC, opracowanie technik kodowania informacji o ruchu wykorzystujących autorskie metody predykcji FPDBP i IPDBP wymagało zastosowania odmiennego podejścia. W szczególności, w kodekach MV-HEVC i MVC zastosowano wyraźnie odmiennie mechanizmy predykcji z kompensacją ruchu. Powoduje to m. in., iż tryby kodowania *Direct* oraz *Skip* znane z kodeka MVC nie występują w kodeku MV-HEVC, a więc autorskie techniki kodowania informacji o ruchu opracowane dla kodeka MVC nie mogą być tu zastosowane. Mając na uwadze powyższe spostrzeżenia, najbardziej naturalnym sposobem rozszerzenia kodeka MV-HEVC o zaproponowane w rozprawie mechanizmy między-widokowej predykcji informacji o ruchu było zmodyfikowanie adaptacyjnej listy predyktorów informacji o ruchu pozwalającej w bardzo efektywny sposób zakodować cały zestaw informacji o ruchu opisującej daną jednostkę kodowania CU (ang. *Coding Unit*). Wspomniana lista predyktorów w kodeku MV-HEVC stanowi uporządkowany zbiór predyktorów uszeregowanych w kolejności zależnej od prawdopodobieństwa ich użycia. W zaproponowanym przez autora rozwiązaniu, oryginalna lista predyktorów stosowanych w kodeku MV-HEVC do predykcji informacji o ruchu dla trybu kodowania bloku o nazwie *merge*, została rozszerzona o metody predykcji między-widokowej opisane w niniejszej rozprawie. Dzięki takiemu podejściu pierwotna struktura kodeka MV-HEVC pozostaje niezmienną, a liczba i koszt dostępnych trybów kodowania CU są zachowane. Dodatkowo, w pełni wykorzystana zostaje hierarchiczna struktura kodeka, a autorskie predyktory informacji o ruchu mogą bez trudu zostać użyte do kodowania CU o wszystkich dostępnych rozmiarach bez konieczności dodatkowych modyfikacji składni strumienia. Kolejną zaletą zaproponowanego rozwiązania jest możliwość elastycznej zmiany pozycji na liście predyktorów, na której umieszczany jest autorski predyktor informacji o ruchu, a w efekcie

optymalizacja kodeka wielowidokowego ze względu na maksymalizację osiąganych zysków kompresji. **W niniejszej rozprawie zaproponowano i przebadano trzy najbardziej obiecujące warianty uszeregowania listy predyktorów kodeka wielowidokowego MV-HEVC rozszerzonej o autorski między-widokowy predyktor informacji o ruchu.** W *wariancie 1* predyktor między-widokowy umieszczony został na początku listy (najmniejszy koszt sygnalizacji). Analogicznie, w *wariancie 2* predyktor znajduje się na drugiej pozycji na liście, natomiast w *wariancie 3* – na pozycji 3.

**Implementacje powyższych technik kodowania informacji o ruchu w kodeku MV-HEVC wykonano w oprogramowaniu opracowanym oryginalnie przez Politechnikę Poznańską [Dom11], bazującym na oprogramowaniu modelu testowego o nazwie HM 3.0 [MP11a] stworzonym w ramach prac komisji MPEG nad techniką kodowania nowej generacji HEVC [MP11, Dom12].**

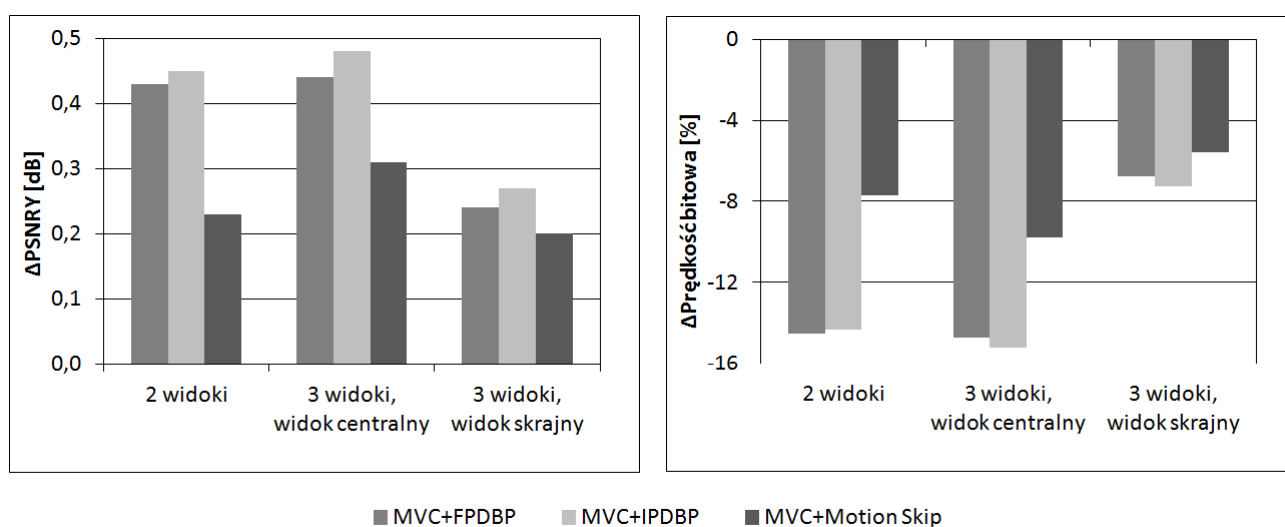
## **2.4. Wpływ zaproponowanych metod predykcji na efektywność kompresji kodeków wielowidokowych**

### **2.4.1. Wpływ na efektywność kompresji kodeka MVC**

Zaprezentowane w rozprawie badania przeprowadzone z użyciem ograniczonego zbioru testowego, pozwoliły na wybór wariantu zmodyfikowanego kodeka MVC, wykorzystującego autorskie metody predykcji między-widokowej FPDBP oraz IPDBP, o największej efektywności kompresji. Badania przeprowadzono dla ośmiu wariantów kodeka, stanowiących implementację nowych technik kodowania informacji o ruchu zaproponowanych przez autora. **Wyniki przedstawione w rozprawie pokazały, iż zastosowanie autorskich metod predykcji, niezależnie od zastosowanego wariantu kodeka, skutkuje poprawą efektywności kompresji kodeka wielowidokowego.**

Wariant kodeka osiągający największe zyski kodowania został następnie poddany szczegółowym badaniom w celu określenia wpływu zaproponowanych metod predykcji między-widokowej FPDBP oraz IPDBP na efektywność kompresji kodeka MVC. Rys. 2.2 przedstawia średnie zyski kodowania uzyskane względem oryginalnego kodeka MVC, obliczone z użyciem miar Bjontegaarda, zgodnie z zaleceniami komisji MPEG. Wyniki zaprezentowane w rozprawie pokazują, iż zastosowanie predyktora FPCBP umożliwia osiągnięcie średniej redukcji prędkości

bitowej wynoszącej 14,5 [%] w przypadku kodowania widoku zależnego sekwencji dwu-widokowej oraz 14,3[%] i 7,7 [%] w przypadku kodowania dwóch widoków zależnych sekwencji trzy-widokowej. W przypadku predyktora IPDBP wartości te wynoszą odpowiednio: 14,7 [%] dla kodowania widoku zależnego sekwencji dwu-widokowej oraz 15,2 [%] i 9,8 [%] dla kodowania sekwencji trzy-widokowej. **Autor wykazał zatem, iż możliwe jest zwiększenie efektywności reprezentacji informacji o ruchu w kodowaniu trójwymiarowych sekwencji wizyjnych w formacie MVD, poprzez wykorzystanie nowych metod między-widokowej predykcji informacji o ruchu w kodeku MVC.**

