Instytut Telekomunikacji Multimedialnej
Wydział Informatyki i Telekomunikacji
Politechnika Poznańska

Rozprawa doktorska

Systemy wizyjne swobodnego punktu widzenia o wysokiej postrzeganej jakości usług

Adam Grzelka

Promotor: prof. dr hab. inż. Marek Domański
Promotor pomocniczy: dr hab. inż. Olgierd Stankiewicz

Poznań 2024

Rodzicom
Hannie i Jerzemu

Żonie Aleksandrze

Prof. drowi hab. inż. Markowi Domańskiemu
i zespołowi Instytutu Telekomunikacji Multimedialnej

Spis treści

Streszczenie

Rozprawa dotyczy systemów swobodnego punku widzenia, które są systemami wizyjnymi umożliwiającymi użytkownikowi wybór kierunku i miejsca obserwowania prezentowanej sceny. Autor podejmuje za cel zapewnienie wysokiej postrzeganej jakości usług w takich systemach, który realizuje poprzez szereg propozycji oryginalnych rozwiązań. Rozwiązania te dotyczą między innymi optymalizacji ustawienia kamer systemu wielokamerowego, zbadania wpływu stratnej kompresji, opóźnień transmisji do terminala użytkownika oraz realizacji nowego systemu wielokamerowego i serwera swobodnej nawigacji.

Wśród najważniejszych osiągnięć przedstawionych w pracy wymienić można:

- opracowanie modeli teoretycznych wpływu przysłonięć na jakość widoku wirtualnego oraz zależności jakości widoku wirtualnego od jakości widoków przesyłanych wraz z ich eksperymentalną walidacją,
- budowę nowego, kompletnego systemu wielokamerowego oraz serwera swobodnej nawigacji,
- eksperymentalne zbadanie różnych aspektów systemów swobodnego punktu widzenia w kontekście zapewnienia wysokiej postrzeganej jakości usług.

Abstract

Free-viewpoint Video Systems with High Quality of Experience

The dissertation deals with free-viewpoint video systems, which allow a user to choose the location and viewing direction for observing the presented scene. The author proposes a set of original solutions to obtain a high-quality experience in such systems. These proposals relate to optimizing the positioning of the cameras in a multi-camera system, investigating the impact of lossy compression, the transmission delay to the user terminal, and implementing a new multi-camera system and a free-viewpoint server.

The major achievements presented in the paper include:

- development of theoretical models of the effect of occlusions on the quality of the virtual view and the dependence of the quality of the virtual view on the quality of the transmitted views and their experimental validation,
- construction of a new, complete multi-camera system and a free navigation server,
- experimental investigation of various aspects of free viewpoint systems in the context of providing high perceived quality of service.

Słownik terminów

W niniejszej rozprawa przyjęto konwencję pisania zmiennych kursywą. Wyjątek stanowią zmienne będące nazwami metryk (np. PSNR).

3DoF — Three degrees of freedom — trzy stopnie swobody — termin określający systemy wizyjne umożliwiające użytkownikowi swobodny obrót wokół trzech osi w scenie trójwymiarowej

6DoF — Six degrees of freedom — sześć stopni swobody — termin określający systemy wizyjne umożliwiające użytkownikowi swobodny obrót wokół trzech osi w scenie trójwymiarowej wraz z możliwością przemieszczania się

CTC — Common Test Conditions — zbiór warunków prowadzenia eksperymentów określany przez grupę MPEG dla badań nad daną techniką poddawaną normalizacji

DIBR — Depth-Image-Based Rendering — technika generowania obrazu wirtualnego wykorzystująca informację o głębi w scenie

FTV — Free-viewpoint Television — telewizja swobodnego punktu widzenia

HMD — Head-Mounted Display — wyświetlacz nagłowny

IEC — International Electrotechnical Commission — Międzynarodowy Komitet Elektrotechniczny

ISO — International Organization for Standardization — Międzynarodowa Organizacja Normalizacyjna

ITU-R — International Telecommunication Union (Radiocommunication Sector) — Międzynarodowy Związek Telekomunikacyjny (Sektor Radiokomunikacji)

ITU-T — International Telecommunication Union (Telecommunication Standarization Sector) — Międzynarodowy Związek Telekomunikacyjny (Sektor Normalizacji Telekomunikacji)

IV-PSNR — Immersive Video PSNR — miara jakości obrazu bazująca na wskaźniku PSNR i dostosowana do charakterystyki obrazów syntezowanych

JPEG — Joint Photographic Experts Group — grupa robocza ISO/IEC zajmująca się rozwojem kodowania fotografii, jak również technika stratnej kompresji obrazów nieruchomych

MJPEG — Motion JPEG — technika stratnej kompresji sekwencji wizyjnej korzystająca z techniki JPEG

MPEG — Moving Picture Experts Group — grupa robocza ISO/IEC zajmująca się rozwojem i normalizacją technik kodowania multimediów

MVD — Multiview Video plus Depth — obraz wielowidokowy wraz z głębią — sposób reprezentacji sekwencji wielowidokowych

MIV — MPEG Immersive Video — norma kodowania wszechogarniającej treści wizyjnej — ujęta w normie ISO / IEC 23090, część 12

PSNR — Peak Signal to Noise Ratio — miara jakości obrazu określona przez stosunek mocy maksymalnej sygnału do mocy szumu

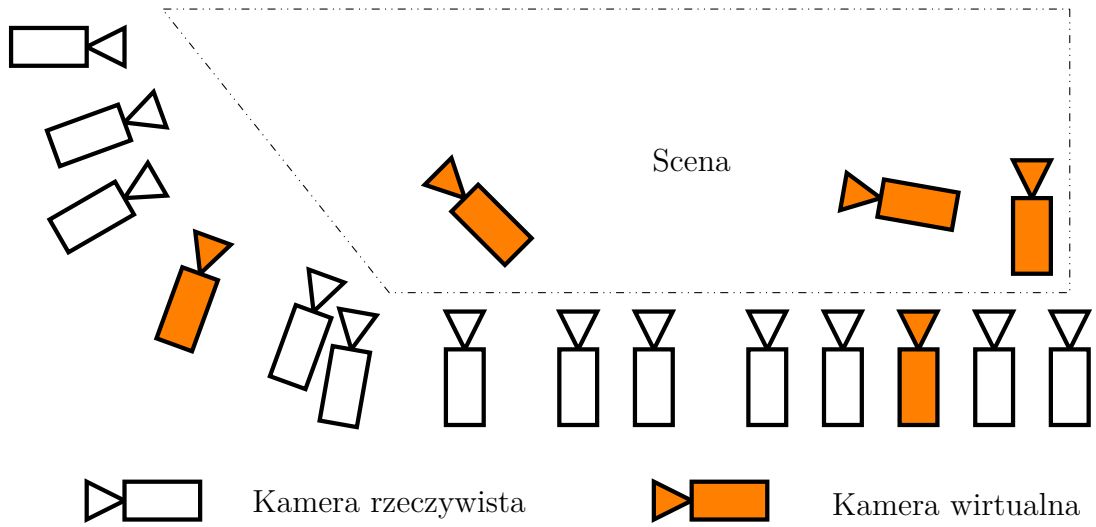
OPUS — technika stratnej kompresji dźwięku, pełniej określana jako Opus Interactive Audio Codec

QoE — Quality of Experience — postrzegana jakość usług

1 Wprowadzenie

1.1 Systemy swobodnego punktu widzenia

Niniejsza rozprawa dotyczy systemów swobodnego punktu widzenia oraz swobodnej nawigacji (rysunek 1.1). Systemy takie umożliwiają użytkownikowi wybór kierunku i miejsca, z którego chce obserwować prezentowaną scenę. Wyświetlany obraz może pochodzić bezpośrednio z kamery nagrywającej scenę (nazywanej **kamerą rzeczywistą**) lub też być stworzony sztucznie (algorytmicznie), tak aby odzwierciedlać obraz, który byłby widziany w danym miejscu przez nieistniejącą fizycznie kamerę (**kamerę wirtualną**). Obraz z kamery wirtualnej jest wyliczany na podstawie danych zebranych z kamer rzeczywistych, a jego parametry (miejsce, kierunek, rozmiar ramki) mogą być wybierane w odbiorniku. Użytkowników systemu może być wielu i mogą oni niezależnie i swobodnie przemieszczać się w scenie. Taką funkcjonalność określa się mianem **swobodnej lub wirtualnej nawigacji** (ang. free navigation, virtual navigation). Badane systemy są rozwinięciem klasycznych systemów dostarczania treści wizyjnej do odbiorcy [Pal+17][Dom+16a][Dom+17], znanych w literaturze jako telewizja swobodnego punktu widzenia (ang. Free-Viewpoint Television — FTV) [Tan10][FBP06], oferując funkcjonalność określaną również jako wizja ze swobodnym punktem widzenia (ang. Free-Viewpoint Video — FVV) [LTT15][LZ17].



Rysunek 1.1: Rzut z góry — system swobodnego punktu widzenia

Niniejsza praca skupia się na systemach swobodnej nawigacji ukierunkowanych na prezentację scen rzeczywistych. Z tego względu zakładane jest wykorzystanie **danych rzeczywistych**, czyli rejestrowanych za pomocą systemów wielokamerowych. Aby nie ograniczać ogólności rozważań, nie są rozpatrywane formaty i metody specyficzne dla trójwymiarowej grafiki komputerowej, np. rendering na podstawie modeli siatkowych. Nie wyklucza to jednak wykorzystywania **danych sztucznych**, np. na potrzeby eksperymentów czy w celu pozyskania sekwencji testowych. Wytworzenie takich danych jest dużo łatwiejsze niż danych rzeczywistych, a zaprojektowana wirtualna scena pozwala na umieszczenie dowolnej liczby i rodzaju kamer. Nagrywanie (wytwarzanie sztucznych danych) można łatwo powtórzyć i zmodyfikować którekolwiek parametry, np. ustawienie systemu wielokamerowego, typy wykorzystanych kamer, rozmieszczenie obiektów. Sekwencje wygenerowane sztucznie dodatkowo mogą zawierać dane będące informacją przestrzenną sceny — są one często odniesieniem w wielu pracach naukowych, gdyż powinny zawierać poprawną wartość reprezentacji przestrzennej (ang. ground-truth). Takie sztucznie wytworzone sekwencje (np. [Kov+15][SS13][Goo+14], rozdział 3.1) w tej pracy również zostały wykorzystane.

Badane systemy umożliwiają swobodną nawigację w wirtualnej scenie reprezentującej rzeczywiste wydarzenie, dynamicznie zmienne w czasie (np. imprezę sportową, koncert, widowisko) [Zit+04][Goo+12]. Zagadnienia związane z takimi systemami są przedmiotem prac badawczych, w tym prowadzonych przez grupy robocze ISO/IEC, dotyczących normalizacji m.in. trójwymiarowej reprezentacji sceny, kompresji, transmisji treści wszechkierunkowej [Yam01][Boy+21][KSN13][48220][00420]. Nowe systemy telekomunikacji bezprzewodowej 5G poprawiają ich warunki rozwoju, gdyż zwiększają potencjalne grono odbiorców, mobilność oraz dostępność usługi swobodnego punktu widzenia [Pé+22][Zhe20][LB19].

Główną funkcjonalnością omawianych systemów jest umożliwienie użytkownikowi swobodnego nawigowania w wirtualnej scenie. W tym celu wykorzystuje się **terminal użytkownika**, służący do prezentacji wirtualnego widoku i sterowania swobodne nawigowaną pozycją. Taki terminal może być zrealizowany na wiele sposobów, co wiąże się ze zróżnicowanymi wymaganiami dotyczącymi danych dostarczanych z systemu. Terminal użytkownika używający do prezentacji treści ekran monitora lub telefonu wymaga obrazu o odpowiednim rozmiarze ramki. Urządzenie użytkownika, pozwalające na uzyskanie efektu trójwymiarowego, takiego jak monitor autostereoskopowy [Ure+11], monitor stereoskopowy (migawkowy, polaryzacyjny) [Gen13] czy coraz

popularniejszy wyświetlacz nagłowny (ang. HMD — Head-mounted display) [Shi02], wymaga dostarczenia treści wielowidokowej lub reprezentacji trójwymiarowej. W systemach swobodnego punktu widzenia wyznaczenie danych dla dowolnego urządzenia użytkownika jest możliwe poprzez np. umieszczenie w scenie dowolnej i odpowiedniej kamery lub wielu kamer wirtualnych.

Funkcjonalność systemu swobodnego punktu widzenia jest podobna do usług dostępnych w systemach wykorzystujących wirtualną rzeczywistość, takich jak gry komputerowe lub symulatory lotu. Wirtualna rzeczywistość wytworzona na podstawie grafiki komputerowej pozwala twórcy systemu na dowolną prezentację treści użytkownikowi. Taka funkcjonalność jest bardzo często udostępniana (przez twórcę) użytkownikowi systemu wykorzystującemu wirtualną rzeczywistość, np. w formie wyboru wyświetlanego wirtualnego widoku czy możliwości swobodnej nawigacji w scenie gry komputerowej. Uzyskanie takiego samego efektu dla danych rzeczywistych jest obecnie możliwe np. dla statycznej sceny przy zastosowaniu skanowania laserowego (LIDAR) [Lv+17]. Skaner pozwala na zebranie danych w postaci spójnej chmury punktów o precyzyjnie zmierzonych położeniach. Punkty te mogą zostać zaprezentowane użytkownikowi w formie statycznej sceny. Jakość tych danych zależy od liczby zebranych punktów, a skanowanie może trwać nawet kilka godzin. Otrzymanie podobnego efektu dla sekwencji obrazów w przypadku scen zarejestrowanych systemem wielokamerowym jest możliwe dzięki np. **estymacji map głębi**, czyli rekonstrukcji informacji przestrzennej na podstawie zarejestrowanych widoków. Takie systemy są przedmiotem badań podjętych w tej rozprawie.

Podsumowując, praca dotyczy systemów swobodnej nawigacji, w których możliwe jest dostarczanie danych zarejestrowanych rzeczywistym systemem wielokamerowym. Celem jest poszerzenie stanu wiedzy w celu poprawy jakości dostarczanej usługi swobodnego punktu widzenia w takich systemach.

1.2 Postrzegana jakość usług

Prace nad pierwszymi algorytmami umożliwiającymi syntezę wirtualnego widoku oraz pierwsze systemy wielokamerowe pozwalające na wielowidokową akwizycję powstały już przed rokiem 2000 [KRN97][Ved+00]. Składały się one z kilku do nawet kilkudziesięciu kamer i pozwalały na akwizycję i przenoszenie obiektu do wirtualnej sceny, a część działała również w czasie rzeczywistym [Mat+00][Car+03][YWB02]. Mimo że pierwsze próby stworzenia takich systemów zostały podjęte wiele lat temu

[Tan10][FBP06][Goo+12][Zit+04], uzyskanie wysokiej jakości usługi wirtualnej nawigacji wciąż jest znacznym wyzwaniem. Szczególną trudność sprawia realizacja takiej usługi dla wydarzeń nagrywanych i realizowanych na żywo. Przyczyną wspomnianych trudności jest m.in. duża złożoność obliczeniowa procesu tworzenia obrazów wirtualnych, zniekształcenia wprowadzane przez kolejne etapy przetwarzania wizji wielowidokowej czy wrażliwość użytkowników na opóźnienia wyświetlanego widoku, szczególnie w przypadku użycia terminali użytkownika nowej generacji takich jak wyświetlacze nagłowne. Rozwój badanych systemów pozwala na rejestrację scen w coraz lepszej jakości, a nowe algorytmy przetwarzania danych i ich prezentacji umożliwiają poprawę postrzeganej jakości.

Jakość obrazu najczęściej definiuje się przez parametry związane z obiektywną jakością obrazu (np. metrykę PSNR) oraz parametrami jego wyświetlania (liczba obrazów na sekundę, procent gubionych ramek itp.). Tak definiowana jakość nie jest odpowiednia dla przypadku systemów telewizji swobodnego punktu widzenia. Podstawowe problemy utrudniające uzyskanie wysokiej jakości usługi w systemach wirtualnej rzeczywistości wynikają z wad obrazu wyświetlanego użytkownikowi systemu. Dlatego w pracy podjęto rozważania dotyczące postrzeganej jakości usług (QoE — Quality of Experience), definiowanej w odniesieniu do parametrów subiektywnych (subiektywna jakość obrazu, subiektywna dokuczliwość opóźnień itp.). Wady te mogą być powodowane m.in. przez następujące czynniki:

- zniekształcenia wynikające z zastosowania kompresji stratnej,
- błędy estymacji głębi,
- błędy syntezy obrazu wirtualnego,
- opóźnienia przetwarzania i reprezentacji obrazu wirtualnego.

Jakość obrazu prezentowana użytkownikowi zależy też od rodzaju odbiornika, który może maskować część zniekształceń w związku ze sposobem swojego działania lub algorytmem prezentacji wizji [OFI12]. Jakość ta może zmniejszać się również na skutek konfliktu jednoocznych i dwuocznych wskazówek głębi, małą rozdzielczość matryc i widoków czy małą wartość luminancji [Gen13][Dom10]. Oglądanie obrazu stereoskopowego może też być męczące dla wzroku [How11], co również wpływa negatywnie na postrzeganą jakość.