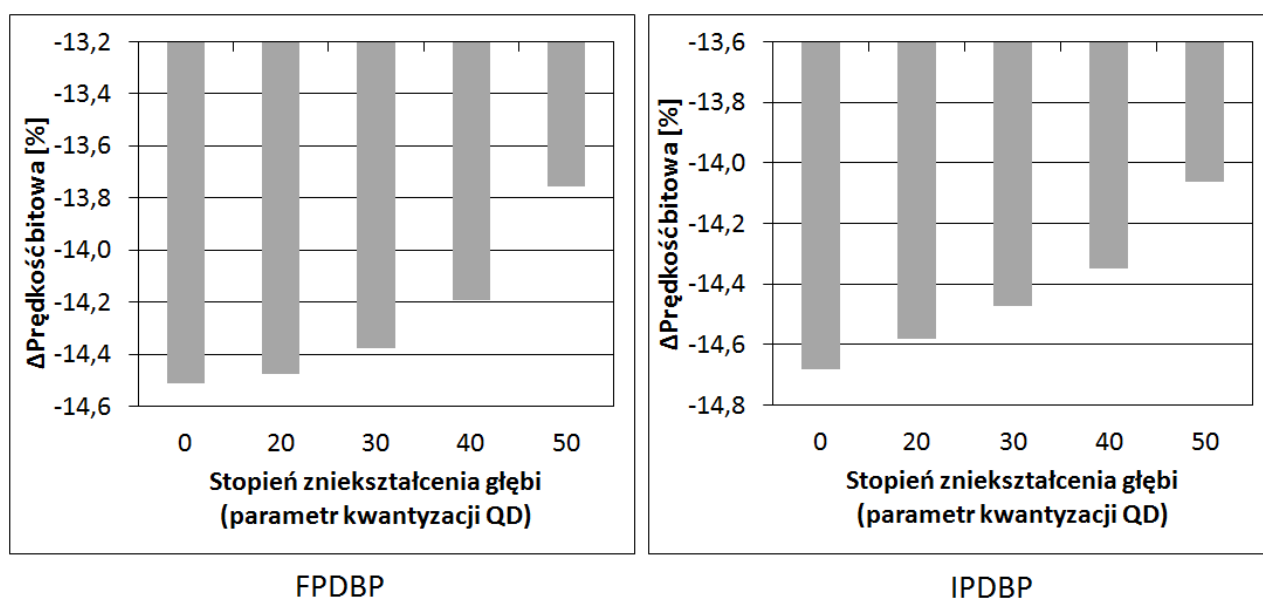


Rys. 2.2. Zyski kodowania (miary Bjontegaarda) osiągnięte poprzez zastosowanie autorskich metod predykcji FPDBP i IPDBP względem oryginalnego kodeka MVC. Dla porównania: zyski osiągnięte przez konkurencyjną metodę *Motion Skip*.

Zaproponowane metody zostały również eksperymentalnie porównane z konkurencyjną techniką między-widokowej predykcji informacji o ruchu, znanej jako *Motion Skip*. Technika ta została zaproponowana podczas prac komisji MPEG nad normą MVC. Przeprowadzone badania wykazały jednak, iż efektywność kompresji osiągnięta przez kodek MVC wykorzystujący autorskie algorytmy FPDBP oraz IPDBP przewyższa efektywność kompresji algorytmu *Motion Skip* (rys. 2.2). Jak pokazano w rozprawie, średnia redukcja prędkości bitowej osiągnięta przez metody zaproponowane przez autora okazała się być o 2,1-7,9 [pp] większa niż ta osiągnięta przez algorytm *Motion Skip*. **Wykazano zatem, iż możliwe jest opracowanie technik reprezentacji informacji o ruchu, konkurencyjnych względem metod opisanych w literaturze i rozwijanych równoległe do prac autora.**

Przeprowadzone badania pozwoliły również zaobserwować, iż **wykorzystanie między-widokowych metod predykcji opisanych w niniejszej rozprawie powoduje częstsze wybieranie przez koder trybów kodowania makrobloku nie wymagających przesyłania informacji o ruchu**, a więc trybów *Direct*, *Skip*, *IVD* oraz *IVS*. Wzrost częstości wyboru wspomnianych trybów wynosi średnio 8,4-10,6 [pp]. To z kolei skutkuje zwiększeniem efektywności kompresji kodeka wielowidokowego.

W rozprawie dokonano również oceny eksperymentalnej wpływu jakości głębi na efektywność kompresji z użyciem między-widokowych metod predykcji zaproponowanych przez autora. Wyniki badań wskazują, iż zastosowanie w predykcji między-widokowej głębi poddanej kompresji stratnej skutkuje zmniejszeniem zysków kodowania osiągniętych przez autorskie metody predykcji, jednak zmiana jest niewielka (rys. 2.3).

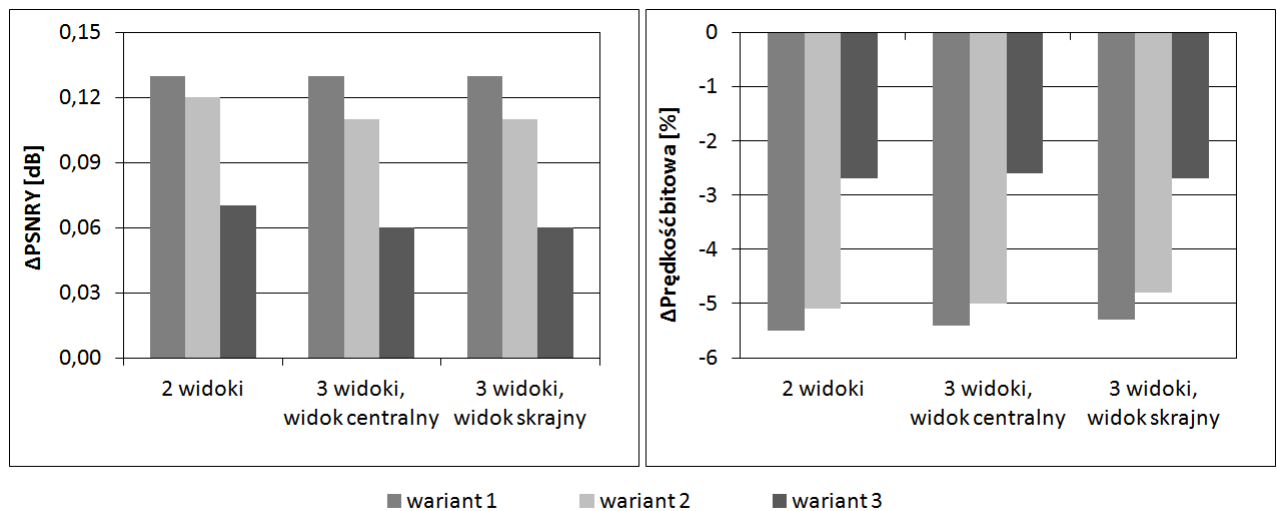


Rys. 2.3. Przykładowe zyski kodowania (miary Bjontegaarda) osiągnięte poprzez zastosowanie autorskich metod predykcji FPDBP oraz IPDBP względem oryginalnego kodeka MVC w przypadku użycia głębi o różnej jakości, określonej parametrem kwantyzacji dla głębi QD.

## 2.4.2. Wpływ na efektywność kompresji kodeka MV-HEVC

Wpływ autorskich metod predykcji między-widokowej na efektywność kompresji kodeka MV-HEVC przebadano analizując trzy warianty kodeka, będące implementacją trzech zaproponowanych w rozprawie technik kodowania informacji o ruchu dla kodeka MV-HEVC. Każdy z wariantów odpowiada innemu sposobowi uszeregowania listy predyktorów kodeka MV-HEVC rozszerzonej o autorski predyktor między-widokowy. Na rys. 2.4 przedstawiono średnie

zyski kodowania dla poszczególnych wariantów zmodyfikowanego kodeka MV-HEVC względem kodeka oryginalnego, obliczone z użyciem miar Bjontegaarda, zgodnie z zaleceniami komisji MPEG. Uzyskane wyniki eksperymentalne pokazują, iż, pomimo zaawansowanych mechanizmów predykcji używanych przez kodek MV-HEVC, zastosowanie autorskich metod predykcji międzywidokowej umożliwia osiągnięcie średniej redukcji prędkości bitowej wynoszącej 5,3-5,5 [%] w przypadku kodowania widoku zależnego sekwencji dwu-widokowej oraz 4,8-5,1[%] i 2,6-2,7 [%] w przypadku kodowania dwóch widoków zależnych sekwencji trzy-widokowej. **Autor wykazał zatem, iż możliwe jest zwiększenie efektywności reprezentacji informacji o ruchu w kodowaniu trójwymiarowych sekwencji wizyjnych w formacie MVD, poprzez wykorzystanie nowych metod między-widokowej predykcji informacji o ruchu w kodeku MV-HEVC.** Analiza różnych wariantów kodeka pozwoliła także stwierdzić, iż wielkość odnotowanych zysków kodowania zależy od tego, która z autorskich technik kodowania informacji o ruchu jest użyta.



Rys. 2.4. Zyski kodowania (miary Bjontegaarda) osiągnięte poprzez zastosowanie autorskich metod predykcji między-widokowej względem oryginalnego kodeka MV-HEVC.

Badania eksperymentalne przeprowadzone dla kodeka MV-HEVC wskazują również dwa źródła zysków kodowania osiąganych poprzez zastosowanie autorskich algorytmów predykcji między-widokowej. Dostępność nowego, wydajnego predyktora informacji o ruchu **powoduje wzrost ogólnej liczby jednostek kodowania CU, dla których nie jest konieczne przesyłanie informacji o ruchu.** Dodatkowo, własność autorskich algorytmów predykcji polegająca na niezależnym dziedziczeniu informacji o ruchu dla każdego punktu kodowanej CU **zapobiega w wielu przypadkach konieczności podziału tej jednostki kodowania na mniejsze części,** redukując w ten sposób liczbę przesyłanych bitów, niezbędnych do zasygnalizowania takiego podziału.

## 3. Podsumowanie pracy oraz wnioski

### 3.1. Podsumowanie pracy

Zgodnie z postawionymi celami, autor pracy skupił swoje starania na opracowaniu nowych, efektywnych metod reprezentacji informacji o ruchu dla potrzeb kodowania trójwymiarowych sekwencji wizyjnych. Celem było stworzenie algorytmów zapewniających lepszą efektywność kompresji w porównaniu z istniejącymi technikami reprezentacji informacji o ruchu.

W tym celu sformułowany został problem reprezentacji informacji o ruchu w naturalnych trójwymiarowych sekwencjach wizyjnych zawierających dodatkowo informację o głębi. Problem ten przeanalizowany został zarówno pod kątem predykcji, jak i kodowania informacji o ruchu w trójwymiarowych kodekach wizyjnych.

Opracowanie efektywnych metod reprezentacji informacji o ruchu dla potrzeb kodowania trójwymiarowych sekwencji wizyjnych wymagało przeprowadzenia następujących badań eksperymentalnych:

- Przeanalizowano istniejące techniki kodowania wielowidokowych sekwencji wizyjnych, ze szczególnym akcentem na najbardziej zaawansowane algorytmy predykcji informacji o ruchu znane z literatury (rozdział 2).
- Wskazano słabości istniejących technik predykcji między-widokowej (rozdział 3).
- Przeprowadzono szereg eksperymentów wstępnych, które pozwoliły na opracowanie autorskich metod predykcji między-widokowej z wykorzystaniem informacji o głębi: Forward Projection Depth-Based Prediction (FPDBP) oraz Inverse Projection Depth-Based Prediction (IPDBP) (rozdział 3 **Błąd! Nie można odnaleźć źródła odwołania.**).
- Przeanalizowano wydajność zaproponowanych algorytmów FPDBP oraz IPDBP w przypadku wykorzystania ich do predykcji informacji o ruchu, a także porównano je z predyktorem medianowym, stosowanym powszechnie w dzisiejszych hybrydowych kodekach wizyjnych (rozdział 4).

Na podstawie przeprowadzonych badań stwierdzono, iż zaproponowane algorytmy predykcji między-widokowej z wykorzystaniem informacji o głębi umożliwiają bardzo dokładną predykcję informacji o ruchu. W szczególności, rozważane autorskie metody predykcji w wielu przypadkach osiągnęły większą wydajność niż predyktor medianowy znany z literatury. W rezultacie, autor

postanowił wykorzystać potencjał zaproponowanych algorytmów do poprawy efektywności kompresji wielowidokowych kodeków wizyjnych stanowiących obecny stan wiedzy. W tym celu podjęto następujące działania:

- Opracowano metody efektywnego wykorzystania zaproponowanych algorytmów predykcji informacji o ruchu dla wielowidokowych kodeków wizyjnych MVC oraz MV-HEVC (rozdział 5).
- Na podstawie powyższych metod, algorytmy FPDBP oraz IPDBP zostały zaimplementowane w oprogramowaniu modelowym kodeków MVC i MV-HEVC, a także przebadane eksperymentalnie w celu sprawdzenia ich użyteczności dla potrzeb kodowania wielowidokowych sekwencji wizyjnych (rozdział 6).