Podczas badań prezentowanych w niniejszej pracy skupiono się na pomiarze dwóch typów zniekształceń wpływających na postrzeganą jakość usługi (QoE). Pierwszy związany jest z jakością dostarczanego widoku i reprezentuje wszystkie wady powodujące błędne wartości punktów wirtualnego obrazu. Drugi typ związany jest z kierunkiem i pozycją prezentowanego widoku oraz dokuczliwością i wpływem opóźnień prezentacji treści w systemach swobodnego punktu widzenia.

Jakość wyświetlanego obrazu została wyznaczona metrykami takimi jak PSNR oraz IV-PSNR. Pierwsza z nich miarą szeroko stosowaną do pomiaru jakości obrazów, druga to metryka dostosowana do widoków wirtualnych i systemów swobodnego widzenia [DD19] (rozdział 4 oraz 5). Metryki oceny jakości sekwencji wizyjnych zostały przedstawione w rozdziale 3.2.

Jakość swobodnej nawigacji zmierzono z wykorzystaniem wyświetlacza nagłownego. Ruch użytkownika, który z niego korzysta, sprawia, że system musi reagować na pozycję i obrót głowy. Uzyskanie wysokiej jakości usługi swobodnej nawigacji w wyświetlaczu nagłownym wymaga więc możliwie płynnego i natychmiastowego wyświetlenia żądanego obrazu. Konflikt między ruchem głowy a bodźcami ze zmysłu wzroku, w związku z opóźnieniem widoku dostarczonego przez wyświetlacz nagłowny, powoduje silne uczucie dyskomfortu [Grz+19a]. W rozdziale 6 przedstawiono wyniki testów subiektywnych z wykorzystaniem wyświetlacza nagłownego.

1.3 Cele i tezy rozprawy

1.3.1 Cele rozprawy

Głównym celem pracy jest znalezienie rozwiązań zapewniających postrzeganą wysoką jakość usług w systemach swobodnej nawigacji. Niniejsza rozprawa ma na celu poszerzenie stanu wiedzy w tej dziedzinie i opracowanie sposobów, które umożliwiają wysoką jakość w praktycznych systemach. Autor przeanalizował, jak poszczególne elementy systemu wpływają na postrzeganą przez użytkownika jakość usługi swobodnej nawigacji. W kolejnych rozdziałach zostały przedstawione badania nad następującymi zagadnieniami w kontekście ich wpływu na postrzeganą jakość:

- przysłonięcia i ustawienie kamer w systemie wielokamerowym (rozdział 4),
- kompresja stratna przesyłanych widoków (rozdział 5),
- opóźnienia dostarczania treści do wyświetlacza nagłownego (rozdział 6),

- budowa praktycznego systemu wielokamerowego oraz implementacja serwera swobodnej nawigacji (rozdział 7).

1.3.2 Tezy rozprawy

Tezy rozprawy są następujące:

- Rejestracja złożonych scen wizualnych przy ustawieniu kamer wokół sceny parami o małej odległości bazowej umożliwia uzyskanie syntezy wirtualnych widoków o lepszej jakości, niż gdy ta sama liczba kamer rzeczywistych jest rozmieszczona równomiernie wokół sceny.

  Przez złożoną scenę rozumie się taką scenę, w której średnio ponad 25% punktów jest przysłoniętych w obrazach kamer rzeczywistych ustawionych równomiernie wokół sceny.

- Jakość obrazów wirtualnych uzyskiwanych ze zdekodowanych widoków zależy silnie od jakości tych widoków i w stosunkowo niewielkim stopniu od zastosowanej techniki kompresji stratnej.

- W systemach wirtualnej rzeczywistości z wyświetlaczami nagłownymi dokuczliwość opóźnień pomiędzy ruchem użytkownika a odświeżeniem ekranu jest większa dla ruchu rotacyjnego niż dla ruchu translacyjnego.

- Za pomocą konsumenckich kamer i typowego komputera osobistego można zbudować system wirtualnej rzeczywistości działający w czasie rzeczywistym, przy założeniu, że model sceny przestrzennej jest przygotowywany wcześniej.

2 Przegląd stanu wiedzy

2.1 Wstęp

W systemach swobodnej nawigacji do akwizycji obrazów wykorzystuje się systemy wielokamerowe [Tan+12][Sta+18a], które mogą być uzupełniane przez kamery głębi [Hor+16][SFO11]. W rozważaniach pracy przyjmuje się, że głębie są wyznaczane przez analizę obrazów, co w praktyce jest bardzo częstym rozwiązaniem.

W systemach swobodnej nawigacji dla wszystkich kamer systemu wielokamerowego należy wyznaczyć:

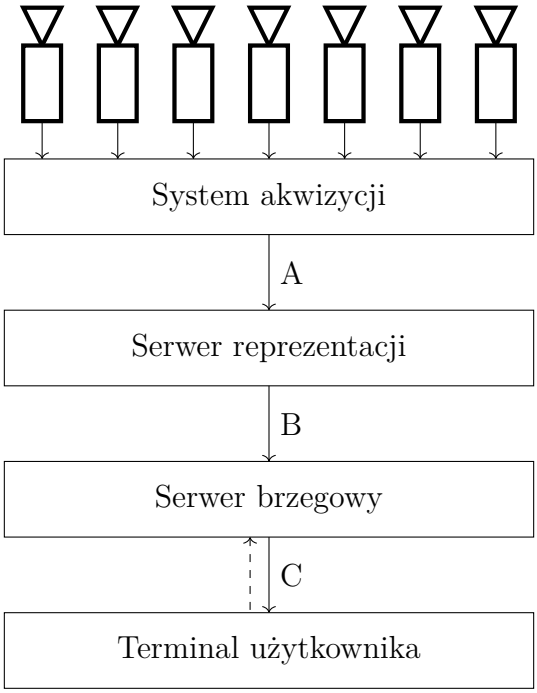
- parametry wewnętrzne (ang. intrinsic parameters) - indywidualne parametry kamery, które nie zależą od jej położenia, takie jak położenie punktu środkowego obrazu, długość ogniskowa obiektywu czy parametry abberacji optycznej obiektywu,
- parametry zewnętrzne (ang. extrinsic parameters) - parametry systemu wielokamerowego określające położenie i ukierunkowanie każdej z kamer w przestrzeni.

Powyższe parametry są wyznaczane w procesie kalibracji systemu wielokamerowego [Zha99][Dom10][Cyg02], który najczęściej obejmuje rejestracje specjalnych sekwencji zawierających np. punkty charakterystyczne, tablice kalibracyjne lub obiekty o znanym kształcie trójwymiarowym. Wiele metod ich wyznaczania pozwala na jednoczesne wyznaczenie zarówno wewnętrznych, jak i zewnętrznych parametrów kamer rzeczywistych [Li+13][Zhu12][Zha99][GSY17][AD03]. Są one najczęściej reprezentowane w postaci macierzowej [HZ04][Cyg02] i zawierają pełną informację o kamerach systemu swobodnej nawigacji. Parametry kamer wirtualnych mogą być dowolne. Wewnętrzne są najczęściej zgodne z wymaganiami wyświetlacza w terminalu użytkownika systemu, a zewnętrzne — są modyfikowane poprzez usługę swobodnej nawigacji.

W niniejszej pracy sformułowanie **obraz z kamery** będzie oznaczało dane zarejestrowane przez kamerę. Słowo **kamera** będzie pełną informacją o kamerze, czyli m.in. o parametrach zewnętrznych i wewnętrznych, polu widzenia czy wielkości rejestrowanej ramki. Kamera wraz z obrazem będzie określana terminem **widok**. Każde z powyższych pojęć może odnosić się zarówno do rzeczywistych, jak i wirtualnych kamer, obrazów, widoków.

2.2 Systemy swobodnej nawigacji

Na rysunku 2.1 został zaprezentowany ogólny schemat systemu swobodnej nawigacji [Goo+12][Tan+12][Dom+15a][Dom+16a][Sta+18d]. W takim systemie terminal użytkownika wyświetla wirtualne widoki odpowiadające za usługę swobodnej nawigacji w scenie oraz możliwe jest wykorzystanie różnego typu kamer, np. perspektywicznych, półsferycznych, sferycznych oraz urządzeń rejestrujące głębię sceny (np. kamer głębi lub sensorów wykorzystywanych w konsolach do gier) [Yan+15][Jia+18]. System akwizycji zbiera dane z kamer oraz je uzupełnia, tworząc pełną informację o wykorzystanych kamerach, ich parametrach oraz całym systemie. Odpowiada również za sterowanie kamerami oraz ich synchronizację (rozdział 2.8).



Rysunek 2.1: Ogólny schemat systemu swobodnej nawigacji z zaznaczeniem trzech rozważanych rodzajów łącz (A, B, C) [Dom+16a][Dom+15a]

Dane z systemu akwizycji przekazywane są do **serwera reprezentacji** przez łącze **A**. Serwer ten wyznacza reprezentację przestrzenną sceny. W zależności od rodzaju kamer informacja o reprezentacji sceny może być estymowana na podstawie widoków z kamer (estymacja głębi — rozdział 2.5), częściowo estymowana, poprawiana lub uwspólniana, jeśli system wielokamerowy zawierał kamery głębi lub inne urządzenia rejestrujące głębię sceny [Jia+18][Xia+13][Yan+15]. Dane wytwarzane

przez serwer reprezentacji mogą być reprezentowane na wiele różnych sposobów. Najpopularniejszy to reprezentacja poprzez wiele obrazów i towarzyszących im map głębi (MVD, ang. Multiview Video plus Depth) [MMW11][Feh03][Mie18]. Inną reprezentacją jest chmura punktów, kiedy to dane ze wszystkich widoków tworzą jeden nieuporządkowany zbiór punktów wraz z ich atrybutami [Wu+17][WHC13]. Kolejnymi możliwymi sposobami reprezentacji informacji trójwymiarowej zarejestrowanej sceny są przestrzenie promieni [Kim+13][DGM16] albo reprezentacje obiektowe [MSH06][Smo+05].

Opisane w tej pracy badania wykorzystują reprezentację sceny za pomocą wielu widoków z kamer i odpowiadających im map głębi (MVD, ang. Multiview Video plus Depth) oraz syntezę obrazu wirtualnego na ich podstawie (DIBR, ang. Depth-Image-Based Rendering) [Sta14][Mie18][Dzi18].

Dane są przesyłane przez łącze **B** do **serwera brzegowego** (ang. edge server), który jest odpowiedzialny za przygotowanie oraz dystrybucję danych przez łącze **C** do **terminala użytkownika**. Proces ich przetwarzania przez serwer zależy od sposobu komunikacji przez łącze C, który jest wynegocjowany przez podłączone terminale. Dwa podstawowe schematy wymiany informacji w tym łączu zostały przedstawione na rysunku 2.2. Serwer brzegowy może przetwarzać dane poprzez syntezę wirtualnego widoku zgodnego z pozycją użytkownika w scenie i wysyłać wyświetlany widok (łącze C1). Może też kodować i wysyłać wszystkie bądź wybrane obrazy bez przetwarzania, co powoduje, że terminal użytkownika lokalnie dokonuje syntezy wirtualnego widoku (łącze C2). W zależności od mocy obliczeniowej serwera brzegowego oraz wymaganych przez terminale użytkowników prędkości bitowych pojedynczy serwer może jednocześnie obsługiwać pewną liczbę użytkowników [Sta+18d][Grz+19a].

Użytkownicy systemu mogą posługiwać się różnymi typami terminali odbiorczych. Urządzenia te potrafią wyświetlać obraz na ekranie (np. monitora, urządzenia mobilnego) lub korzystać z urządzenia pozwalającego na uzyskanie efektu trójwymiarowego takiego jak: monitor autostereoskopowy [Ure+11], monitor stereoskopowy (migawkowy, polaryzacyjny) [Gen13] czy coraz popularniejszy wyświetlacz nagłowny (ang. HMD — Head-Mounted Display) [Shi02].

2.3 Kompresja

Techniki kompresji można sklasyfikować np. z uwagi na typ kodowanych danych:

- pojedynczego widoku (np. H.265/MPEG-H High Efficiency Video Coding [Sul+12]),
- wizji wielowidokowej (np. MV-HEVC [Han+15]),
- wizji trójwymiarowej (np. 3D-HEVC [Tec+16]),
- reprezentacji wizji wszechkierunkowej (np. MPEG Immersive Video MIV [Boy+21]),
- chmury punktów (np. Video-Based Point Cloud Compression V-PCC [Gra+20]).

Powszechnie stosowane techniki kompresji pojedynczego widoku zazwyczaj korzystają ze schematu kodowania hybrydowego z kompensacją ruchu. Normy dekodowania strumienia wizyjnego zostały opracowane m.in. przez grupy robocze ISO/IEC MPEG, np. MPEG-2 Video/H.262 [Sik97], MPEG-4 Advanced Video Coding/H.264 [Wie+03], MPEG-H High Efficiency Video Coding/H.265 [Sul+12] i MPEG-I Part 3 Versatile Video Coding/H.266 [Bro+21]. Popularne są również techniki kompresji opracowane przez konsorcja firm, np. VP9 [Muk+13] lub AV1 [Che+18].

Dla każdej z wymienionych norm zazwyczaj istnieją przykładowe implementacje kodera, które pozwalają np. na szybkie i wydajne uruchamianie na wybranych platformach sprzętowych [wX265][wX264][Wie+21]. Jest to możliwe poprzez m.in. wykorzystanie instrukcji wektorowych, takich jak: SIMD (ang. Single Instruction Multiple Data) lub MMX (ang. MultiMedia eXtensions), zrównoleglanie obliczeń, wykorzystanie szybkich algorytmów wyboru trybu kodowania [Hu+14][She+13]. Takie kodery posiadają przygotowane zestawy zdefiniowanych ustawień (ang. preset), które dodatkowo pozwalają zmniejszyć lub wydłużyć czas kodowania, co wpływa na współczynnik kompresji i uzyskiwaną jakość [Wie+21].

W literaturze można również znaleźć prace na temat rozszerzeń wspomnianych norm kompresji pojedynczego widoku. Pozwalają one np. na kompresję wielu widoków [Han+15][VWS11] lub kompresję wizji trójwymiarowej [Tec+16].

W obszarze zagadnień związanych z niniejszą pracą doktorską należy wymienić również technikę kompresji wizji wszechogarniającej MPEG Immersive Video (MIV)

[48220][Boy+21]. Dotyczy ona wykorzystania dowolnej techniki kompresji pojedynczego widoku do przesłania wizji wszechkierunkowej sceny. Koder MIV przetwarza scenę i przygotowuje metadane oraz atlasy dla dekodera MIV. Atlasy są obrazami, które zawierają kompozycję fragmentów reprezentacji sceny trójwymiarowej, zebranych z różnych widoków. Kompozycji tej dokonuje się aby uniknąć wielokrotnego powielania tych samych fragmentów sceny w różnych widokach, co umożliwia zmniejszenie nadmiarowości reprezentacji trójwymiarowej. W przypadku granicznym, atlas może zawierać pełnię informacji związanej z danym widokiem, jednak informacja ta nie jest wówczas powielana w innych atlasach. Atlasy poddaje się kompresji z pomocą dowolnej techniki pojedynczego widoku. Technika ta może działać w dwóch konfiguracjach. Domyślnie przesyłane są dwa roadzje altasów. Pierwsze z nich to obrazy z kamer, drugi stanowią odpowiadające im mapy głębi. Druga z konfiguracji techniki wizji wszechogarniającej MIV to tryb GA (ang. Geometry Absent), w której to wszystkie przesyłane atlasy zawierają obrazy z kamer, a głębia jest estymowana w odbiorniku. Sprawia to, że znaczna część przetwarzania danych jest przenoszona z kodera do dekodera, przez co mechanizm dekodowania jest bardziej rozbudowany i czasochłonny. Proces przetwarzania wykonywany w dekoderze działającym w trybie GA obejmuje cztery główne kroki:

- dekodowanie metadanych,
- dekodowanie wizji,
- estymację głębi,
- syntezę widoku wirtualnego.

Metadane zawierają parametry kamer, inne krytyczne informacje o widokach oraz (w trybie GA) parametry, które są przydatne na etapie estymacji map głębi i pozwalają na zmniejszenie złożoności obliczeniowej [Gar+22].

Do odrębnych technik kompresji należą techniki kompresji chmury punktów [Gra+20][Bui+21], ale niniejsza praca doktorska nie wykorzystuje reprezentacji sceny w takim formacie. W tej rozprawie wykorzystano reprezentację sceny za pomocą wielu widoków z kamer oraz odpowiadających im map głębi (MVD, ang. Multiview Video plus Depth).

W systemach swobodnego punktu widzenia techniki kompresji stratnej wykorzystuje się w łączach oznaczonych na rysunku 2.1 przez litery A, B oraz C. Ze

względu na zróżnicowane wymagania łączy oraz specyfikację zastosowań często wymaga się wykorzystania innych technik kompresji bądź co najmniej różnych trybów i konfiguracji tych samych technik.