W fazie projektowania zwrócono szczególną uwagę na ograniczenie złożoności struktury wynikowego kodeka oraz starano się zminimalizować liczbę modyfikacji składni niezbędnych do zaimplementowania autorskich algorytmów predykcji.

Przeprowadzone eksperymenty pozwoliły stwierdzić, iż algorytmy FPDBP i IPDBP opracowane przez autora umożliwiają poprawę efektywności kompresji wielowidokowych kodeków wizyjnych stanowiących obecny stan wiedzy. Większe użycie trybów kodowania bloków o najniższym koszcie wpłynęło na wzrost efektywności kompresji w przypadku kodowania widoków pobocznych wielowidokowych sekwencji wizyjnych. W przypadku kodeka MVC obiektywne miary zniekształcenia wykazały średnią redukcję prędkości bitowej o 7,7-15,2 [%] dla pojedynczego widoku pobocznego. Podobnie, średnia redukcja prędkości bitowej zmierzona dla kodeka MV-HEVC wynosi 2,6-5,5 [%]. Powyższe zyski kodowania obliczono z użyciem miar Bjontegaarda [Bjo01], stosowanych powszechnie w ramach prac komisji Moving Picture Experts Group (MPEG) Międzynarodowej Organizacji Normalizacyjnej (ang. *International Organization for Standardization*, ISO).

Zaproponowane metody zostały eksperymentalnie porównane z inną techniką międzywidokowej predykcji informacji o ruchu, znanej jako Motion Skip. Technika ta została zaproponowana podczas prac komisji MPEG nad normą MVC. Przeprowadzone badania wykazały jednak, iż efektywność kompresji osiągnięta przez kodek MVC wykorzystujący autorskie algorytmy FPDBP oraz IPDBP przewyższa efektywność kompresji algorytmu Motion Skip. Średnia redukcja prędkości bitowej osiągnięta przez metody zaproponowane przez autora okazała się być o 2,1-7,9 [pp] większa niż ta osiągnięta przez algorytm Motion Skip.

W celu zbadania wydajności zaproponowanych algorytmów predykcji przeprowadzono eksperymenty o łącznym czasie trwania przekraczającym 50 000 godzin. Obliczenia



przeprowadzono z użyciem jednego z najszybszych serwerów obliczeniowych, jakie były komercyjnie dostępne w tym czasie.

Algorytmy między-widokowej predykcji z wykorzystaniem informacji o głębi zaproponowane przez autora zostały również włączone jako element propozycji technologii kodowania trójwymiarowych sekwencji wizyjnych zgłoszonej przez Politechnikę Poznańską, opracowanej w odpowiedzi na wezwanie komisji MPEG [MP11b] (rozdział 7). Zaproponowany trójwymiarowy kodek wizyjny osiągnął wybitne rezultaty i, wraz z propozycją Instytutu Fraunhofera – Instytutu H. Hertza z Berlina, został oceniony jako najlepszy w klasie kodeków kompatybilnych z techniką kodowania HEVC. W konsekwencji, metoda predykcji informacji o ruchu zaproponowana przez autora została włączona do modelu odniesienia dla nowej technologii kodowania (ang. *Test Model under Consideration*) [ISO11a], który stanowi punkt startowy dla procesu tworzenia nowej normy kodowania trójwymiarowych sekwencji wizyjnych kompatybilnych z techniką kodowania HEVC, opracowywanej w ramach prac komisji MPEG. Oznacza to, iż rozwiązania zaproponowane w tej pracy mają dużą szansę by stać się częścią powstającej normy kodowania trójwymiarowych sekwencji wizyjnych.

Należy również zaznaczyć, iż zaproponowane autorskie techniki FPDBP oraz IPDBP są pierwszymi algorytmami predykcji między-widokowej wykorzystującymi informację o głębi, umożliwiającymi predykcję informacji o ruchu z dokładnością do punktu obrazu. Są to również jedne z pierwszych skutecznych algorytmów między-widokowej predykcji informacji o ruchu zaproponowanych dla trójwymiarowego kodeka wizyjnego, kompatybilnego z techniką kodowania HEVC.

Celem rozprawy było opracowanie nowych technik kompresji trójwymiarowych sekwencji wizyjnych dla sekwencji wielowidokowych w formacie MVD. Autor zaproponował dwie metody między-widokowej predykcji trójwymiarowych sekwencji wizyjnych: FPDBP oraz IPDBP. Pokazano, iż zaproponowane predyktory pozwalają na zredukowanie prędkości bitowej wymaganej do transmisji wspomnianych sekwencji wizyjnych w porównaniu z systemami uznawanymi za reprezentatywne dla obecnego stanu wiedzy poprzez zapewnienie bardziej efektywnej reprezentacji informacji o ruchu oraz wykorzystanie dostępnej informacji o głębi. Zaproponowane techniki zostały zbadane eksperymentalnie w celu oceny ich wpływu na efektywność kompresji istniejących oraz przyszłych wielowidokowych kodeków wizyjnych.

Przedstawione w pracy wyniki, będące rezultatami przeprowadzonych badań, pozwalają stwierdzić, iż tezy postawione w rozprawie zostały udowodnione. Dowiedziono również, że zaproponowane algorytmy między-widokowej predykcji z wykorzystaniem informacji o głębi

poprawiają efektywność reprezentacji informacji o ruchu w kodowaniu trójwymiarowych sekwencji wizyjnych w formacie MVD. Pokazano także, iż opracowane techniki reprezentacji informacji o ruchu są konkurencyjne względem metod opisanych w literaturze i rozwijanych w trakcie prac autora.

## **3.2. Oryginalne osiągnięcia zaprezentowane w rozprawie**

### **Główne osiągnięcia pracy:**

- Opracowanie oryginalnych algorytmów między-widokowej predykcji informacji o ruchu z wykorzystaniem informacji o głębi dla kodowania trójwymiarowych sekwencji wizyjnych: Forward Projection Depth-Based Prediction (FPDBP) oraz Inverse Projection Depth-Based Prediction (IPDBP) (rozdział 3.2).
- Zaproponowanie oryginalnych algorytmów reprezentacji informacji o ruchu dla wielowidokowych kodeków wizyjnych uznawanych za reprezentatywne dla obecnego stanu wiedzy: MVC i MV-HEVC (rozdział 5).
- Ocena eksperymentalna różnych strategii sygnalizacji dla zaproponowanych technik predykcji FPDBP oraz IPDBP (rozdział 6.3.1 oraz rozdział 6.4).
- Ocena eksperymentalna zaproponowanych technik predykcji względem wielowidokowych kodeków wizyjnych MVC i MV-HEVC (rozdział 6.3.2 oraz rozdział 6.4).
- Porównanie eksperymentalne zaproponowanych technik FPDBP oraz IPDBP względem trybu Motion Skip kodeka JMVM (rozdział 6.3.2).
- Opracowanie oryginalnego algorytmu reprezentacji informacji o ruchu dla powstającej normy kodowania trójwymiarowych sekwencji wizyjnych oraz jego ocena eksperymentalna (rozdział 7).

### **Inne oryginalne osiągnięcia zaprezentowane w rozprawie:**

- Badania eksperymentalne udziału indywidualnych elementów składni w strumieniu wytwarzanym przez wielowidokowy hybrydowy kodek wizyjny MVC (rozdział 4.1).

- Porównanie dokładności zaproponowanych metod między-widokowej predykcji informacji o ruchu z wykorzystaniem informacji o głębi z predyktorem medianowym, stosowanym powszechnie w dzisiejszych hybrydowych kodekach wizyjnych (rozdział 4.3).
- Zaproponowanie algorytmu wykrywania przysłoneń *o-mask* (rozdział 3.3).
- Adaptacja algorytmu *z-test* jako szybkiego algorytmu wykrywania przysłoneń o niskiej złożoności (rozdział 3.3).
- Opracowanie metod uzupełniania obszarów nieprzypisanych: *FILLmax*, *FILLmin* oraz *FILLSim* o niskiej złożoności (rozdział 3.4).
- Ocena eksperymentalna wpływu jakości głębi na wydajność zaproponowanych algorytmów reprezentacji informacji o ruchu w wielowidokowym kodeku wizyjnym MVC (rozdział 6.3.3).
- Badania eksperymentalne nad złożonością zaproponowanych algorytmów reprezentacji informacji o ruchu w wielowidokowych kodekach wizyjnych MVC oraz MV-HEVC (rozdział 6.5).

### 3.3. Ogólne wnioski

W rozprawie przedstawiono tezę, iż wykorzystanie korelacji pomiędzy polami ruchu sąsiednich widoków wielowidokowej sekwencji wizyjnej oraz dostępności informacji o głębi opisującej scenę wizyjną powinno umożliwić poprawę efektywności reprezentacji informacji ruchu w kodowaniu wielowidokowych sekwencji wizyjnych.

W nowoczesnych koderach wizyjnych, informacja o ruchu wraz z danymi sterującymi stanowią znaczącą część strumienia. Wyniki badań eksperymentalnych przeprowadzonych przez autora wskazują, iż w przypadku wielowidokowych sekwencji wizyjnych, wspomniane elementy składni stają się dominującą częścią strumienia dla widoków pobocznych, zwłaszcza w przypadku strumieni o małej prędkości bitowej. Powyższe obserwacje pokazują wyraźnie, że rozwój bardziej efektywnych metod reprezentacji informacji o ruchu jest celowy. Udoskonalenia w tym obszarze powinny wpłynąć na zwiększenie efektywności kodowania wielowidokowych sekwencji wizyjnych.

W rezultacie, w ostatnich latach powstał szereg dedykowanych metod efektywnej reprezentacji informacji o ruchu w wielowidokowych strumieniach wizyjnych. Starania skupiono w szczególności na wykorzystaniu korelacji pomiędzy polami ruchu sąsiednich widoków w

wielowidokowych sekwencjach wizyjnych. Jednakże, w technikach opracowanych dotychczas do określenia zależności pomiędzy różnymi widokami stosowano wyłącznie globalny wektor rozbieżności lub prostą, blokową estymację rozbieżności. W wielu przypadkach, takie podejście może prowadzić do znacznych błędów, zwłaszcza jeśli jednakowa wartość rozbieżności stosowana jest do reprezentacji rozbieżności dla obszarów sceny wizyjnej o różniącej się znacznie odległości od kamery. Podobnie, zastosowanie lokalnej wartości rozbieżności przypisanej do bloków obrazu może również powodować niedokładności. Wynika to z faktu, iż podział obrazu na prostokątne bloki nie gwarantuje spójności takiego podziału z krawędziami obiektów występujących w scenie wizyjnej. Jednocześnie, dostępność dodatkowej informacji o głębi umożliwia zastosowanie zaawansowanych technik modelowania w przestrzeni trójwymiarowej znanych z grafiki komputerowej jako Depth Image Based Rendering (DIBR). W ten sposób możliwe staje się niezależne odwzorowanie położenia poszczególnych punktów tworzących scenę wizyjną w różnych widokach, a w konsekwencji wykorzystanie tej informacji do dokładniejszego reprezentowania widoku kodowanego na podstawie widoków odniesienia. Na podstawie powyższych obserwacji, w niniejszej rozprawie zaproponowano dwie techniki predykcji informacji o ruchu na podstawie informacji o głębi, umożliwiające odwzorowanie z dokładnością do punktu obrazu.

Uzyskane wyniki eksperymentalne dowiodły, że zaproponowane udoskonalenia reprezentacji informacji o ruchu pozwalają na osiągnięcie znacznych zysków kompresji w wielowidokowych kodekach wizyjnych uznanych za reprezentatywne dla obecnego stanu techniki. Zastosowanie między-widokowych predyktorów zaproponowanych przez autora pozwala na redukcję prędkości bitowej dla kodera MVC średnio o 7.7-15.2 [%] względem kodera niezmodyfikowanego. Dla porównania, konkurencyjna technika między-widokowej predykcji informacji o ruchu znana jako *Motion Skip*, zaproponowana w ramach prac komisji MPEG nad normą MVC, pozwala na uzyskanie zysków kodowania wynoszących 5.6-7.3 [%]. Znaczne zyski kodowania zaobserwowano również dla wielowidokowych kodeków wizyjnych nowej generacji, opartych o technikę kodowania HEVC. Osiągnięta średnia redukcja prędkości bitowej wynosi w tym przypadku 2.6-5.5 [%]. W opinii autora, osiągnięte rezultaty należy uznać za bardzo dobre, zwłaszcza, gdy uwzględnimy fakt, iż, śledząc rozwój kolejnych generacji kodeków wizyjnych, możemy zaobserwować tendencję polegającą na redukcji prędkości bitowej o ok. 50 [%] co 9 lat. Taki postęp wymaga zaangażowania niemałych środków, tysięcy osobo-lat pracy oraz wiąże się zazwyczaj ze znacznymi zmianami dotychczas stosowanej technologii. Na tym tle, rezultaty odnotowane dla technik predykcji opracowanych przez autora, będących zaledwie nowym narzędziem, a nie odrębną technologią kodowania, zasługują na bardzo pozytywną ocenę.

Na podstawie badań eksperymentalnych nad zaproponowanymi technikami między-widokowej predykcji informacji o ruchu wykorzystującymi informację o głębi, wyciągnięto także następujące wnioski:

- Rozważane autorskie metody predykcji w wielu przypadkach osiągnęły lepsze rezultaty niż predyktor medianowy znany z literatury.
- Zastosowanie autorskich algorytmów predykcji skutkuje większym użyciem trybów kodowania niewymagających transmisji informacji o ruchu do dekodera, a także, w wielu przypadkach, zapobiega konieczności dalszego podziału kodowanych bloków na mniejsze części.. W oczywisty sposób wpływa to na poprawę efektywności kompresji kodeka.
- Osiągnięte zyski kodowania wynikające z użycia zaproponowanych technik predykcji są większe dla małych prędkości bitowych. Wynika to z faktu, iż koszt zakodowania informacji o ruchu w takim przypadku jest większy.
- Ograniczanie jakości głębi ma generalnie negatywny wpływ na skuteczność zaproponowanych algorytmów predykcji, zmniejszając efektywność kompresji kodeka. Jednakże, wpływ na osiągane zyski kodowania jest nieznaczny.
- Wpływ obu zaproponowanych technik predykcji na efektywność kompresji wielowidokowych kodeków wizyjnych uznawanych za obecny stan techniki jest bardzo zbliżony.