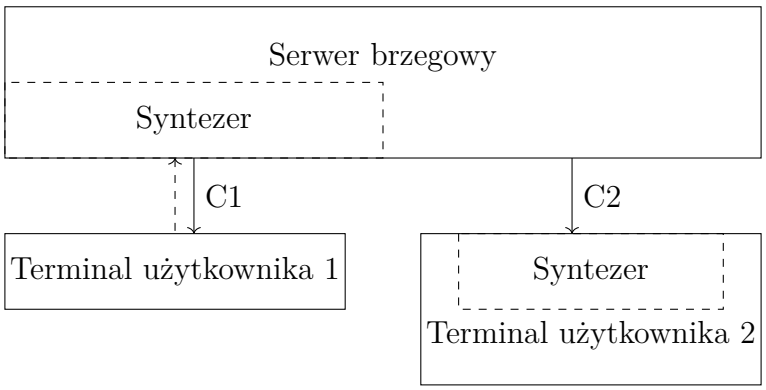
A jest łączem przesyłającym jedynie obraz z poszczególnych kamer systemu wielokamerowego. Oczywistym rozwiązaniem jest zatem niezależna kompresja obrazów z kamer systemu z wykorzystaniem techniki kompresji pojedynczego widoku. Takie rozwiązanie jest korzystne i praktyczne, gdyż często nie jest możliwe wyprowadzenie danych z kamery bez kompresji, a większość kamer ma wbudowany koder stratny. Mogą zostać wykorzystane techniki kompresji pojedynczego widoku np. AVC, HEVC, VVC, VP9, AV1. System akwizycji uzupełnia informacje o kamerach i parametrach systemu. Widoki przesyłane są do serwera reprezentacji.

Wpływ kompresji stratnej w łączu A na jakość całego systemu przebadano w rozdziale 5. Zakodowane dane zostały wykorzystane do estymacji map głębi oraz do syntezy wirtualnych widoków.

B jest łączem, które służy do przesyłania przestrzennej reprezentacji zarejestrowanej sceny. W niniejszej rozprawie wykorzystano reprezentację sceny za pomocą widoków oraz odpowiadających im map głębi (MVD, ang. Multiview Video plus Depth). Technika kompresji zastosowana dla takiej reprezentacji może być wspólna dla wszystkich widoków oraz map głębi z wykorzystaniem np. rozszerzeń wielowidokowych, rozszerzeń trójwymiarowych koderów (np. MV-HEVC, 3D-HEVC) [Han+15][wHEVC] lub koderów dedykowanych dla reprezentacji wszechkierunkowej, takich jak MPEG Immersive Video (MIV) [48220]. Alternatywnym rozwiązaniem może być niezależna kompresja każdego z obrazów z kamer oraz map głębi. Dla innych rodzajów reprezentacji również istnieją dedykowane techniki kompresji takie jak np. kodery dla chmur punktów [Gra+20][Bui+21], ale niniejsza prac skupia się na reprezentacji MVD, która jest powszechnie wykorzystywana dla danych rzeczywistych.

C jest łączem, w którym możliwe jest wiele schematów transmisji oraz syntezy wirtualnego widoku. Dwa podstawowe schematy łącza C zostały przedstawione na rysunku 2.2. W pierwszym z nich (łącze **C1**) synteza wirtualnego widoku realizowana jest przez serwer brzegowy. Przesyłane przez łącze C1 dane to syntezowany widok lub widoki wirtualne. W takim łączu mogą zostać zastosowane techniki kompresji pojedynczego widoku (przedstawione dla łącza A). Z uwagi na opóźnienie w swobodnej nawigacji w wirtualnej scenie, techniki te powinny zapewniać mały maksymalny czas

opóźnienia pomiędzy wprowadzeniem ramki na łącze (początek kompresji) a zdekodowaniem, czyli krótki czas dekodowania ramek względem kodowania (ang. Low Delay) [OFI12][Grz+19a][OFI12]. W drugim schemacie (łącze **C2** na rysunku 2.2) synteza wirtualnego widoku realizowana jest przez terminal użytkownika, a serwer brzegowy przesyła reprezentacje wizji wielowidokowej, wizji trójwymiarowej lub reprezentacji wizji wszechkierunkowej sceny. Kodery, jakie mogą zostać wykorzystane, to rozszerzenia wielowidokowe koderów (np. MV-HEVC, 3D-HEVC) [Han+15][wHEVC] lub kodery dedykowane dla reprezentacji wszechkierunkowej, takie jak MPEG Immersive Video (MIV) [48220].

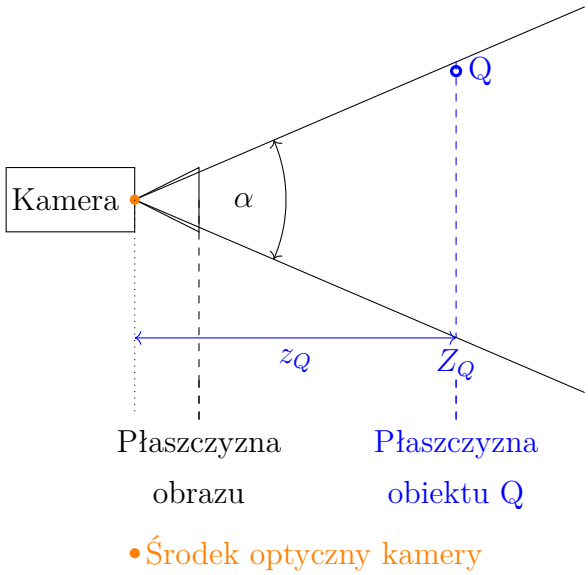


Rysunek 2.2: Synteza wirtualnego widoku i dwa schematy realizacji łącza C

2.4 Mapy głębi

Wartość głębi reprezentuje informację o odległości każdego z punktów do kamery w pewnym widoku. Dane te mogą zostać wyznaczone metodami grafiki komputerowej (dla danych sztucznych), zarejestrowane (np. przez kamerę głębi) lub wyliczone na podstawie analizy innych widoków za pomocą algorytmów estymacji głębi (np. dla danych rzeczywistych). Mapy głębi danych sztucznych są najczęściej określane sformułowaniem mapy głębi odniesienia (ang. ground-truth). Są to wyznaczone wartości głębi, które powinny zawierać poprawną wartość odległości do obiektów. W przypadku rzeczywistych sekwencji akwizycja map głębi odniesienia jest teoretycznie możliwa np. za pomocą kamer głębi. Takie rozwiązanie sprawia jednak wiele trudności z uwagi na m.in. możliwość wzajemnego zakłócania się kamer głębi, trudności w rejestrowaniu scen zewnętrznych, małą rozdzielczość, problematyczną

Rysunek 2.3: Przykładowe widoki z kamer (po lewej) oraz wizualizacje mapy głębi w skali szarości (po prawej stronie)



Rysunek 2.4: Rzut z góry — głębia (z_Q) punktowego obiektu Q kamery perspektywicznej

synchronizację z kamerami systemu wielokamerowego czy odbicia promieniowania podczerwonego od powierzchni [Pla+16][Kur19][Mie18].

W literaturze można znaleźć wiele rozwiązań hybrydowych wykorzystujących np. sensory głębi (sprzedawane wraz z konsolami do gier), które bazują na wstępnym modelowaniu sceny i jej tła, lub skanowaniu laserowym [Sch+17][Xia+13][Son+14][Kur19]. Wykorzystanie takich sensorów jest zazwyczaj problematyczne z wyżej wymienionych powodów dotyczących kamer głębi. Wiele z takich sensorów nie ma możliwości zarejestrowania głębi z wymaganą liczbą ramek na sekundę, podobnie jak w przypadku stosowania technologii skanowania laserowego. Mimo tych problemów, trwają prace nad rozwiązaniami, które pozwalają na budowę prostych wielokamerowych systemów wykorzystujących hybrydowe systemy wielokamerowe [Jia+18][Ber+20].

Estymacja map głębi oraz ich reprezentacja za pomocą widoków i towarzyszących im map głębi (MVD, ang. Multiview Video plus Depth) jest powszechnie wykorzystywana w systemach swobodnego punktu widzenia zarówno dla danych rzeczywistych, jak i sztucznych. Większość parametrów estymowanej mapy głębi (takich jak rozmiar ramki, pole widzenia, prędkość ramkowa) jest zgodna z obrazem z kamery. Na rysunku 2.3 przedstawiono obrazy z kamer oraz wizualizację mapy głębi w skali szarości. Pierwsze dwie sekwencje to dane sztuczne oraz mapy wyznaczone

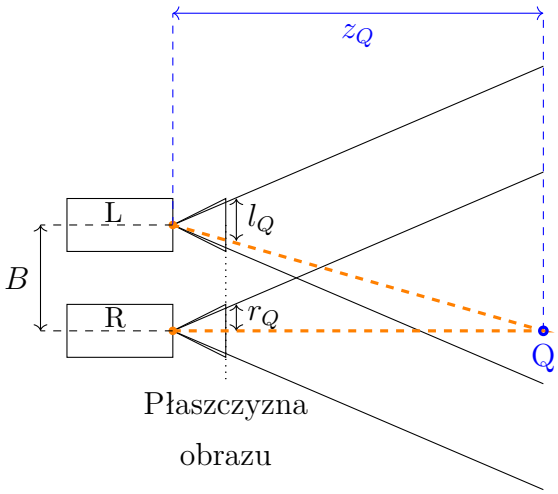
na podstawie modeli komputerowych, kolejne dwie sekwencje to dane rzeczywiste i mapy głębi wyliczone algorytmicznie za pomocą algorytmów estymacji map głębi.

Definicja mapy głębi może różnić się np. dla danych zarejestrowanych przez kamery głębi albo dla kamer półsferycznych czy sferycznych (rodzaje kamer zostały przedstawione w rozdziale 3.1). Najczęściej dla kamery perspektywicznej stosuje się model kamery otworkowej, czyli wyidealizowany model, w którym kamera jest sprowadzana do punktu nazywanego środkiem optycznym kamery [Dom10][Ike14]. Głębia punktowego obiektu Q, czyli z_Q została zaprezentowana na rysunku 2.4 w rzucie dwuwymiarowym z góry. Wartości głębi reprezentują z_Q będące odległością środka optycznego kamery od płaszczyzny Z_Q, która jest równoległą płaszczyzną do płaszczyzny przetwornika, gdzie znajduje się punktowy obiekt Q.

2.5 Estymacja głębi

Mapy głębi dla danych rzeczywistych typowo są estymowane na podstawie widoków uzyskanych przez system akwizycji. Każdy z zarejestrowanych przez kamery obiektów może zostać umiejscowiony w przestrzeni trójwymiarowej, jeśli został zarejestrowany również przez inną kamerę. Algorytmy estymacji map głębi można sklasyfikować ze względu na liczbę widoków, które są wykorzystywane do estymacji. W najprostszym ujęciu mapy głębi są wyznaczane niezależnie dla każdej pary kamer.

Porównanie obrazów polega na znalezieniu dla każdego punktu odpowiadającego mu punktu w widoku drugiej kamery. Obszar poszukiwań wynika z położenia kamer względem siebie, a zakres wszystkich możliwych położeń punktu w drugim widoku tworzy linię nazywaną epipolarną [KWZ99]. Linia ta może biec pod dowolnym kątem, a dla liniowego ustawienia kamer (wyjaśnienie rozdział 3.1) linie epipolarne pokrywają się z liniami (wierszami) obrazu [Dom+14b][CS09][Str+05]. Takie ustawienie zostało zaprezentowane w rzucie z góry na rysunku 2.5.



z_Q — głębia
B — odległość pomiędzy środkami optycznymi kamer (bazowa)
l_Q — współrzędne obiektu Q w płaszczyźnie obrazu kamery lewej (L)
r_Q — współrzędne obiektu Q w płaszczyźnie obrazu kamery prawej (R)
• — środek optyczny kamery

Rysunek 2.5: Rzut punktowego obiekt Q o głębi z_Q na płaszczyznę obrazu lewego (współrzędne l_Q) i prawego (współrzędne r_Q)

Znając różnicę położeń tego samego obiektu punktowego w dwóch widokach $r_Q - l_Q = d_Q$, czyli rozbieżność (ang. disparity), możliwe jest wyznaczenie położenia w przestrzeni trójwymiarowej. W przypadku dwóch kamer ustawionych liniowo i oddalonych od siebie o B (odległość bazowa, ang. baseline), głębię z_Q obiektu Q można obliczyć za pomocą wzoru:

$$z_Q = f \cdot B \cdot \frac{1}{r_Q - l_Q} = \frac{f \cdot B}{d_Q}, \tag{2.1}$$

gdzie f jest długością ogniskową obiektywu [HZ04][MMW11][Sta14]. Wynikiem takiego obliczania map głębi jest wiele niezależnych przestrzennie wyników, które nie zapewniają spójności przestrzennej wyznaczonych danych [Zha+19][Guo+19] [Dom+15d][Mie18].

Zaawansowane algorytmy wykorzystują te same zależności, ale pozwalają na jednoczesną estymację głębi dla wszystkich zarejestrowanych widoków z kamer. Powstała informacja jest spójna międzywidokowo dzięki wykorzystaniu wspólnej minimalizacji błędów. Algorytmy tego typu można podzielić na klasyczne metody oraz metody wykorzystujące uczenie maszynowe.

Klasyczne algorytmy pozwalające na wspólną wielowidokową estymację głębi działają najczęściej na zasadzie minimalizacji funkcji celu [BK04][Sta14], a powszechnie używany do tego jest algorytm cięcia grafu (ang. Graph Cut) [KZ04][KZ02]. Wspomniane wyżej funkcje błędu i spójności wykorzystywane są do budowy grafu, który reprezentuje rozwiązywany problem estymacji głębi. Graf ten jest dzielony przez algorytm, co odpowiada przypisywaniu punktom obrazów z kamer wartości głębi.

Algorytmy wykorzystujące uczenie maszynowe pozwalają zazwyczaj na wspólną estymację głębi dla wszystkich widoków i najczęściej działają w trzech krokach: ekstrakcji cech, konstruktora funkcji celu oraz sieci optymalizującej funkcje celu [Sch+17]. Wiele algorytmów ze względu na duże zapotrzebowanie na pamięć nie generuje map głębi w pełnej rozdzielczości oraz nie umożliwia estymacja dla bardzo wielu widoków [Yao+18][Wei+21]. Podejście wykorzystujące algorytmy uczenia maszynowego jest bardziej obiecujące dla syntezy wirtualnego widoku z pominięciem estymacji map głębi [Wan+21][Mo+23]. Metody estymacji głębi wykorzystujące uczenie maszynowe nie są wykorzystywane w niniejszej pracy, gdyż w zakładanym systemie swobodnej nawigacji są mniej efektywne z uwagi na jakość generowanego wyniku, czas działania i potrzebną moc obliczeniową [Rav+22]. Poza tym, w niniejszej rozprawie założono, że architektura systemu zawiera serwer reprezentacji, co uniemożliwia wykorzystanie metod z pominięciem tego kroku.

Jest wiele prac, które dotyczą rozwoju klasycznych wielowidokowych metod estymacji głębi. Do najważniejszych kierunków należą dzieła na temat:

- spójności czasowej [SDW15][KPP13],
- segmentacji [Li+18][Shi+16],
- zrównoleglania obliczeń [Pen+15][Mie18].

Metody spełniające powyższe wytyczne pozwalają na uzyskanie wysokiej jakości map głębi, spójnych międzywidokowo i czasowo, odpowiednich dla systemów swobodnego punktu widzenia [Mie18].

Można znaleźć strony, które zawierają bazę sekwencji testowych służących do porównania metod estymacji głębi między sobą na nieznanych danych sztucznych lub rzeczywistych zarejestrowanych eksperymentalnymi systemami wielokamerowymi [wMid][wKITTI][wMidm][Sch+17].

Niniejsza praca nie obejmuje rozwoju metod estymacji map głębi. W jej ramach wykorzystuje się istniejące algorytmy, w szczególności rozwijane przez grupy robocze ISO/IEC MPEG, oprogramowanie DERS (ang. Depth Estimation Reference Software) oraz IVDE (ang. Immersive Video Depth Estimation)[Sta+13][Mie18][Rog+19].

2.6 Synteza widoków wirtualnych

Jeden z możliwych sposobów klasyfikacji metod syntezy wirtualnego widoku to podział na rodzaje wykorzystywanej reprezentacji przestrzennej sceny. Wśród możliwych rozwiązań można wymienić metody syntezy z reprezentacji sceny w postaci:

- chmury punktów [Ceu+18][Wu+17],
- przestrzeni promieni [Tan05][Kim+13],
- modeli trójwymiarowych obiektów lub segmentów [Dub+18][Smo+05],
- reprezentacji MVD (ang. Multiview Video plus Depth) [Dzi+19][SD20],
- sieci neuronowej (NeRF, ang. neural radiance field)[Mil+20][Wu+23].

Metody wykorzystujące modelowanie sceny za pomocą sieci (np. NeRF ang. Neural Radiance Field) pozwalają na syntezę wirtualnego widoku z pomięciem kroku reprezentacji przestrzennej [Mil+20][Irs+23][Mo+23][Wan+21]. Często pozwalają one również na generowanie mapy głębi, ale nie uczestniczy ona bezpośrednio w procesie syntezy wirtualnego widoku. Zazwyczaj synteza wirtualnego widoku dla nowej sceny wymaga ponownego uczenia sieci, lecz można znaleźć metody, w których pomija się ten krok np. dla scen zewnętrznych [Irs+23]. Tego typu rozwiązania nie są wykorzystywane w pracy z uwagi na dużą złożoność obliczeniową oraz częstą potrzebę uczenia sieci dla nowej sekwencji. Poza tym założono architekturę systemu, w której występuje przestrzenna reprezentacja sceny realizowana przez serwer reprezentacji.

Jednym z rodzajów syntezerów wirtualnego widoku z reprezentacji MVD są syntezery dostosowane jedynie do liniowego ustawienia kamer [Aki+15][HL15] lub takie, które wykorzystują ograniczoną liczbę widoków rzeczywistych [LHL13][Jin+16]. Te rozwiązania są mało praktyczne w badanych systemach ze względu na dostępność wielu widoków o różnym ustawieniu.

W niniejszej pracy wykorzystano metody syntezy wirtualnych widoków przystosowane do przetwarzania wielu widoków oraz dla dowolnego ustawienia kamer, tj.

[SD20][Dzi+16c][Dzi+19]. Najważniejsze etapy działania syntezerów to rzutowanie punktów widoków rzeczywistych do widoku wirtualnego oraz wypełnianie obszarów, które nie zostały przerzutowane [Dzi18].

Algorytmy syntezy wirtualnych widoków oraz badania nad ich rozwojem nie znajdują się w zakresie niniejszej rozprawy. Zastosowano w niej istniejące algorytmy syntezy widoków wirtualnych, w szczególności rozwijane przez grupy robocze ISO/IEC MPEG i wykorzystujące wieloobrazową reprezentację sceny i towarzyszące im mapy głębi (MVD, ang. Multiview Video plus Depth).

2.7 Rozmieszczenie wielu kamer w scenie

Rozmieszczenie kamer w scenie powinno zapewniać postrzeganą wysoką jakość usługi swobodnej nawigacji oraz prostą budowę systemu. Większa liczba kamer w systemie zwiększa ilość danych, które muszą zostać przetworzone oraz podnosi koszt budowy i obsługi systemu. Optymalizacja rozmieszczenia kamer powinna zatem pozwolić na użycie możliwie najmniejszej ich liczby w takim rozmieszczeniu, które zapewni postrzeganą wysoką jakość usługi swobodnej nawigacji.

Problem optymalizacji ustawienia kamer rozpatrywano w literaturze dla różnych zagadnień związanych z grafiką komputerową i śledzeniem obiektów [RK17][CD08] [OM02][QL15][GPS07]. Dokładniej, różne ustawienia kamer zostały wzięte pod uwagę dla systemów CAVE (ang. Cave Automatic Virtual Environment) czyli projekcji obrazu w pomieszczeniu wokół widza [RK17], śledzenia obiektów [CD08] oraz precyzyjnej rekonstrukcji powierzchni obiektów [OM02][GPS07]. Niestety, dla rozważanych technik wymagana jest większa ilość informacji wejściowej (np. geometria, kształt, ruch, liczba obiektów w scenie), niż jest to dostępne w przypadku systemów swobodnej nawigacji.

Wiele pozycji dotyczy problemu rozmieszczenia jedynie dwóch kamer. Wnioski wyciągnięte przez autorów [YNT10][NM19] potwierdzają wiele znanych i przewidywalnych czynników, takich jak: indywidualne predyspozycje użytkownika, większą atrakcyjność ustawienia łukowego względem liniowego, pozytywny wpływ liczby dostępnych widoków czy negatywny wpływ obciążenia sieci i opóźnień w sterowaniu. Prace [Dim+19][WLH04] podkreślają duży wpływ pozycji widoku wirtualnego w systemach związanych ze zdalnym sterowaniem czy robotyką. Propozycje optymalizacji ustawienia kamer podjęte w publikacjach [Saf+13][FSL17] rozpatrują spadek jakości systemu dla różnych obrotów dwóch kamer. Rozważania te są istotne, szczególnie

kiedy znany jest kształt i rozmieszczenie obiektów w scenie. Z tego względu mają one bardzo ograniczone praktyczne zastosowanie w systemach swobodnej nawigacji. Istnieją prace dotyczące błędu estymacji głębi, metod jej szacowania i ich złożoności obliczeniowej, optymalizacji odległości pomiędzy kamerami w stereoparze oraz określania błędu w przestrzeni trójwymiarowej (również z uwagi na niedokładne ustawienie kamer lub błędną kalibrację)[CS09][Foo+13][Bn15][ZB11]. Wiele publikacji zawiera również wyniki prac eksperymentalnych, podczas których korzystając np. z laserowego pomiaru odległości [San+17][Che+21b] lub z przedmiotów o znanym rozmiarze w polu widzenia kamer [Gra+13][Bn15][Kuk+19], weryfikuje się wyprowadzone teoretyczne zależności. Wszystkie te prace mają ograniczone zastosowanie w systemach swobodnej nawigacji, gdyż dotyczą systemów składających się jedynie z dwóch kamer.

W innych publikacjach rozpatruje się problem rozmieszczenia kamer dla systemów dozoru wizyjnego. Pełne pokrycie pola widzenia przy możliwie niewielkiej liczbie kamer, obserwacja obszarów krytycznych z kilku pozycji, wykorzystywanie możliwości kamer dozorowych takich jak ruchome głowice czy powiększenie optyczne, jest ważnym i praktycznym zagadnieniem badawczym. Problemy podjęte w takich pracach [Liu+14][KSG19] mają również ograniczone zastosowanie w przypadku systemów swobodnej nawigacji, gdyż nie uwzględniają procesu rekonstrukcji informacji przestrzennej, niezbędnego dla systemów swobodnej nawigacji.

Rozmieszczenie kamer w scenie zostało również zbadane dla systemów, które m.in. śledzą ruch i położenie obiektów w scenie [RK17][RK15], dotyczących rekonstrukcji powierzchni [QL15],lub starających się modelować błędną wartość głębi spowodowaną niedokładną synchronizacją kamer [DSO17]. Problem dokładności wyznaczania pozycji punktów w przestrzeni trójwymiarowej jest również badany w dla sensorów głębi [LG18], w systemach hybrydowych [Kur19] oraz dla urządzeń wykorzystywanych przez konsole do gier [Yan+15][Che+21a].

W literaturze nie ma prac, które zajmują się rozmieszczeniem kamer w systemach wielokamerowych. Wpływ zmiany ustawienia tych kamer na proces estymacji map głębi oraz syntezy wirtualnego widoku, z uwagi na uzyskiwaną postrzeganą jakość, został przedstawiony w rozdziale 4.

2.8 Synchronizacja

Synchronizacja systemu wielokamerowego pozwala na koordynację momentów wyzwalania migawki kamery i akwizycję ramek w tej samej chwili. Dokładność synchronicznej akwizycji sekwencji wielowidokowej wpływa na spójność zarejestrowanych danych i jakość systemu. Brak synchronizacji i późniejsze algorytmiczne jej wyznaczenie poprzez znalezienie odpowiadających w czasie ramek [Hua+19][WML14], jest również wykorzystywane w pewnych obszarach badawczych. Nie wymaga to żadnych dedykowanych urządzeń, a jedynie dodatkowego kroku przetwarzania danych. Rezultaty mogą wystarczać do uzyskania wysokiej jakości rekonstrukcji przestrzennej, szczególnie jeśli zarejestrowana sekwencja nie zawiera szybko przemieszczających się obiektów. W skrajnym przypadku (statyczna scena) wszystkie kamery systemu mogą zostać zastąpione jednym ruchomym aparatem, a problem synchronizacji zostaje wyeliminowany. W innym przypadku i dynamicznej (ruchomej) scenie, synchronizacja kamer jest niezbędna w celu osiągnięcia postrzeganej wysokiej jakości [Dim18][Kur19].

Szacowanie dokładności, z jaką system synchronizacji powinien wyzwalać wszystkie kamery, związane jest z ich rodzajem oraz dynamiką ruchu w scenie, czyli prędkością obiektów i ich odległością od kamer systemu [Dim18][Kur19]. Jednym z możliwych rozwiązań, zapewniających synchroniczność wyzwalania i akwizycji ramek w systemie, jest synchronizacja czasu w kamerach. Wyzwalanie może wówczas odbywać się w tych samych momentach. Wiadomo, że powszechnie wykorzystywane protokoły synchronizacji, takie jak NTP (ang. Network Time Protocol) [Mar+10] czy SNTP (ang. Simple Network Time Protocol) [Mil06b], zapewniają milisekundową dokładność synchronizacji czasu [Mil06a]. Nie jest to wystarczająca dokładność, a synchronizacja kamer musi wykorzystywać mniej powszechne precyzyjne protokoły synchronizacji, takie jak PTP (ang. Precision Time Protocol) [wPTP20][SS15] czy White Rabbit [Lip+18], które mogą zapewnić dokładność synchronizacji czasu poniżej mikrosekundy [Hor+17]. Rozwiązania te wymagają urządzeń (m.in. sieciowych, kamer) wspierających tę normę, ponieważ konieczne jest znakowanie czasu otrzymywanych i wysyłanych ramek synchronizacyjnych przez sieć komputerową. Innym możliwym rozwiązaniem jest budowa dedykowanego systemu synchronizacji wyzwalania ramek lub nawet linii obrazu kamer z migawką szczelinową (ang. rolling shutter) [DSO17][Jia+18]. W jednym i drugim przypadku synchronizacja wymaga nie tylko rozwiązań na poziomie oprogramowania, ale również dedykowanych

urządzeń (kart sieciowych ze wsparciem dla np. PTP lub dedykowanych urządzeń synchronizacyjnych).

Błąd synchronizacji przekłada się na uzyskiwaną jakość systemu i jest obserwowany razem z innymi problemami takimi jak np. niedokładność estymacji parametrów wewnętrznych i zewnętrznych kamery, efekt rozmycia krawędzi szybko przemieszczających się obiektów [Dzi18] i brak możliwości wyznaczenia poprawnej mapy głębi z uwagi na przesunięcia przemieszczających się obiektów. Błędy synchronizacji przekładają się również na błąd wyznaczania parametrów zewnętrznych w przypadku metod kalibracji wykorzystujących analizę sekwencji z ruchomym obiektem. Sposobem rozwiązania tego problemu są alternatywne metody wyznaczania parametrów zewnętrznych kamer systemu (np. analiza punktów charakterystycznych lub dla sceny bez poruszających się obiektów) bądź zapewnienie synchronizacji z dokładnością, która pozwala na pominięcie tego błędu [Liu+09][GSY17][Kur19][Dom+14a].

3 Metodologia badań

3.1 Sekwencje testowe

W rozprawie bardzo ważną rolę odgrywają eksperymentalne walidacje wyników rozwiązań. Eksperymenty służą ponadto do uchwycenia różnych zjawisk w złożonych systemach. Przyjęta w dziedzinie przetwarzania i kompresji wizji wielowidokowej metodologia badań obejmuje zbiory sekwencji testowych. Takie sekwencje do celów badawczych i normalizacyjnych udostępniają grupy MPEG działające w ramach ISO/IEC JTC1/SC29 [Ter+20][Teh+18]. Zbiór sekwencji testowych wykorzystywanych podczas pracy nad rozprawą doktorską został przedstawiony w tabeli 3.1.

Tablica 3.1: Wykorzystane sekwencje testowe

Nazwa sekwencji	Rozmiar ramki	Liczba ramek	Liczba widoków	Treść	Ustawienie kamer	Rzutowanie obrazu
BBB Butterfly Arc [Kov+15]	1280 × 768	120	91	Sztuczna	Łukowe	Perspektywiczne
BBB Butterfly Linear [Kov+15]	1280 × 768	120	91	Sztuczna	Liniowe	Perspektywiczne
BBB Flowers Arc [Kov+15]	1280 × 768	120	91	Sztuczna	Łukowe	Perspektywiczne
BBB Flowers Linear [Kov+15]	1280 × 768	120	91	Sztuczna	Liniowe	Perspektywiczne
BBB Rabbit Arc [Kov+15]	1280 × 768	120	91	Sztuczna	Łukowe	Perspektywiczne
BBB Rabbit Linear [Kov+15]	1280 × 768	120	91	Sztuczna	Liniowe	Perspektywiczne
Bee [SS13]	1920 × 1088	1	182	Sztuczna	Liniowe	Perspektywiczne
Champagne [TFF08]	1280 × 960	500	80	Rzeczywista	Liniowe	Perspektywiczne
Chess [00620]	2048 × 2048	300	10	Sztuczna	Inne	Półsferyczne
ChessPieces [00620]	2048 × 2048	300	10	Sztuczna	Inne	Półsferyczne
ClassroomVideo [Kro18]	4096 × 2048	120	15	Sztuczna	Inne	Sferyczne (Dookólne)
Dog [TFF08]	1280 × 960	300	80	Rzeczywista	Liniowe	Perspektywiczne
Fan [00620]	1920 × 1080	97	15	Sztuczna	Macierzowe	Perspektywiczne
Frog [00620]	1920 × 1080	300	13	Rzeczywista	Liniowe	Perspektywiczne
Group [00620]	1920 × 1080	99	21	Sztuczna	Inne	Perspektywiczne
Hijack [00620]	4096 × 2048	300	10	Sztuczna	Inne	Półsferyczne
Kitchen [00620]	1920 × 1080	97	25	Sztuczna	Macierzowe	Perspektywiczne
Mirror [RD20]	1920 × 1080	97	15	Sztuczna	Macierzowe	Perspektywiczne
Museum [DBT18]	2048 × 2048	300	24	Sztuczna	Inne	Półsferyczne
Painter [00620]	2048 × 1088	300	16	Sztuczna	Macierzowe	Perspektywiczne
Pantomime [TFF08]	1280 × 960	500	80	Rzeczywista	Liniowe	Perspektywiczne
(Poznań) Carpark [00620]	1920 × 1088	250	9	Rzeczywista	Liniowe	Perspektywiczne
(Poznań) Fencing [00620]	1920 × 1080	250	10	Rzeczywista	Łukowe	Perspektywiczne
(Poznań) Hall [00620]	1920 × 1088	500	9	Rzeczywista	Liniowe	Perspektywiczne
Poznań-People360 [Sta+18b]	4096 × 2048	1	3	Sztuczna	Inne	Sferyczne (Dookólne)
(Poznań) Street [00620]	1920 × 1088	250	9	Rzeczywista	Liniowe	Perspektywiczne
Poznań Volleyball [Dom+18a]	1920 × 1080	300	34	Rzeczywista	Inne	Perspektywiczne
San Miguel [Goo+14]	1920 × 1080	1	200	Sztuczna	Liniowe	Perspektywiczne

Składa się on ze zróżnicowanych rzeczywistych i sztucznych sekwencji o różnej liczbie kamer, ustawieniu, rozmiarze ramki. Sekwencje te są wykorzystywane w wielu pracach [Ceu+18][LZS18][Fuj+06][Zit+04] i w większości stanowią zbiór wykorzystywany od wielu lat przez grupy ISO/IEC JTC1/SC29 [Ter+20][Teh+18]. Zbiór ten zawiera także pozycje, które zostały stworzone przez zespół naukowców obecnego Instytutu Telekomunikacji Multimedialnej Politechniki Poznańskiej (Poznań Street, Poznań Volleyball, Poznań-People360, Poznań Hall, Poznań Fencing, Poznań Carpark). W przygotowaniu części z tych sekwencji (Poznań Volleyball, Poznań Fencing) autor brał udział, m.in. w opracowaniu scenariusza, konfiguracji sprzętu, weryfikacji nagrań, wyznaczeniu parametrów zewnętrznych oraz wstępnych map głębi i wirtualnych widoków. Przykładowe obrazy z kamer zostały przedstawione na rysunku 3.1.

- **Liniowe ustawienie** kamer (nazywane też ustawieniem kanonicznym [Str+05]) oznacza system wielokamerowy, w którym środki optyczne kamer znajdują się na jednej linii prostopadłej do osi optycznych, które są równoległe względem siebie (np. rysunek 4.1a). W takim ustawieniu osie optyczne są współpłaszczyznowe.

- **Ustawienie łukowe** jest modyfikacją ustawienia liniowego — osie optyczne są współpłaszczyznowe, ale przecinają się w jednym punkcie, a środki optyczne leżą na wycinku okręgu.

- **Ustawienie macierzowe** oznacza kamery ustawione w macierz np. 5×5 w taki sposób, że kamery tworzą wiele systemów liniowych — poziomo, pionowo, po przekątnej.

Typowa kamera najczęściej wykorzystuje **rzutowanie perspektywiczne** obrazu. **Rzutowanie półsferyczne** dotyczy kamery o bardzo szerokim kącie widzenia wynoszącym około 180 stopni. Kamera sferyczna lub dookólna (360) to kamera rejestrującą scenę we wszystkich kierunkach, czyli wykorzystująca **rzutowanie sferyczne**. Przedstawiony podział związany jest z typem zarejestrowanych obrazów i nie dotyczy precyzyjnie jego sposobu rzutowania i przechowywania zarejestrowanych danych. Sekwencje z kamerami półsferycznymi oraz sferycznymi, które zostały użyte w tej pracy, najczęściej korzystają z rzutowania walcowego równoodległościowego (ang. ERP, Equirectangular Projection) [Sta+18c].

(a) Champagne (b) Dog (c) Pantomime

(d) Bee (e) San Miguel

(f) ClassroomVideo (g) Poznań-People360

(h) Museum (i) Chess

Rysunek 3.1: Przykładowe obrazy z sekwencji testowych

3.2 Metryki oceny jakości sekwencji wizyjnych

Najbardziej wiarygodne badania, które dotyczą oceny jakości, związane są z testami subiektywnymi. W literaturze można znaleźć wiele metod subiektywnej oceny jakości dla różnych zastosowań, np. oceny jakości telewizji cyfrowej, detekcji obiektów czy siły zniekształceń kodowania stratnego [Pau+13][Biba][Bibc][Bibd][Bibb]. Jakość subiektywna określania jest przez współczynnik MOS (ang. Mean Opinion Score) w założonej skali, a testy mogą być przeprowadzone w różnych scenariuszach. Do najważniejszych należą [Biba][Bibc]:

- metoda oceny pojedynczego bodźca SS (ang Single Stimulus),
- metoda oceny bezwzględnej jakości wizji ACR (ang. Absolute Category),
- metoda ACR z ukrytym odniesieniem ACR-HR (ang. Absolute Category Rating-Hidden Reference),
- metoda oceny degradacji jakości DCR (ang. Degradation Category Rating),
- metoda porównawcza PC (ang. Pair Comparison).