Ocena eksperymentalna efektywności zaproponowanych między-widokowych algorytmów reprezentacji informacji o ruchu prowadzi do wniosku, iż techniki zaproponowane przez autora mogą również zostać z powodzeniem zaadaptowane do predykcji innych elementów składni w kodowaniu wielowidokowych sekwencji wizyjnych. Jest to jeden z możliwych kierunków dalszych badań.

Obecnie trwają prace nad nową normą kodowania trójwymiarowych sekwencji wizyjnych. Techniki między-widokowej predykcji informacji o ruchu przedstawione w niniejszej rozprawie zostały włączone do modelu odniesienia dla nowej technologii kodowania (ang. *Test Model under Consideration*) [ISO11a], który stanowi punkt startowy dla procesu tworzenia nowej normy kodowania trójwymiarowych sekwencji wizyjnych kompatybilnych z techniką kodowania HEVC, opracowywanej w ramach prac komisji MPEG. W rezultacie, rozwiązania zaproponowane w tej pracy mają dużą szansę by stać się częścią powstającej normy kodowania trójwymiarowych sekwencji wizyjnych.

# Literatura

- [Bjo01] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves", VCEG Contribution VCEG-M33, Austin, USA, April 2001.
- [Bos11] F. Bossen, "Common test conditions and software reference configurations", ISO/IEC JTC1/SC29/WG11, JCTVC-F900, Geneva, Switzerland, July 2011.
- [Chen07] Y.-W. Chen, W.-H. Peng, S.-Y. Lee, "Adaptive macroblock and motion skip flags for Multiview Video Coding (MVC)", Joint Video Team (JVT), Doc. JVT-X047, Geneva, Switzerland, June 2007.
- [Chen09] Y. Chen, P. Pandit, S. Yea, "WD 4 reference software for MVC", Joint Video Team (JVT), Doc. JVT-AD207, Geneva, Switzerland, February 2009.
- [Dom98] Domański M., "Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych", Wydawnictwo Politechniki Poznańskiej, Poznań 1998.
- [Dom10] M. Domański, „Obraz cyfrowy”, Wydawnictwo Komunikacji i Łączności, 2010.
- [Dom11] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "Multiview HEVC – experimental results", ISO/IEC JTC1/SC29/WG11, JCTVC-G582, Geneva, Switzerland, November 2011.
- [Dom12] M. Domański, „HEVC – nowa generacja technik kompresji obrazu”, Przegląd Telekomunikacyjny No. 4/2012, vol. LXXXV, p. 103, 2012.
- [Feck05] U. Fecker, A. Kaup, "Statistical analysis of multi-reference block matching for dynamic light field coding", 10th International Fall Workshop - Vision, Modeling, and Visualization (VMV), pp. 445-452, Erlangen, Germany, November 2005.
- [Guo06] X. Guo, Y. Lu, F. W. W. Gao, "Inter-view direct mode for Multiview Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, No. 12, pp. 1527–1532, December 2006.
- [ISO10] ISO/IEC JTC1/SC29/WG11, "Joint call for proposals on video compression technology", Doc. N11113, Kyoto, Japan, January 2010.
- [ISO11] ITU-T and ISO/IEC JTC1, "Advanced video coding for generic audiovisual services", ITU-T Rec. H.264 – ISO/IEC 14496-10 AVC, 2011.
- [ISO11a] ISO/IEC MPEG, "Technology under consideration for HEVC based 3D video coding", MPEG N12350, December 2011.
- [Kauf07] P. Kauff, N. Atzpadin, C. Fehn, M. Mueller, O. Schreer, A. Smolic, R. Tanger, "Depth Map Creation and Image Based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability", Signal Processing: Image Communication, Vol. 22, No. 2, Special Issue on 3D Video and TV, pp. 217-234, February 2007.
- [Koo06] H. S. Koo, Y. J. Jeon, B. M. Jeon, "Motion skip mode for MVC", Joint Video Team (JVT), Doc. JVT-U091, Hangzhou, China, October 2006.
- [Koo07] H. S. Koo, Y. J. Jeon, B. M. Jeon, "MVC motion skip mode", Joint Video Team (JVT), Doc. JVT-W081, San Jose, USA, April 2007.
- [Lin07] S. Lin, S. Gao, H. Yang, L. Xiong, "RDV based MVC motion skip mode", Joint Video Team (JVT), Doc. JVT-Y036, Shenzhen, China, October 2007.
- [Mart06] E. Martinian, A. Behrens, J. Xin, A. Vetro, "View synthesis for multiview video compression", in Proc. Picture Coding Symposium PCS, Beijing, China, 2006.
- [Merk07] P. Merkle, A. Smolic, K. Muller, T. Wiegand, "Efficient prediction structures for Multiview Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 11, pp.1461-1473, Nov. 2007.
- [MP08] "Joint multiview video model (JMVM) 8.0", JVT-AA207, Geneva, Switzerland, April 2008.