W rozdziale 6 przeprowadzono testy subiektywne zgodnie ze scenariuszem ACR-HR, kierując się wytycznymi zawartymi w normach ITU-T Rec. P.910 [Bibc] oraz ITU-R Rec. BT500 [Biba].

W pracach naukowych jakość sekwencji wizyjnej, zgodnej z subiektywną oceną użytkowników, określa się liczbowo. Dąży się do tego, gdyż przeprowadzanie testów subiektywnych jest skomplikowanym procesem, angażującym wiele osób i wydłużającym proces oceny jakości. Wykorzystywanie metryk oceny jakości umożliwia przyspieszenie oceny procesu przetwarzania sekwencji wizyjnej poprzez możliwość obiektywnego porównania wyników. Dokonuje się tego najczęściej w celu określenia wielkości zniekształceń spowodowanych różnego rodzaju przetwarzaniem, takim jak np. kompresja stratna, transkodowanie, transmisja ze stratami pakietów.

Proces pomiaru może się odbywać z dostępem do sekwencji oryginalnej (niezniekształconej) bądź bez dostępu. Pierwsze z podejść jest często wykorzystywane w trakcie badań naukowych, drugie jest użyteczne np. we wdrożonych systemach transmisyjnych, gdzie dostęp do oryginalnych danych nie jest możliwy. Przesłankami do wyznaczenia liczbowej oceny jakości w systemach bez dostępu do sekwencji oryginalnej są typowe zniekształcenia silnego kodowania lub błędy transmisji, takie

jak np. efekt blokowy, rozmycie krawędzi, efekt dzwonienia, migotanie, zamrożenie wyświetlanego obrazu. Zastosowanie takich metryk ogranicza się do sytuacji, kiedy oczekujemy pewnego typu zniekształceń. W celu oceny jakości w pracy wykorzystano metryki z dostępem do danych oryginalnych [Dom10][Akr14]. Jednakże w systemach swobodnej nawigacji algorytmy estymacji map głębi oraz synteza widoków wirtualnych sprawiają, że w widoku wirtualnym mogą pojawić się zniekształcenia, które nie są spotykane w przypadku klasycznych systemów przetwarzania czy kodowania obrazu.

Najpopularniejszą metryką oceny jakości jest PSNR (ang. Peak Signal-to-Noise Ratio), stosowany powszechnie dla obrazów statycznych oraz sekwencji [Dom10]. Występują jednak rodzaje zniekształceń takie jak np. lokalne braki w syntezowanym obrazie, które powodują silne negatywne odczucia subiektywne nieodwzorowane przez wartość PSNR — w pracach naukowych można znaleźć inne ich rodzaje, jak np. SSIM (ang. Structural similarity) [Wan+04a], VQM (ang. Video Quality Metrics) [PW04], VIF (ang. Visual Information Fidelity) [SB06] czy VMAF (ang. Video Multimethod Assessment Fusion) [wNetflix]. Wszystkie one wykorzystują porównanie z oryginalnym obrazem i są często stosowane, a w literaturze można znaleźć również ich rozszerzenia i modyfikacje (np. VQM_VFD [PCB14] lub modyfikacje metryki VMAF uwzględniające modele czasoprzestrzenne [BLB19]). VMAF wykorzystuje sieci neuronowe i algorytmy uczenia, co sprawia, że zniekształcenia, na jakie jest wyczulona, zależą od zbiorów sekwencji wykorzystywanych w procesie trenowania i są najczęściej związane z pracą kodera oraz wyborem najlepszego trybu pracy. Podobnie metryka VQM wykorzystuje ekstrakcję cech, która może opierać się na różnych modelach, np. ogólnym, wideokonferencji, telewizyjnym. Ich modyfikacje, uwzględniające m.in. zawartość sceny, opóźnienia, obrazy dookólne, to np. MS-SSIM (ang. multiscale structural similarity) [WSB03], VQM_VFD (ang. video quality model for variable frame delay) [PCB14], WS-PSNR (ang. weighted-to-spherically-uniform peak signal-to-noise ratio) [SLY17], PSNR-HVS-M (ang. peak signal-to-noise ratio human visual system metrics) [Pon+07]).

Można zatem stwierdzić, że żadna z zaprezentowanych wyżej metryk nie jest dostosowana do szczególnych zniekształceń wynikających z syntezy wirtualnego obrazu. Obraz syntezowany posiada specyficzne wady, które wynikają z różnej charakterystyki barwnej lub punktowych przesunięć obiektów związanych z zaokrągleniami podczas obliczeń [Dzi18]. Ze względów wspomnianych powyżej, w rozprawie zdecydowano się

wykorzystać metrykę IV-PSNR (ang. Immersive Video Peak Signal-to-Noise Ratio). Metryka ta jest dostosowana do obrazów syntezowanych i skorelowana z odczuciami subiektywnymi, wynikającymi z wyżej przedstawionych zniekształceń. Jest ona używana przez ekspertów grupy ISO/IEC MPEG podczas prac nad nową techniką kompresji wizji wszechkierunkowej MIV. Do najważniejszych jej cech należy brak wrażliwości na niewielkie przesunięcie obiektów poprzez uwzględnianie potencjalnego przesunięcia punktu odniesienia (ang. Corresponding Pixel Shift) oraz różnicy charakterystyki barwnej kamer (ang. Global Component Difference) [49520][Dzi+22]. Badania dotyczące opracowania i rozwoju metryk oceny jakości subiektywnej nie znajdują się w głównym nurcie badań objętych rozprawą, jednakże jej autor jest współtwórcą metryki IV-PSNR, co stanowi dodatkowe osiągnięcie.

W rozdziale 4 oraz 5 wyniki zostały porównane z wykorzystaniem metryki PSNR. W rozdziale 5 wykorzystano dodatkowo metrykę IV-PSNR.

3.3 Definicja metryk PSNR oraz IV-PSNR

W niniejszej pracy wykorzystano metryki PSNR oraz IV-PSNR. Pierwsza z nich jest powszechnie stosowana we wszystkich badaniach naukowych związanych z przetwarzaniem obrazu oraz wizji [Poy12][Dom10][Cyg02]. Można ją również sklasyfikować jako szybką do wyznaczenia i łatwą do implementacji. Wartość PSNR dowolnej składowej barwnej, oznaczonej jako c, wyznacza się następującym wzorem:

$$\mathrm{PSNR}_c = 10 \cdot \log_{10} \left(\frac{\left(2^b - 1\right)^2}{\mathrm{MSE}_c} \right), \tag{3.1}$$

gdzie b to liczba bitów reprezentacji wartości próbek, a MSE_c to średni błąd kwadratowy składowej barwnej c obrazu I względem obrazu odniesienia J, który jest obrazem bez zniekształceń:

$$\mathrm{MSE}_c = \frac{1}{W \cdot H} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \left(I_c^{x,y} - J_c^{x,y}\right)^2, \tag{3.2}$$

gdzie W oraz H to rozmiary ramki obrazu.

Metryka IV-PSNR jest modyfikacją metryki PSNR, mającą na celu uzyskanie większej zgodności z ocenami subiektywnymi jakości obrazów syntezowanych. Metryka IV-PSNR uwzględnienia potencjalną różnicę charakterystyki barwnej widoków (ang. Global Component Difference) oraz potencjalne przesunięcie punktów odniesienia

(ang. Corresponding Pixel Shift), co zostało wyjaśnione poniżej. Wyznaczanie metryki IV-PSNR dla każdego z komponentów jest podobne do wyznaczania metryki PSNR. Dla obrazów I i J oraz składowej barwnej c oblicza się ją za pomocą wzoru:

$$\text{IV-PSNR}_c^{I\to J} = 10 \cdot \log_{10}\left(\frac{\left(2^b-1\right)^2}{\text{IV-MSE}_c^{I\to J}}\right), \tag{3.3}$$

gdzie $IV\text{-}MSE_c$ to odpowiednik średniego kwadratowego błędu dla składowej barwnej c i metryki $IV\text{-}PSNR$. Jego wartość jest wyliczana z uwzględnieniem otoczenia odpowiadającego punktu, czyli uwzględniając potencjalne przesunięcia punktów odniesienia, zgodnie ze wzorem:

$$\text{IV-MSE}_c^{I\to J} = \frac{1}{W\cdot H}\sum_{y=0}^{H-1}\sum_{x=0}^{W-1} \min_{\substack{w\in[x-B,x+B]\\ h\in[y-B,y+B]}} \left(I_c^{x,y} - J_c^{w,h} + GCD_c^{I\to J}\right)^2, \tag{3.4}$$

gdzie B jest wielkością bloku przeszukiwania odpowiadającego punktu (najczęściej $B = 2$, co odpowiada blokowi 5×5). $GCD_c^{I->J}$ jest średnią różnicą charakterystyki barwnej widoków, wyznaczaną na podstawie średniej różnicy wartości próbek składowych barwnych, zgodnie z wzorem:

$$\text{GCD}_c^{I\to J} = \frac{1}{W\cdot H}\sum_{y=0}^{H-1}\sum_{x=0}^{W-1}\left(I_c^{x,y} - J_c^{x,y}\right), \tag{3.5}$$

Wartości metryki IV-PSNR składowych barwnych przestrzeni YC_bC_r są uśredniane ze współczynnikami zgodnie ze wzorem:

$$\text{IV-PSNR}_{YC_bC_r}^{I\to J} = \frac{\text{IV-PSNR}_Y^{I\to J}\cdot w_Y + \text{IV-PSNR}_{C_b}^{I\to J}\cdot w_{C_b} + \text{IV-PSNR}_{C_r}^{I\to J}\cdot w_{C_r}}{w_Y + w_{C_b} + w_{C_r}}. \tag{3.6}$$

Domyślnie wartość współczynników dla luminancji to $w_Y = 4$, a dla każdej z chrominancji to wartości $w_{Cb} = w_{Cr} = 1$. Przybliżają one liczbę punktów, która występuje dla często wykorzystywanego formatu próbkowania 4:2:0 [Dzi+22].

Ponieważ funkcja wartości błędu średniego kwadratowego $IV\text{-}MSE_c^{I\to J}$ nie jest symetryczna względem $IV\text{-}MSE_c^{J\to I}$, to również wyrażenie IV-PSNR ma tą cechę, czyli zwykle:

$$\text{IV-PSNR}^{I\to J} \neq \text{IV-PSNR}^{J\to I}. \tag{3.7}$$

W celu zapewnienia symetrii metryki IV-PSNR wybierana jest mniejsza z jej wartości wyliczanej dla obrazu I względem J oraz J względem I, zgodnie ze wzorem:

$$\text{IV-PSNR}(I,J) = \min\left(\text{IV-PSNR}^{I\to J}_{YUV}, \text{IV-PSNR}^{J\to I}_{YUV}\right). \tag{3.8}$$

W pracy przyjęto, że IV-PSNR jest obliczany w domyślny sposób, czyli dla wszystkich składowych przestrzeni barwnej YC_bC_r i łączony zgodnie z wzorem 3.6. Przyjęto również, że PSNR jest wyznaczany wyłącznie dla luminancji ($c = Y$), co jest powszechnie stosowaną praktyką w literaturze, np. [Ceu+18],[Dzi+16c],[HL15],[LZS18] i [Dzi18].

Przy powyższych opisach metryk PSNR i IV-PSNR, dla zwiększenia ich czytelności, pominięto kwestię uwzględnienia różnych metod rzutowania wykorzystywanych w badanych obrazach. Ich uwzględnienie jest konieczne w przypadkach, gdy dane rzutowanie przypisuje różną istotność różnym obszarom obrazu. Przykładowo, w rzutowaniu walcowym równoodległościowym (ang. ERP, Equirectangular Projection) [Sta+18c], punkty obrazu leżące na równiku mają znacznie większe znaczenie dla całkowitej jakości, niż punkty obrazu leżące na północnych i południowych równoleżnikach, gdyż w skrajnym przypadku odpowiadają one zaledwie pojedynczym punktom (biegunom) w scenie. Dlatego, zależnie od danej metody rzutowania obrazu, różnym punktom przypisuje się wagi [Dzi+22], które są uwzględniane podczas liczenia sumarycznych błędów średniokwadratowych, MSE, oraz IV-MSE, odpowiednio dla PSNR i IV-PSNR. Dla sekwencji z kamerami sferycznymi i półsferycznymi (rozdział 5.4.4) wartość metryki IV-PSNR jest wyznaczana w taki właśnie zmodyfikowany sposób, z uwzględnieniem rzutowania. Przykładowo, dla wspomnianego rzutowania ERP, ważona metryka PSNR nosi nazwę WS-PSNR [SLY17].

Dokładny sposób wyznaczania IV-PSNR oraz WS-PSNR można znaleźć m.in. w publikacjach [Dzi+22] oraz [SLY17]. Opis ten uwzględnia również problem poszukiwania odpowiadającego punktu dla różnych formatów próbkowania chrominancji w metryce IV-PSNR, który w niniejszej pracy również został pominięty ze względu na oszczędność miejsca i małą istotność dla wyników rozprawy.

4 Optymalizacja ustawienia kamer w systemach swobodnej nawigacji

4.1 Opis problemu

System akwizycji, który rejestruje dane za pomocą wielu kamer umieszczonych wokół sceny, umożliwia użytkownikowi usługę swobodnej nawigacji [Tan10]. Pierwszy etap przygotowania wirtualnego widoku to estymacja map głębi [MMW11][Sta+13]. Jest ona możliwa dla wszystkich obszarów sceny, jeśli zostały one zarejestrowane przez co najmniej dwie kamery systemu. Próbki, które nie są widoczne w co najmniej dwóch obrazach, będą nazywane **przysłoniętymi**. Dla takich obszarów nie jest możliwe wyznaczenie mapy głębi poprzez znalezienie odpowiadających próbek w widokach pozostałych kamer. Wartości mapy głębi dla obszarów przysłoniętych mogą zostać wyznaczone np. na podstawie obszarów sąsiednich lub błędnego dopasowania, co powoduje spadek jakości wyznaczanych danych. Niska jakość map głębi oraz błędy estymacji wpływają na proces syntezy wirtualnych widoków i będą powodować pojawienie się wad wirtualnego obrazu. Dodatkowo proces syntezy takiego widoku dokonuje przetwarzania przez m.in. filtrację (np. dla obszarów, do których zostało przerzutowanych kilka próbek) i wypełnianie (np. dla obszarów bez przerzutowanych próbek) [SD20][Dzi+19]. Może to powodować zmianę jakości wyświetlanego widoku, szczególnie jeśli mapy głębi zawierają dużo punktowych obszarów z niepoprawną wartością głębi lub małą jej dokładnością.

Celem tego rozdziału pracy jest znalezienie zależności określającej, jak położenie kamer w scenie wpływa na jakość systemu swobodnej nawigacji. Modyfikacji zostanie poddana pozycja kamer środkowych, aby skrajne kamery systemu niezmiennie rejestrowały ten sam obszar, który stanowi całą scenę. Celem optymalizacji ustawienia kamer będzie uzyskanie wysokiej jakości widoków wirtualnych. Zostanie ona zmierzona poprzez zmianę uśrednionej wartości metryki oceny jakości dla syntezowanych widoków.

4.2 Wielokamerowa akwizycja

Aby zmniejszyć negatywne efekty powodowane przede wszystkim przez rzadkie rozmieszczenie kamer (przysłonięcia oraz inne warunki akwizycyjne), autorzy

[Dom+16f] zaproponowali pogrupowanie ich w pary zamiast równomiernego rozmieszczenia wokół sceny. W tym podejściu kamery z tej samej pary rejestrują scenę z bardzo podobnej perspektywy. Można wymienić czynniki, które sprawiają, że mapa głębi estymowana z pary kamer powinna zawierać mniej błędów i sprawiać, że syntezowane wirtualne widoki będą miały postrzeganą wyższą jakość. Nieduża odległość pomiędzy kamerami w parze sprawia, że bardzo niewiele fragmentów sceny jest przysłoniętych, a warunki oświetleniowe są podobne. Kolejny z czynników związany jest m.in. z powierzchniami odbijającymi, półprzezroczystymi oraz nielambertowskimi odbiciami, które mogą pojawić się w naturalnych sekwencjach. Mniejsza odległość pomiędzy kamerami sprawia, że obserwowane obiekty nierównomiernie odbijające, rozpraszające światło, półprzeźroczyste lub świecące wyglądają podobnie.

Ustawianie kamer w pary ma również skutki negatywne. Główny z nich związany jest z dokładnością estymowanej mapy głębi. Niewielka odległość bazowa kamer powoduje, że wyznaczana mapa głębi ma małą dokładność [Mie18]. Rozbieżności w obrazach takich kamer, które są umieszczone bardzo blisko siebie, mogą ograniczać się np. do kilku okresów próbkowania, co przekłada się na tyle samo rozróżnialnych wartości mapy głębi (rozwinięcie tej zależności zostało przedstawione w rozdziale 4.4). Teoretycznie, dużą dokładność estymowanej mapy głębi można uzyskać, porównując obrazy dwóch kamer, które są bardzo odległe od siebie, czyli posiadają dużą odległość bazową. Jednakże widoki, zarejestrowane przez takie kamery, są mniej do siebie podobne, co często uniemożliwia estymację mapy głębi.

Wyżej wymienione sprzeczne zjawiska wpływają na estymację głębi, a tym samym na jakość syntezowanych wirtualnych widoków. W celu syntezy wirtualnych widoków o wysokiej jakości należy znaleźć kompromisową odległość pomiędzy kamerami, która jednocześnie wpływa na podobieństwo widoków oraz na dokładność wyznaczanej mapy głębi. Do określenia tej zależności została użyta metryka PSNR [Wan+04b] ze względu na powszechność wykorzystania oraz możliwość porównania wyników (np.w grupie syntezowanych tą samą metodą rezultatów, podobnie jak w [Sha+14][Wan+17]). Dlatego zaproponowano pomiar zmiany wartości metryki, co sprawia, że wyniki dla IV-PSNR (lepiej dostosowanej do widoków syntezowanych [Dzi+22]) byłyby podobne. Zmiana tej wartości została oznaczona przez Δ, obliczona pomiędzy wartościami PSNR widoku wirtualnego dla kamer ustawionych w pary (oznaczonych indeksem p — ang. pair) oraz tych samych kamer rozmieszczonych równomiernie (oznaczonych indeksem u — ang. uniform distribution) i wyrażona

jako suma dwóch składowych, związanych z dwoma opisanymi wyżej zjawiskami:

$$\Delta_{p-u}\text{PSNR} = \Delta_{p-u}\text{PSNR}_b + \Delta_{p-u}\text{PSNR}_o, \tag{4.1}$$

gdzie $\Delta_{p-u}\text{PSNR}_b$ jest zmianą jakości wynikającą z innej odległości pomiędzy parami kamer (b - ang. base), a $\Delta_{p-u}\text{PSNR}_o$ jest zmianą jakości wynikającą z innej liczby przysłonięć (o - ang. occlusions), wyrażonymi jako różnica między wartościami PSNR dla kamer ustawionych równomiernie oraz w pary. Różnica ta jest oznaczona przez Δ_{p-u}. W równaniu 4.1 oba składniki odpowiadające zmianom dokładności estymacji map głębi oraz liczby przysłonięć związanych z ustawieniem kamer w pary są dokładniej omówione w rozdziałach 4.4 oraz 4.5.

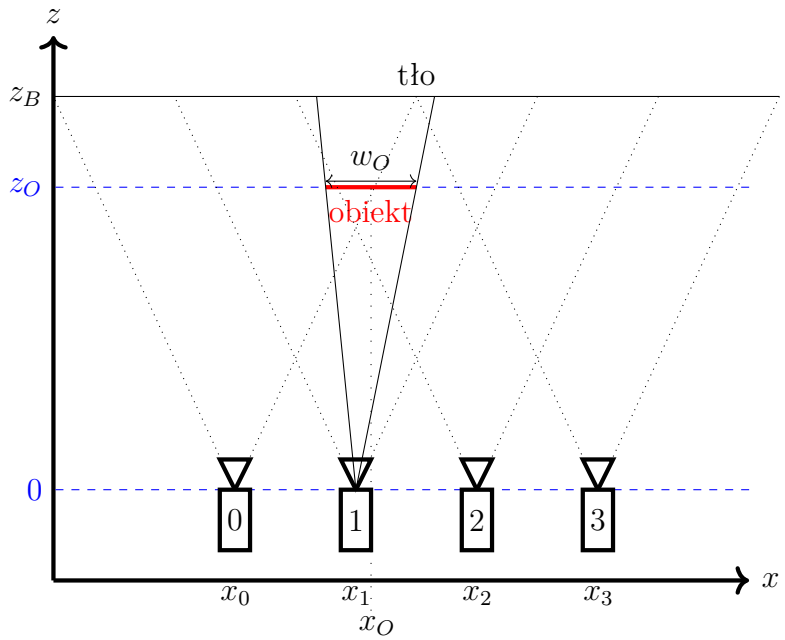
Wpływ różnic oświetlenia oraz odbić w scenie na jakość wirtualnego widoku nie zostały uwzględnione w tym rozdziale. Istnieje wiele prac na ten temat (np. [ZYZ13][NDC17]) i jest to ważne zagadnienie, które w przyszłości może uzupełnić przedstawione badania. W niniejszym rozdziale założono, że wszystkie odbicia są lambertowskie, źródła światła punktowe, a sekwencje nie zawierają obiektów półprzezroczystych oraz odbijających. Jest to również cecha większej części sekwencji ze zbioru testowego, który w punkcie 4.7 wykorzystano do potwierdzenia wyników teoretycznych (sekwencje naturalne, w których mogą wystąpić wyżej wymienione zjawiska to jedynie sekwencje Dog oraz Champange).

4.3 Założenia

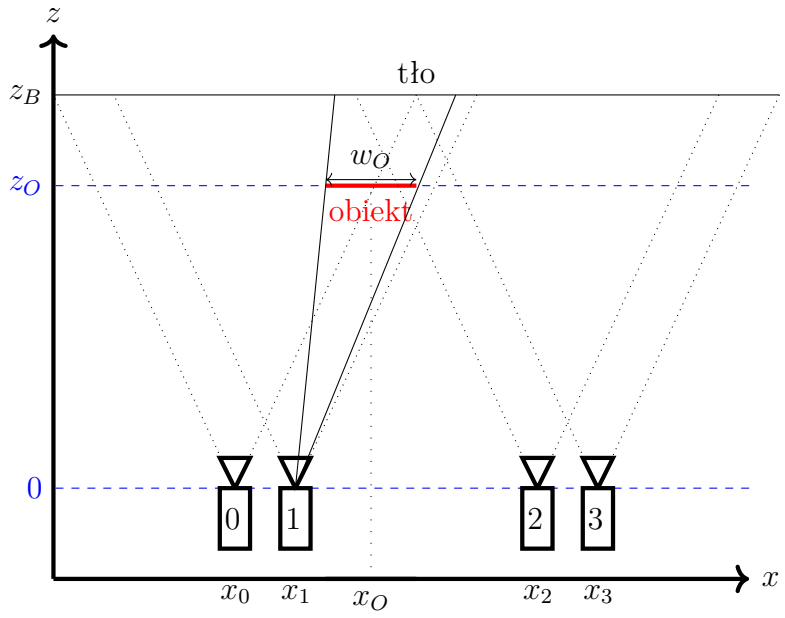
4.3.1 Model systemu z przysłonięciami w scenie

W tym rozdziale został wprowadzony model systemu wielokamerowego. Wykorzystano go kolejnych rozdziałach do wyznaczenia modelu wpływu odległości pomiędzy kamerami na jakość wirtualnego widoku (rozdział 4.4) oraz w części eksperymentalnej (rozdział 4.7).

Oszacowanie zmiany jakości $\Delta_{p-u}\text{PSNR}$, wynikające z pogrupowania kamer w pary (zamiast równomiernego ustawienia), zostało przeprowadzone dla modelu systemu, którym jest rzut z góry. Składa się on z 4 kamer i został przedstawiony na rysunkach 4.1a oraz 4.1b. Kamery zostały umieszczone na płaszczyźnie $z = 0$ oraz w punktach od x_0 do x_3 w kierunku osi x. Wszystkie mają takie same parametry (m.in. pole widzenia ang. Angle of View, długość ogniskową). Na pierwszym planie scena zawiera pojedynczy obiekt, który zasłania tło. Obiekt ten ma szerokość w_O,



(a) Model systemu wielokamerowego o równomiernym rozmieszczeniu kamer



(b) Model systemu wielokamerowego o rozmieszczeniu kamer w parach

Rysunek 4.1: Modele systemu wielokamerowego

a położenie jego środka wynosi x_O,z_O. Tło jest umieszczone w odległości z_B od kamer. Przedstawiony wyżej model został wykorzystany do oszacowania zmiany PSNR z umieszczania kamer w parach Δ_{p-u}PSNR.

Dla prostego modelu z rysunku 4.1a zbiór próbek widocznych dla kamery numer 1 pokazano na rys. 4.2 i oznaczono czerwoną linią kropkowaną. Dla kamery oznaczonej indeksem i na każdy obiekt przypadają cztery charakterystyczne punkty: $F_L(i,z)$, $F_R(i,z)$, które są przecięciem granic pola widzenia (FOV) kamery i z dowolną płaszczyzną P umieszczoną w odległości z, oraz $O_L(i,z)$, $O_R(i,z)$, które oznaczają przecięcia linii łączących kamerę i z najbardziej wysuniętym w lewo i prawo punktem obiektu z płaszczyzną w odległości z. Dla przedstawionego modelu punkt może znajdować się w odległości $z = z_P$ lub $z = z_O$. Wszystkie próbki, które są widoczne dla kamery i, można zdefiniować jako:

$$\mathbb{C}_i = \mathbb{B}_i \cup \mathbb{O}_i, \tag{4.2}$$

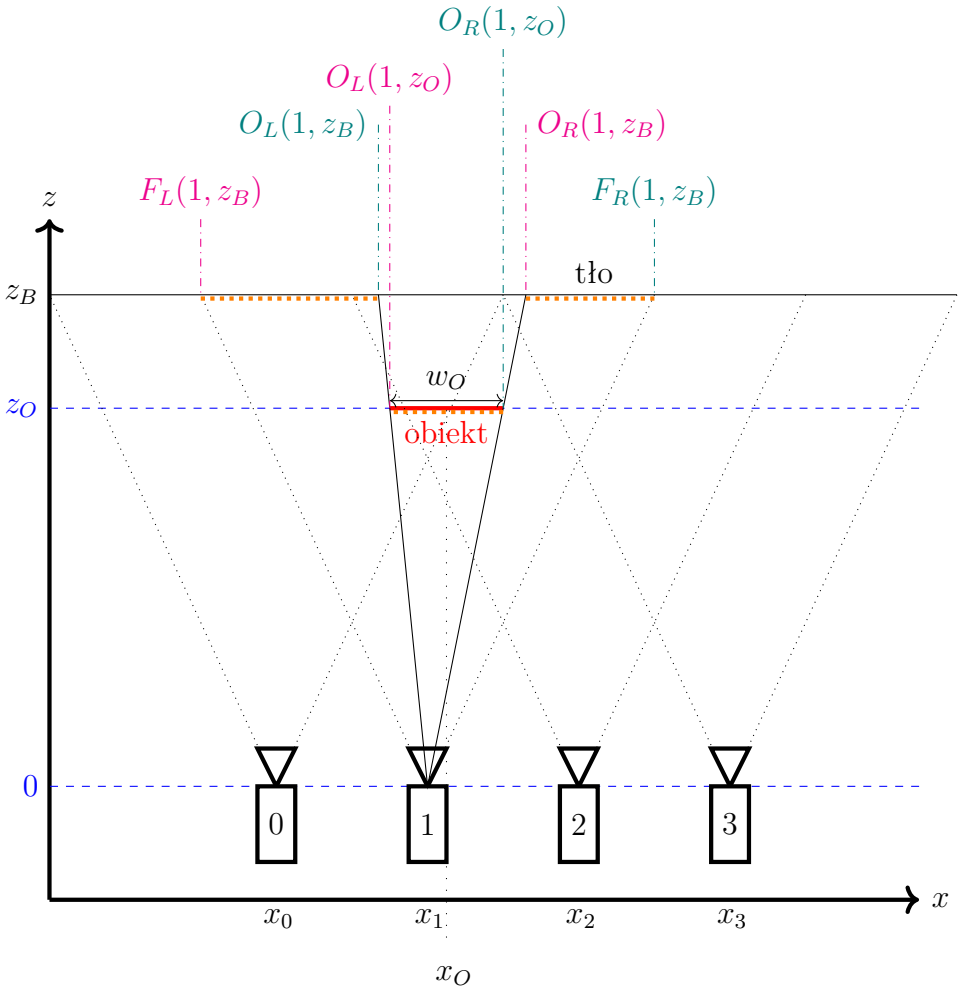
gdzie $\mathbb{B}_i$ jest zbiorem próbek tła zarejestrowanym przez kamerę i, a $\mathbb{O}_i$ oznacza zbiór próbek pierwszego planu widoczny w kamerze i:

$$\mathbb{B}_i = [F_L(i,z_B), F_R(i,z_B)] - (O_L(i,z_B), O_R(i,z_B)), \tag{4.3}$$

$$\mathbb{O}_i = [F_L(i,z_O), F_R(i,z_O)] \cap [O_L(i,z_O), O_R(i,z_O)]. \tag{4.4}$$

Wartości w nawiasie kwadratowym oznaczają zbiór próbek pomiędzy współrzędnymi w osi poziomej, np. $F_L(i,z_B)$ oraz $F_R(i,z_B)$. Dla sceny przedstawionej na rysunku 4.2 obszar widoczny z kamery $\mathbb{C}_1$ oznaczono przerywaną czerwoną linią.

Dla oszacowania zmiany $\Delta_{p-u}PSNR$ przyjęto założenia, które umożliwiają porównanie wyników z eksperymentem przedstawionym w rozdziale 4.7. Parametry te przybliżają rzeczywiste sceny wykorzystywane w części eksperymentalnej przedstawionej w rozdziale 4.8. Założono, że kamery 0 i 3 nie zmieniają swojego położenia oraz zdefiniowano odległości pomiędzy nimi: $x_0 = 0$, $x_3 = 3$. Sprawia to, że dla równomiernego położenie kamer, wartości x_1 oraz x_2 wynoszą odpowiednio 1 oraz 2. Dla ustawienia kamer w pary natomiast założono, że położenie kamer 1 oraz 2 wynosi odpowiednio 0,4 oraz 2,6. Przyjęto, że posiadają one pole widzenie 70 stopni oraz 1920 elementów światłoczułych, co odpowiada poziomej wielkości ramki dla formatu HD. Założono, że tło znajduje się w odległości $z_B = 6$. W celu modelowania różnych wielkości przysłonięć, szerokość obiektu w_O oraz położenie środka obiektu z_O zmieniają się odpowiednio od 0,2 do 2,8 oraz 0,2 do 5,8.



Rysunek 4.2: Model systemu wielokamerowego o równomiernym rozmieszczeniu kamer — czerwona linia kropkowana to obszar widoczny przez kamerę numer 1

4.3.2 Model jakości syntezowanego widoku

W rozdziale 4.3.2 został wprowadzony model średniej jakości syntezowanego widoku. Jest on wykorzystywany do oszacowania wpływu przysłonięć w scenie i odległości pomiędzy kamerami na jakość wirtualnego widoku (rozdziały 4.4 oraz 4.5).

Błędne wartości w estymowanej mapie głębi powodują zazwyczaj poziome przesunięcia fragmentów wirtualnego widoku. W przypadku niejednolitych obszarów sceny przesunięcia te znacznie pogarszają jakość takiego widoku, natomiast w przypadku obszarów jednolitych utrata jakości jest często niezauważalna. W celu oceny wyżej

wymienionych efektów zaproponowano model średniej jakości syntezowanego widoku. Podobieństwo $S(n)$ to wartość od 0 do 1, gdzie wartość jednostkowa oznacza, że syntezowany widok jest dokładnie taki sam jak widok, który byłby zarejestrowany przez kamerę rzeczywistą. Wzór definiujący metrykę podobieństwa $S(n)$, która wyraża różnicę pomiędzy widokiem wirtualnym odniesienia (czyli widokiem bez zniekształceń) a widokiem przesuniętym o n okresów próbkowania względem jego prawidłowej pozycji, tj. z pozycji obliczonej przy użyciu idealnych map głębi jest następujący:

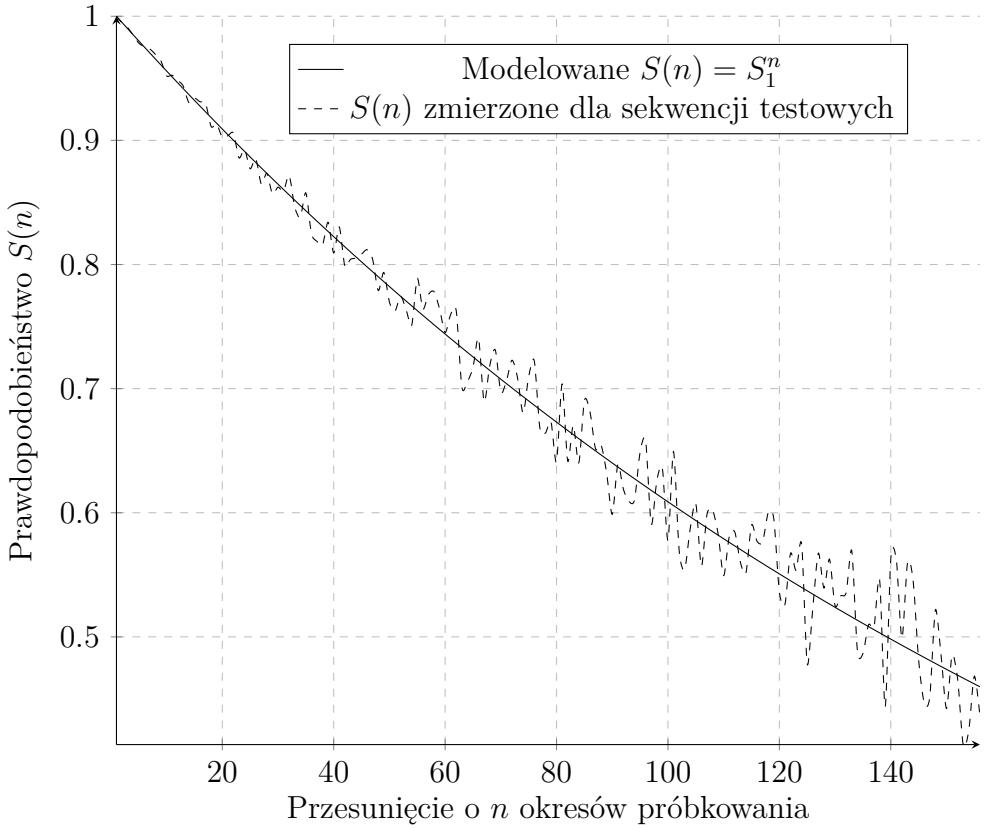
$$S(n) = 1 - \frac{1}{H_{img} \cdot (W_{img} - n)} \sum_{j=1}^{H_{img}} \sum_{i=1}^{W_{img}-n} \frac{(Y(i,j) - Y(i+n,j))^2}{255^2}, \tag{4.5}$$

gdzie $Y(i,j)$ jest wartością luminancji próbki, a 255^2 jest maksymalnym możliwym kwadratem błędu 8-bitowej próbki. W_{img} i H_{img} to odpowiednio szerokość i wysokość obrazu $Y(i,j)$. Iloczyn $H_{img} \cdot (W_{img} - n)$ to liczba próbek obrazów, które można porównać, przesuwając obraz o n okresów próbkowania. Z równania 4.5 wynika, że dla syntezowanego widoku, który jest zniekształcony przez przesunięcie o n okresów próbkowania, PSNR luminancji można wyznaczyć następująco:

$$\mathrm{PSNR}(n) = -10\log(1 - S(n)), \tag{4.6}$$

gdzie punktem odniesienia do obliczanego PSNR jest widok syntezowany z idealnych map głębi. Przy użyciu zestawu sekwencji testowych z wieloma widokami wykorzystanych w rozdziale 4.8 i przedstawionych w tabeli 4.1 zmierzono $S(n)$ dla wartości całkowitych przesunięcia n. Jak pokazano na rysunku 4.3, zmierzony $S(n)$ rozpoczyna się od $S(1) = S_1 = 0,995$ i maleje w przybliżeniu wykładniczo. W dalszych rozdziałach został założony przybliżony model analityczny $S(n)$ zdefiniowany dla rzeczywistych wartości przesunięcia $n > 0$:

$$S(n) = S_1^n. \tag{4.7}$$



Rysunek 4.3: Podobieństwa S(n) między próbkami, które są przesuwane poziomo o n okresów próbkowania [Sta+18a]

4.4 Model wpływu odległości pomiędzy kamerami na jakość wirtualnego widoku

W tym rozdziale przedstawiono wyniki rozważań na temat wpływu odległości pomiędzy kamerami na jakość estymowanej mapy głębi i wirtualnych widoków. W pierwszym etapie rozpatrzona zostanie estymacja jedynie dla pary kamer, a następnie analizę rozszerzono na przypadek ogólny.

Załóżmy, że długości ogniskowe wykorzystanych kamer mają taką samą wartość, która wynosi f, odległość pomiędzy środkami optycznymi kamer to B, a głębia obiektu punktowego — z (jak w rozdziale 2.5). Rozbieżność punktowego obiektu zarejestrowana przez kamery to d. Zakładając $f << z$, otrzymujemy [HZ04][CS09]:

$$z = \frac{f \cdot B}{d}. \tag{4.8}$$

Załóżmy, że w scenie występują dwa obiekty o podobnej wartości głębi — odpowiednio z_1 i z_2. Ich głębia będzie miała inną wartość, jeśli różnica rozbieżności $|d_1 - d_2|$ przekroczy pewną wartość minimalną Δd_{min}:

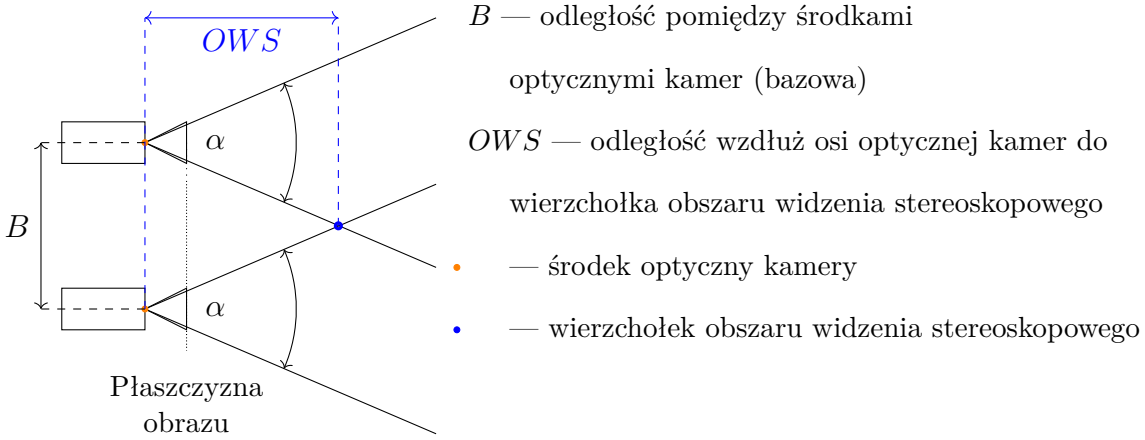
$$|d_1 - d_2| \geq \Delta d_{min}, \tag{4.9}$$

gdzie Δd_{min} jest rozróżnialną zmianą położenia obiektu w rozbieżności, która wynika ze zdolności rozdzielczej kamer. Załóżmy, że wartość ta wynosi jeden okres próbkowania przestrzennego, chociaż z uwagi na m.in. szum i brak ostrości krawędzi wielkość ta może być większa. Po przekształceniu równania 4.8 otrzymujemy: $d_1 = \frac{fB}{z_1}, d_2 = \frac{fB}{z_2}$. Załóżmy, że z jest średnią wartością głębi, wokół której skupione są obiekty o głębi z_1 oraz z_2, oraz że średnia ta będzie średnią geometryczną, czyli $z = \sqrt{z_1 z_2}$. Po przekształceniu równania 4.9 wartości głębi z_1 oraz z_2 można rozróżnić gdy:

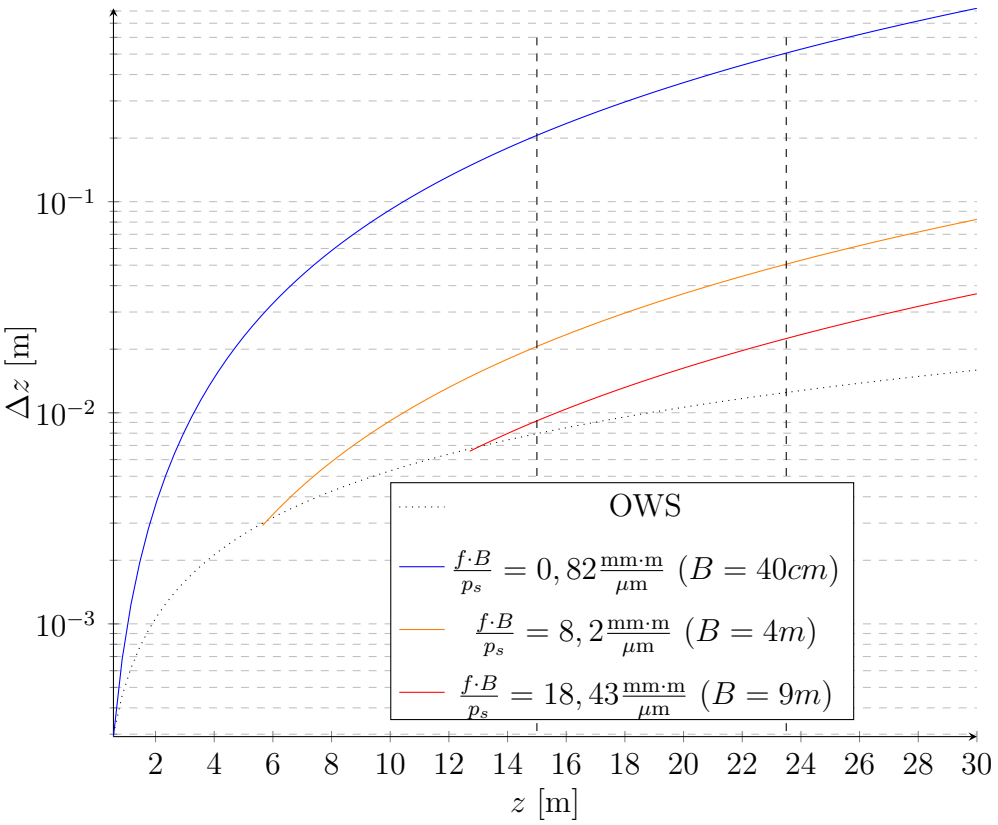
$$\Delta z = |z_1 - z_2| \geq \frac{z^2}{f \cdot B} \Delta d_{min}. \tag{4.10}$$

Δz jest rozróżnialną zmianą położenia obiektu w głębi, która będzie nazywana różnicą głębi.

Dla pary kamer można wyznaczyć punkt, w którym zaczyna się obszar stereoskopowego widzenia (rysunek 4.4). Zgodnie ze wzorem 4.10 głębia oraz różnica głębi są wprost proporcjonalne, więc dla głębi, która jest równa odległości do początku widzenia stereoskopowego ($z = OWS$), różnica głębi (Δz) będzie najmniejsza.



Rysunek 4.4: Wierzchołek obszaru widzenia stereoskopowego dla liniowego ustawienia dwóch kamer



Rysunek 4.5: Różnica głębi Δz w funkcji głębi stereoskopowej z, w nawiasie podana wartość odległości bazowej dla f i Δd_{min} z przykładu

Przykład:
Rozważmy kamerę Basler acA 1920-155uc [wBA], która ma przetworniki w formacie 1920×1200. Załóżmy, że rozróżnialna zmiana położenia w rozbieżności wynosi $\Delta d_{min} = 5,86\mu m$, (co odpowiada rozmiarowi elementu światłoczułego), oraz że zastosowany obiektyw ma ogniskową $f = 16$mm. Pole widzenia kamery w poziomie będzie wówczas wynosić około 39° [wBL].
Na wykresie 4.5 została przedstawiona różnica głębi Δz w zależności od średniej wartości głębi z dla trzech przykładowych odległości pomiędzy środkami optycznymi kamer B (wzór 4.10).
Dla średniej głębi $z = 23,5$m oraz odległości pomiędzy kamerami $B = 40$cm otrzymujemy różnicę głębi $\Delta z \geq 0,5m$, natomiast dla odległości pomiędzy kamerami $B = 4$m otrzymujemy $\Delta z \geq 0,05$m.
Gdy rozważymy obiekt o średniej głębi $z = 15$m, różnica głębi Δz wynosi odpowiednio $0,2$m dla odległości bazowej $B = 40$cm i 10 razy mniej dla odległości bazowej wynoszącej $B = 4$m — zgodnie ze wzorem 4.10.

Powyższe rozumowanie wyjaśnia fakt, że mapa głębi może być estymowana z dużą dokładnością, gdy odległość pomiędzy kamerami jest duża. Dlatego też, ze względu na różnice głębi (czyli dokładność estymacji mapy głębi), w przypadku wielu kamer, głębia powinna być estymowana z wykorzystaniem dwóch, które są najdalej położone w systemie.

W scenach z przysłonięciami poszczególne obszary mogą nie być rejestrowane przez wszystkie kamery. Analizując model systemu liniowego (przedstawiony w punkcie 4.3.1) z czterech kamer, które zarejestrowały dany obszar, do estymacji mapy głębi należy wybrać te najbardziej oddalone od siebie czyli skrajne. Będzie to sprawiało, że różnica głębi Δz, czyli błąd estymacji mapy głębi, będzie najmniejszy. Dla kamer ułożonych równomiernie można określić wartość B_u oznaczającą największą uśrednioną odległość pomiędzy kamerami dla wszystkich próbek sceny (dla których można wyznaczyć mapę głębi). Analogicznie B_p jest tą samą wartością dla kamer ustawionych w pary. Wyprowadzenie tych wartości znajduje się poniżej i zostało wyrażone wzorem 4.16.

Przy założeniu zbioru próbek widzianych przez każdą kamerę jak we wzorze 4.2, można określić, przez które pary kamer zostały zarejestrowane poszczególne obszary sceny. Rozróżnić można 4 zbiory próbek sceny:

$$\mathbb{K} = \mathbb{C}_0 \cap \mathbb{C}_3, \tag{4.11}$$

$$\mathbb{L} = (\mathbb{C}_0 \cap \mathbb{C}_2) \cup (\mathbb{C}_1 \cap \mathbb{C}_3) - \mathbb{K}, \tag{4.12}$$

$$\mathbb{M} = (\mathbb{C}_1 \cap \mathbb{C}_2) - (\mathbb{K} \cup \mathbb{L}), \tag{4.13}$$

$$\mathbb{N} = ((\mathbb{C}_0 \cap \mathbb{C}_1) \cup (\mathbb{C}_2 \cap \mathbb{C}_3)) - (\mathbb{K} \cup \mathbb{L} \cup \mathbb{M}), \tag{4.14}$$

gdzie $\mathbb{K}$ jest zbiorem próbek widocznych ze skrajnych kamer systemu, $\mathbb{L}$ oznacza zbiór próbek widocznych przez kamerę zewnętrzną jednej z par kamer i kamerę wewnętrzną drugiej pary, $\mathbb{M}$ to zbiór próbek widocznych dla kamer wewnętrznych systemu z dwóch par, a $\mathbb{N}$ jest zbiorem próbek widocznych przez kamery jednej pary. Zbiór próbek $\mathbb{D}$ zdefiniowano przez połączenie wszystkich zbiorów próbek widzianych przez co najmniej dwie kamery, czyli próbek, dla których możliwa jest estymacja map głębi:

$$\mathbb{D} = \mathbb{K} \cup \mathbb{L} \cup \mathbb{M} \cup \mathbb{N}. \tag{4.15}$$

Załóżmy, że średnia odległość pomiędzy kamerami dla równomiernego ich rozmieszczenia B_u będzie średnią ważoną maksymalnych odległości pomiędzy kamerami dla całej sceny:

$$B_u = \frac{B_{max}|\mathbb{K}| + B_1|\mathbb{L}| + B_2|\mathbb{M}| + B_{min}|\mathbb{N}|}{|\mathbb{D}|}, \tag{4.16}$$

gdzie $B_{max} = x_3 - x_0$ to odległość pomiędzy skrajnymi kamerami, B_{min} — między kamerami w obrębie jednej pary kamer, $B_1 = B_{max} - B_{min}$ — odległość pomiędzy kamerami dwóch par, czyli wewnętrznej oraz zewnętrznej, $B_2 = B_{max} - 2B_{min}$ — wewnętrzne kamery systemu z dwóch par. Operator $|\cdot|$ oznacza liczebność zbiorów zarejestrowanych próbek, ale może też być interpretowany jako suma długości wszystkich ciągłych podzbiorów punktów w zbiorze (). Średnia odległość dla kamer ustawionych w pary B_p jest obliczana w taki sam sposób.

Dla równomiernego rozmieszczenia kamer oznaczamy średnią bazę jako B_u, a maksymalne błędne przesunięcie pozycji próbki widoku wirtualnego jako Δp_u. Pozycja próbki może być wyrażona tylko w liczbach całkowitych, dlatego zakładając idealny proces estymacji map głębi oraz syntezy wirtualnego widoku, można założyć, że jedynym źródłem błędów przy syntezie wirtualnego widoku są zaokrąglenia. Dlatego przyjmujemy, że $\Delta p_u = 0,5$. Jak wynika z równania 4.10, dla par kamer, gdzie

średnia baza B_p jest krótsza, rozróżnialna zmiana położenia w głębi, czyli różnica głębi, Δz rośnie. Założono, że dokładność położenia próbki w widoku wirtualnym będzie wynosić $\Delta p_p = \Delta p_u B_u / B_p$. Zmiana jakości tego widoku $\Delta_{p-u}\text{PSNR}_b$, po uwzględnieniu równania 4.6, to:

$$\Delta_{p-u}\text{PSNR}_b = 10log\frac{1 - s_1^{\Delta p_u}}{1 - s_1^{\Delta p_p}}. \tag{4.17}$$

4.5 Model wpływu przysłonięć w scenie na jakość wirtualnego widoku

Rozważamy pierwszy składnik zmiany jakości widoku $\Delta_{p-u}\text{PSNR}$, zdefiniowany w równaniu 4.1. $\Delta_{p-u}\text{PSNR}_o$ Jest to zmiana jakości, która wynika ze zmiany ilości obszarów przysłoniętych dla systemu z parami kamer względem systemu z kamerami rozmieszczonymi równomiernie. W celu oszacowania tej różnicy przeanalizujmy najpierw wpływ przysłonięć na proces syntezy widoku wirtualnego.