- [MP11] “WD3: working draft 3 of High-Efficiency Video Coding”, ISO/IEC JTC1/SC29/WG11, JCTVC-E603, Geneva, Switzerland, March 2011.
- [MP11a] [https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/HM-3.0rc2/](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-3.0rc2/), HEVC reference software, HM 3.0.
- [MP11b] “Call for proposals on 3d video coding technology”, ISO/IEC JTC1/SC29/WG11, MPEG2011/N12036, Geneva, Switzerland, March 2011.
- [Ohm04] J.-R. Ohm, “Multimedia communication technology. representation, transmission and identification of multimedia signals”, Springer, 2004.
- [Pan08] P. Pandit, A. Vetro, Y. Chen, ”JMVM 8 software”, ITU-T and ISO/IEC JTC1, JVT-AA208, Geneva, Switzerland, April 2008.
- [Pat07] S. Pateux, “An excel add-in for computing Bjontegaard metric and its evolution”, VCEG Contribution VCEG-AE07, Marrakech, January 2007.
- [Pfis04] H. Pfister, A. Vetro, H. Sun, “3D TV system”, ISO/IEC JTC1/SC29/WG11 MPEG04/M10624, Munich, Germany, March 2004.
- [Ric02] I. Richardson, “Video codec design”, John Wiley & Sons Ltd., Chichester 2002.
- [Ric10] I. Richardson, “The H.264 Advanced Video Compression standard second edition”, John Wiley & Sons, Chichester 2010.
- [Shi00] Y. Q. Shi, H. Sun, "Image and video compression for multimedia engineering", CRC Press, Boca Raton 2000.
- [Shim07] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura, Y. Yashima, “View scalable multiview video coding using 3D warping with depth map”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17 (11), pp. 1485–1495, 2007.
- [Ska93] W. Skarbek, “Metody reprezentacji obrazów cyfrowych”, Warszawa, 1993.
- [Smo06] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, “3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards,” Proc. of ICME 2006, pp. 2161-2164, Toronto, Canada, July 2006.
- [Smo07] A. Smolic, K. Mueller, P. Merkle, N. Atzpadin, C. Fehn, M. Mueller, O. Schreer, R. Tanger, P. Kauff, T. Wiegand, “Multi-view video plus depth (MVD) format for advanced 3D video systems”, JVT-W100, San Jose, USA, April 2007.
- [Song07] H.-S. Song, W.-S. Shim, Y.-H. Moon, J.-B. Choi, “Macroblock information skip for MVC”, Doc. JVT-V052, Marrakech, Morocco, January 2007.
- [Su06] Y. Su, A. Vetro, A. Smolic, “Common test conditions for multiview video coding”, JVT-T207, Klagenfurt, Austria, 2006.
- [Sue10] K. Suehring, J.-R. Ohm, “Results of breakout on decoder speed”, JCT-VC-A201, Dresden, Germany, April 2010.
- [Sul10] G. J. Sullivan, J.-R. Ohm, “Meeting report of the first meeting of the Joint Collaborative Team on Video Coding”, JCT-VC-A200, Dresden, Germany, April 2010.
- [Tan08] TK Tan, G. Sullivan, T. Wedi, “Recommended simulation conditions for coding efficiency experiments”, VCEG Contribution VCEG-AJ10, San Diego, July 2008.
- [Vet07] A. Vetro, S. Yea, W. Matusik, H. Pfister, M. Zwicker, “Antialiasing for 3D displays”, ITU-T and ISO/IEC JTC1, Doc. JVT-W060, San Jose, USA, April 2007.
- [Vet11] A. Vetro, T. Wiegand, G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard", Proceedings of the IEEE , Vol.99, No.4, pp.626-642, April 2011.
- [Yang09] H. Yang, Y. Chang, J. Huo, “Fine-granular motion matching for inter-view motion skip mode in Multiview Video Coding”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 19 (6), pp. 887–892, 2009.
- [Yea07] S. Yea, A. Vetro, M. Zwicker, “MVC showcase for SEI on scene and acquisition info”, ITU-T and ISO/IEC JTC1, JVT-X074, Geneva, Switzerland, June 2007.

# Dokumentacja dorobku naukowego

## autora

### Materiały konferencji międzynarodowych opublikowane w bazie IEEEXplore

- [1] J. Konieczny, A. Łuczak, "Motion estimation algorithm for scalable hardware implementation", 27th Picture Coding Symposium, PCS 2009, 6-8 maj 2009, Chicago, USA.
- [2] J. Konieczny, M. Domański, "Depth-Based Inter-View Prediction of Motion Vectors for Improved Multiview Video Coding", The true vision - Capture, Transmission and Display of 3D Video, 3DTV-CON 2010, 7-9 czerwca 2010, Tampere, Finlandia.
- [3] J. Konieczny, M. Domański, "Extended Inter-View Direct Mode for Multiview Video Coding", International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011, 22-27 maja 2011, Praga, Czechy.
- [4] J. Konieczny, M. Domański, "Depth-Enhanced Compression for 3D Video", Visual Communications and Image Processing, VCIP 2011, 6-9 listopada 2011, Tainan, Tajwan.
- [5] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "New Coding Technology for 3D Video within Depth Maps as Proposed for Standardization within MPEG", International Conference on Systems, Signals and Image Processing, IWSSIP 2012, 11-13 kwietnia, Wiedeń, Austria, s. 415-418.
- [6] M. Domański, T. Grajek, D. Karwowski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "Coding of Multiple Video+Depth Using HEVC Technology and Reduced Representations of Side Views and Depth Maps", 29<sup>th</sup> Picture Coding Symposium, PCS 2012, 7-9 maja 2012, Kraków, Polska, s. 5-8.
- [7] J. Konieczny, M. Domański, "Depth-Based Inter-View Motion Prediction for HEVC-Based Multiview Video Coding", 29<sup>th</sup> Picture Coding Symposium, PCS 2012, 7-9 maja 2012, Kraków, Polska, s. 33-36.
- [8] J. Stankowski, M. Domański, O. Stankiewicz, J. Konieczny, J. Siast, K. Wegner, "Extensions of the HEVC technology for efficient multiview video coding", 2012 IEEE International Conference on Image Processing, ICIP 2012, 30 września – 3 października, 2012, Orlando, Floryda, USA.
- [9] M. Domański, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "3D video compression by coding of disoccluded regions", 2012 IEEE International Conference on Image Processing, ICIP 2012, 30 września – 3 października, 2012, Orlando, Floryda, USA.
- [10] S. Maćkowiak, J. Konieczny, M. Kurc, P. Maćkowiak, "A complex system for football player detection in broadcasted video", International Conference on Signals and Electronic Systems, ICSES 2010, s.119-122, 7-10 września 2010.



- [11] S. Maćkowiak, J. Konieczny, "Player Extraction in Sports Video Sequences", International Conference on Systems, Signals and Image Processing, IWSSIP 2012, 11-13 kwietnia, Wiedeń, Austria, s. 423-426.

### **Dokumenty posiedzeń grupy MPEG, publikowane w bazie MPEG**

- [12] J. Konieczny, M. Domański, "Inter-View Direct Mode for Multiview Video Coding", ISO/IEC JTC1/SC29/WG11, MPEG2010/M17800, lipiec 2010, Genewa, Szwajcaria.
- [13] J. Konieczny, M. Domański, "Experimental Results for Inter-View Direct Mode with Decompressed Disparity Maps for Multiview Video Coding", ISO/IEC JTC1/SC29/WG11, MPEG2010/M18552, październik 2010, Guangzhou, Chiny.
- [14] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "Multiview HEVC – experimental results", ISO/IEC JTC1/SC29/WG11, JCTVC-G582, Genewa, Szwajcaria, listopad 2011.
- [15] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, "Technical description of Poznań University of Technology proposal for call on 3d video coding technology", ISO/IEC JTC1/SC29/WG11, MPEG/M22697, Genewa, Szwajcaria, listopad 2011.
- [16] K. Wegner, J. Siast, J. Konieczny, O. Stankiewicz, M. Domanski, "Poznan University of Technology tools for 3DV coding integrated into 3D-HTM", ISO/IEC JTC1/SC29/WG11 MPEG 2011 / M23783, San Jose, USA, luty 2012.

### **Wykaz publikacji w czasopismach i tomach zwartych**

- [17] J. Konieczny, A. Łuczak, „Skalowalny algorytm estymacji ruchu dla systemów rozproszonych”, Pomiary, Automatyka, Kontrola 2009, Wydawnictwo PAK, nr 08, s. 684-686, 2009, Warszawa.
- [18] M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, O. Stankiewicz, K. Wegner, „An Experimental Free-view Television System”, Image Processing and Communications Challenges, Wydawnictwo EXIT, Warszawa, 2009, s. 169-176.
- [19] S. Maćkowiak, J. Konieczny, M. Kurc, P. Maćkowiak, "Football Player Detection in Video Broadcast", L. Bolc, R. Tadeusiewicz, and L.J. Chmielewski (eds.): Computer Vision and Graphics, Proc. ICCVG 2010, Lecture Notes in Computer Science, Springer-Verlag (2010).
- [20] S. Maćkowiak, J. Konieczny, M. Kurc, P. Maćkowiak, "System detekcji i śledzenia piłkarzy w transmisjach widowisk sportowych w cyfrowym sygnale wizyjnym", Elektronika 12/2010, s. 13-15.
- [21] A. Łuczak, M. Stępniewska, J. Siast, M. Domański, O. Stankiewicz, M. Kurc, J. Konieczny, "Network-on-Multi-Chip (NoMC) with monitoring and debugging support", Journal of Telecommunications and Information Technology, nr 3/2011, s. 81.
- [22] M. Domański, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, Jakub Stankowski, R. Ratajczak, T. Grajek, "High Efficiency 3D Video Coding using prediction by view synthesis", Submitted to IEEE for publication in a journal.

## **Zgłoszenia patentowe**

- [23] M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, US 13/680652, USA - "Image coding metod", 2012.
- [24] M. Domański, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, US 13/680740, USA - "Method for predicting a shape of an encoded area using a depth map", 2012.
- [25] M. Domański, T. Grajek, J. Konieczny, M. Kurc, A. Łuczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, US 13/680822, USA - "Method for coding of stereoscopic depth", 2012.
- [26] J. Konieczny, M. Domański, PAT1314/P.390327 – „Sposób kodowania informacji o ruchu w sekwencjach wielowidokowych z dostępną informacją o rozbieżności lub głębi stereoskopowej”, 2010.
- [27] M. Domański, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1461/P.397012 – “Sposób przetwarzania obrazu zsyntezowanego”, 2011.
- [28] M. Domański, T. Grajek, J. Konieczny, M. Kurc, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1462/P.397014 – „Sposób alokacji prędkości bitowej przy kodowaniu sekwencji wielowidokowych z informacją przestrzenną”, 2011.
- [29] M. Domański, T. Grajek, J. Konieczny, M. Kurc, A. Łuczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1463/P.397013 – „Sposób wyznaczania parametrów kwantyzacji sterujących dekwantyzatorem sygnału resztkowego obrazu z wykorzystaniem mapy głębi odpowiadającej dekodowanemu obrazowi”, 2011.
- [30] M. Domański, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1464/P.397010 – „Sposób predykcji kształtu obszaru kodowanego z wykorzystaniem map głębi”, 2011.
- [31] M. Domański, T. Grajek, J. Konieczny, M. Kurc, A. Łuczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1465/P.397016 – „Sposób kodowania głębi stereoskopowej”, 2011.
- [32] M. Domański, T. Grajek, J. Konieczny, M. Kurc, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1466/P.397015 – „Zastosowanie cech sygnału różnicowego mapy głębi do sterowania koderem sekwencji wizyjnych z informacją przestrzenną”, 2011.
- [33] M. Domański, T. Grajek, K. Klimaszewski, J. Konieczny, M. Kurc, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1469/P.397011 – „Sposób międzyobrazowej predykcji mapy odległości”, 2011.
- [34] M. Domański, T. Grajek, D. Karwowski, J. Konieczny, M. Kurc, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1473/P.397009 – „Sposób kodowania parametrów kamer”, 2011.
- [35] M. Domański, T. Grajek, K. Klimaszewski, J. Konieczny, M. Kurc, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1474/P.397017 – „Sposób przetwarzania map rozbieżności”, 2011.
- [36] M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, K. Wegner, PAT1475/P.397008 – “Sposób kodowania obrazu”, 2011.

## **Opracowania dla przemysłu**

1. „Opracowanie raportu możliwości funkcjonalnych odbiorników TVC, implementacja oprogramowania testowego MHP oraz sporządzenie koncepcji realizacji projektu telemetrii”, prace badawcze dla INEA S.A., 2008-2009.
2. “State-of-the-art video segmentation and classification algorithms are to be implemented and evaluated in the context of CBR and CBVR”, projekt badawczy realizowany dla Fraunhofer Heinrich Herz Institute, 2009-2010.

## **Nagrody**

*Medal Rektora Politechniki Poznańskiej dla wyróżniającego się absolwenta, 8 października 2008, Poznań.*

## **Autorstwo recenzji**

Recenzja dla czasopism:

EURASIP Journal on Image and Video Processing Research – 1 artykuł.

Recenzja dla konferencji międzynarodowych:

Picture Coding Symposium – 3 artykuły.

## **Udział w projektach badawczych**

1. Projekt Zamawiany – „Usługi i sieci teleinformatyczne następnej generacji – aspekty techniczne, aplikacyjne i rynkowe”, Nr pracy 84-007/PZ, Zam. Ministerstwo Nauki i Szkolnictwa Wyższego, 02.01.2008-31.12.2010.
2. Projekt Rozwojowy – „Model aplikacyjny transmisji obrazów wielowidokowych”, Nr pracy 84-0514/PR, Zam. Ministerstwo Nauki i Szkolnictwa Wyższego, 16.04.2007-15.02.2010.
3. Projekt Badawczy – “State-of-the-art video segmentation and classification algorithms are to be implemented and evaluated in the context of CBR and CBVR”, Zam. Fraunhofer Heinrich Herz Institute, 2009-2010.
4. Projekt Rozwojowy – „Implementacja kodeków wizyjnych o wysokiej efektywności kompresji”, Nr pracy 84-0899/PR, Zam. Ministerstwo Nauki i Szkolnictwa Wyższego.
5. Projekt Badawczy – „Nowe algorytmy kompresji i odtwarzania obrazu telewizji trójwymiarowej”, Nr pracy 84-3949.

## **Cytowania**

[2] cytowane przez:

1. Seungchul Ryu, Jungdong Seo, Jin Young Lee and Kwanghoon Sohn, “Advanced motion vector coding framework for multiview video sequences”, *Multimedia Tools and Applications*, Springer Netherlands, ISSN: 1380-7501, DOI: 10.1007/s11042-011-0930-y, 2011.
2. Seungchul Ryu, Jungdong Seo, Xingang Liu, Jin Young Lee, Hochen Wey, Kwanghoon Sohn, "Analysis of Motion Vector Predictor in Multiview Video Coding System," *Parallel and Distributed Processing with Applications Workshops (ISPAW)*, 2011, s.183-188, maj 2011.

3. Shimizu, S., Kimata, H., Sugimoto, S., Matsuura, N., "Decoder-side macroblock information derivation for efficient multiview video plus depth map coding", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, s. 1-4, maj 2011.
4. Schwarz, H., Wiegand, T., "Inter-view prediction of motion data in multiview video coding", *29<sup>th</sup> Picture Coding Symposium, PCS 2012, 7-9 May 2012, Kraków, Poland*, s. 101–104.