Typowa technika syntezy obrazu [WSD15] oparta na DIBR (Depth-Image-Based Rendering) [SJ15] tworzy wirtualny widok w dwóch etapach. Podczas pierwszego z nich fragmenty obrazów z widoków wejściowych są przenoszone na nowe pozycje w wirtualnym widoku i łączone ze sobą. Na tym etapie niektóre obszary są nieznane, ponieważ zostały przysłonięte we wszystkich widokach wejściowych. Te obszary są wypełniane w drugim kroku [Dzi+19]. Powstały wirtualny obraz składa się z dwóch rodzajów obszarów: syntezowanego i wypełnionego. Jakość wirtualnego widoku w obszarach wypełnionych jest zazwyczaj gorsza niż w syntezowanych, ponieważ wypełnianie opiera się na sąsiednich syntezowanych obszarach, a więc błędy syntezy dodatkowo mogą zostać powiększone przez wypełnianie obszarów przysłoniętych. W celu określenia zmiany jakość wirtualnego widoku w odniesieniu do liczby przysłonięć w scenie autor zaproponował następujący wzór:

$$\Delta_{p-u}\text{PSNR}_o = 10 \cdot log\frac{OCC_u \cdot e_s^2 + (1 - OCC_u) \cdot e_i^2}{OCC_p \cdot e_s^2 + (1 - OCC_p) \cdot e_i^2}, \tag{4.18}$$

gdzie e_s^2 i e_i^2 to średnie błędy kwadratowe odpowiednio w obszarach syntezowanych i wypełnionych (ang. inpainted), OCC_u i OCC_p, stanowią współczynnik udziału liczby próbek obszarów przysłoniętych odpowiednio dla równomiernie rozmieszczonych kamer (indeks dolny u od ang. uniform) i dla par kamer (indeks dolny p od ang.

pair). Obszary przysłonięte to fragmenty sceny, w których nie można było określić głębi, czyli np. zarejestrowane jedynie przez jedną kamerę systemu.

Zbiór wszystkich próbek sceny $\mathbb{S}$ można zdefiniować jako sumę zbiorów próbek zarejestrowanych przez kamery. Nierejestrowane fragmenty tła nie są brane pod uwagę, ponieważ nie będą one również widoczne z żadnego z wirtualnych widoków. Ustawienie kamer w pary może powodować, że środek obiektu nie będzie zarejestrowany przez żadną z nich, ale będzie syntezowany dla wirtualnych kamer. Dlatego, aby zagwarantować, że cały obiekt znajduje się w zbiorze $\mathbb{S}$, definiujemy zbiór $\mathbb{S}'$:

$$\mathbb{S}' = \mathbb{S} \cup [O_L, O_R], \tag{4.19}$$

gdzie O_L i O_R oznaczają najbardziej wysunięty w lewo i prawo punkt obiektu. Zbiór $\mathbb{S}'$ zawiera wszystkie próbki sceny, które mogą pojawić się w wirtualnym widoku pomiędzy kamerami 0 oraz 3.

Liczba przysłonięć w scenie OCC_u będzie zdefiniowana jako liczba próbek o nieokreślonej wartości głębi w stosunku do liczebności całego zbioru — dla równomiernego rozmieszczenia kamer. Jest on obliczony w następujący sposób:

$$OCC_u = \frac{|\mathbb{S}' - \mathbb{D}|}{|\mathbb{S}'|}, \tag{4.20}$$

gdzie $\mathbb{D}$ jest liczebnością zbioru próbek, dla których możliwa jest estymacja mapy głębi (dokładane wyprowadzenie w rozdziale 4.4). Operator $|\cdot|$ oznacza liczebność zbiorów zarejestrowanych przez kamery próbek. OCC_p można obliczyć w podobny sposób.

Założono, że algorytm syntezy wirtualnych widoków wykorzystuje sąsiednie, przerzutowane próbki do wypełnienia obszarów niezsyntezowanych, a średni błąd kwadratowy syntezowanego obszaru (e_s^2) jest k razy mniejszy niż średni błąd kwadratowy wypełnionego obszaru ($e_i^2 = ke_s^2$). Wzór 4.18 można zatem przedstawić następująco:

$$\Delta_{p-u}\mathrm{PSNR}_o = 10 \cdot log\frac{(1 + (1/k^2 - 1) \cdot OCC_u)}{(1 + (1/k^2 - 1) \cdot OCC_p)}. \tag{4.21}$$

Korzystając z modelu systemu wielokamerowego przedstawionego w rozdziale 4.3.1, otrzymujemy dwa zasłonięte obszary po lewej i prawej stronie obiektu. W każdym z nich znajduje się połowa przysłoniętych próbek, czyli $OCC_u/2$ dla równomiernego rozmieszczenia kamer i $OCC_p/2$ dla kamer ustawionych w pary. Podstawowe

techniki syntezy wirtualnego widoku wykorzystują próbki sąsiadujące z obszarem niesyntezowanym, który jest wypełniany. Błąd wypełniania jest zatem związany z podobieństwem pomiędzy sąsiednimi punktami. Większy przysłonięty obszar spowoduje, że wartość średniego błędu wypełniania będzie większa. Niech k oznacza średnie podobieństwo wypełnianych punktów do najbliższej syntezowanej próbki. Załóżmy, korzystając z przyjętego modelu jakości syntezowanego widoku (rozdział 4.3.2), że prawdopodobieństwo pomiędzy kolejnymi wypełnianymi próbkami widoku wirtualnego tworzy szereg geometryczny:

$$S(n) = S_1^n, \tag{4.22}$$

a zatem dla równomiernego rozmieszczenia kamer:

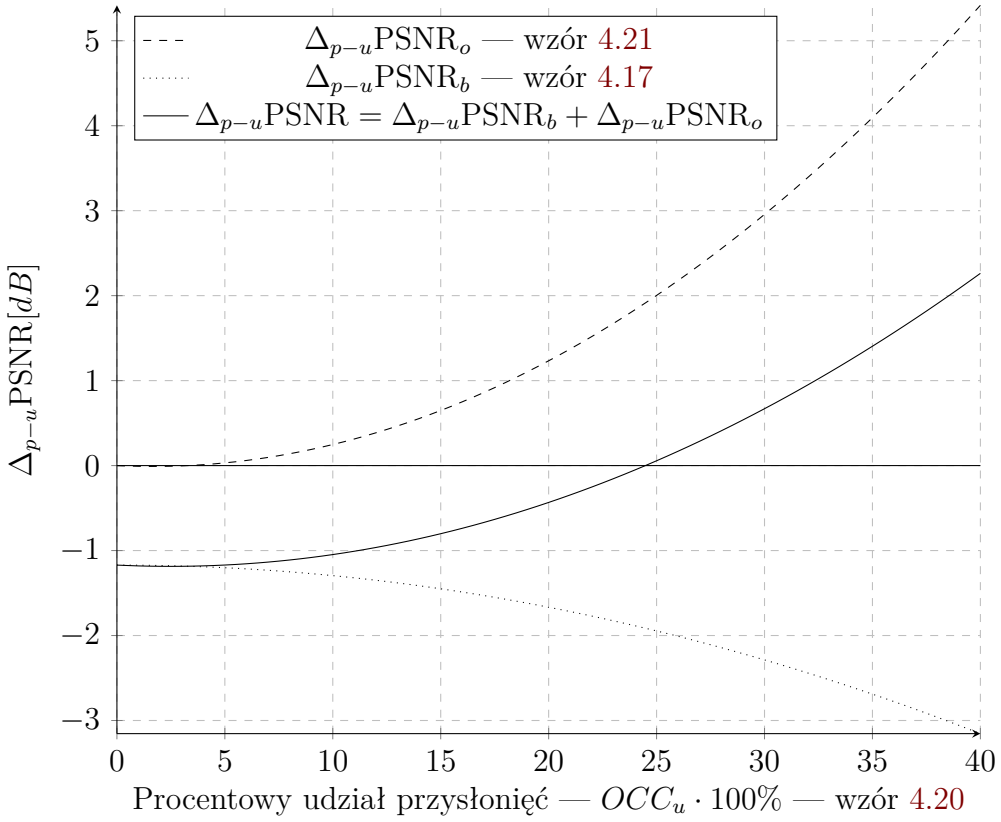
$$k = \frac{\frac{S_1}{1-S_1} \cdot (1 - S_1^{W_{cam} OCC_u/2})}{W_{cam} \cdot OCC_u/2}, \tag{4.23}$$

gdzie W_{cam} jest liczbą światłoczułych elementów w przetworniku kamery. Wyznaczenie k dla kamer ustawionych w pary jest analogiczne. Zmianie ulega jedynie procentowy udział przysłoniętych próbek — OCC_u na OCC_p.

4.6 Wyniki teoretyczne

Rysunek 4.6 przedstawia zmianę wartości Δ_{p-u}PSNR w zależności od liczby przysłonięć w teoretycznej modelowej dwuwymiarowej scenie z przysłonięciami, przedstawionej w rozdziale 4.3.1. Na wykresie zaprezentowano wpływ przysłonięć w scenie $\Delta_{p-u}\text{PSNR}_o$ zgodnie ze wzorem 4.21 z rozdziału 4.5 oraz wpływ odległości pomiędzy kamerami na jakość estymowanych map głębi i widoków wirtualnych $\Delta_{p-u}\text{PSNR}_b$ pokazanych w rozdziale 4.4, zgodnie ze wzorem 4.17.

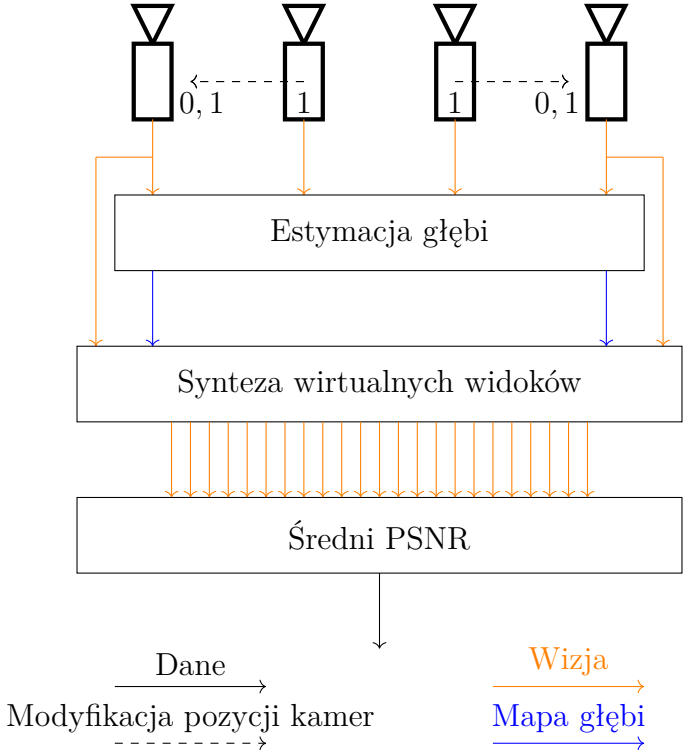
Wyniki pokazują, iż stosowanie par kamer zamiast równomiernego ich rozmieszczenia jest korzystne, gdy procentowy udział przysłoniętych powierzchni jest większy niż około 25%. Przyjęto, że ustawienie kamer w pary zmniejsza dystans pomiędzy nimi o 60% (0,4 znormalizowanej rozbieżności).



Rysunek 4.6: Zmiana wartości PSNR dla par kamer w funkcji $OCC_u \cdot 100\%$ - krzywa teoretyczna [Sta+18a]

4.7 Badania eksperymentalne

Wyniki teoretyczne przedstawione w poprzednim rozdziale (4.6) zostały poddane sprawdzeniu poprzez przedstawiony w tym rozdziale eksperyment. Jego celem jest zweryfikowanie zależności teoretycznej pomiędzy jakością wirtualnego obrazu a rozmieszczeniem kamer. Przyjęto układ 4 kamer (zgodny z modelem systemu — rozdział 4.3.1) ze stałymi położeniami kamer zewnętrznych i zmieniającymi się położeniami 2 kamer wewnętrznych, przedstawiony na rysunku 4.7.



Rysunek 4.7: Schemat eksperymentu ze zmianą ustawienia kamer [Sta+18a]

Odległość pomiędzy kamerami w każdej z par zmienia się od 1 (równomierne rozmieszczenie wszystkich kamer — kamery w pozycjach 0, 1, 2 oraz 3) do 0,1 (rozmieszczenie kamer z małą odległością w parach — kamery w pozycjach 0; 0,1; 2,9; 3). Eksperyment przeprowadzono na zestawie 11 wielowidokowych sekwencji testowych używanych przez grupę MPEG. Sekwencje te posiadają liniowe lub łukowe ustawienie co najmniej 31 kamer. Sekwencje z łukowym ustawieniem są to takie, w których osie optyczne kamer przecinają się w jednym punkcie, ale kąt wycinka łuku pomiędzy skrajnymi kamerami nie przekracza 15°. Są one zbliżone do systemów

liniowych, co pozwala sprawdzić wyprowadzoną zależność teoretyczną dla takich sekwencji oraz zwiększa ilość dostępnego materiału, na którym można przeprowadzić eksperyment. Zostały one zawarte w tabeli 4.1, a ich parametry są przedstawione w rozdziale 3.1.

Tablica 4.1: Procentowa liczba przysłoniętych próbek $OCC_u \cdot 100\%$ oraz zysk z ustawiania kamer w pary $\Delta_{p-u}\text{PSNR}[dB]$ w sekwencjach testowych [Sta+18a]

Identyfikator	Nazwa sekwencji	$OCC_u \cdot 100\%$	$\Delta_{p-u}\text{PSNR}[dB]$	Wykorzystane widoki	Ustawienie kamer
s1	BBB Rabbit Arc [Kov+15]	4,93	-1,33	6, 7... 36	Łukowe
s2	BBB Butterfly Arc [Kov+15]	9,05	-1,05	6, 7... 36	Łukowe
s3	Dog [TFF08]	9,81	-0,29	0, 2... 60	Liniowe
s4	BBB Rabbit Linear [Kov+15]	15,41	-0,19	0, 3... 90	Liniowe
s5	Pantomime [TFF08]	15,61	-0,71	0, 2... 60	Liniowe
s6	BBB Butterfly Linear [Kov+15]	16,53	-1,38	30, 32... 90	Liniowe
s7	BBB Flowers Linear [Kov+15]	29,18	0,73	30, 32... 90	Liniowe
s8	San Miguel [Goo+14]	29,21	1,04	60, 61... 90	Liniowe
s9	Champagne [TFF08]	32,55	1,56	30, 31... 60	Liniowe
s10	Bee [Sen+14]	35,57	1,11	20, 23... 110	Liniowe
s11	BBB Flowers Arc [Kov+15]	38,68	2,12	6, 7... 36	Łukowe

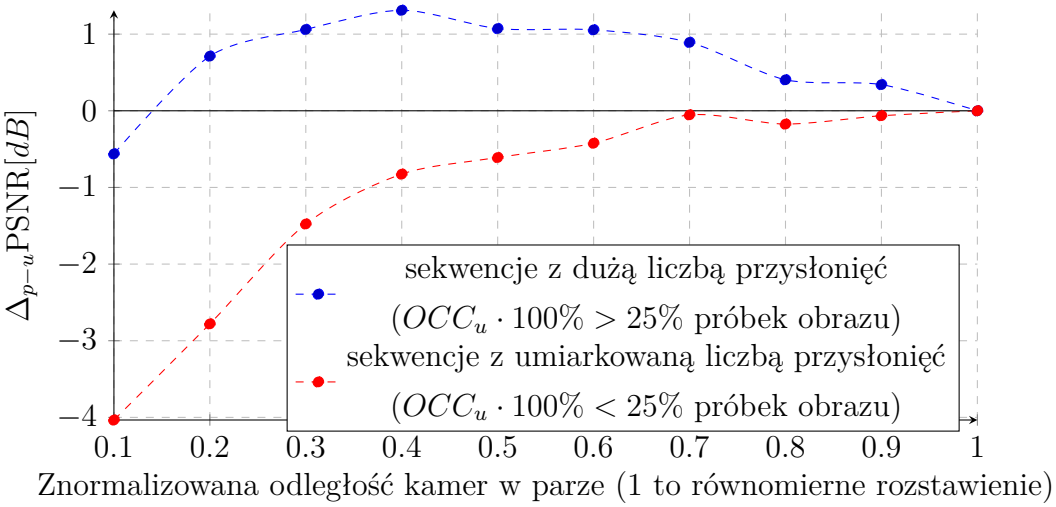
W eksperymencie wirtualne widoki zostały zsyntezowane z wykorzystaniem map głębi estymowanych za pomocą różnych układów 4 kamer. Jakość wirtualnego obrazu została wyznaczona na podstawie średniej wartości metryki PSNR wyznaczonej dla luminancji wszystkich wirtualnych widoków syntezowanych z wykorzystaniem dwóch skrajnych kamer. Taka synteza zapewnia wyłącznie ocenę jakości wirtualnych widoków ze względu na te same pozycje, czyli bez uwzględnienia wpływu zmieniających się odległości pomiędzy kamerami w badanych ustawieniach. Przebieg eksperymentu został przedstawiony na rysunku 4.7. Widokami odniesienia były rzeczywiste zarejestrowane widoki. W każdej sekwencji obliczono średnią liczbę przysłoniętych próbek OCC_u dla równomiernego rozmieszczenia kamer (tabela 4.1). Zostało to zrealizowane poprzez modyfikację oprogramowania syntezy i wyznaczenie liczby próbek, które musiały zostać wypełnione na podstawie sąsiednich. Przyjęto, że sekwencje z nieznacznymi przysłonięciami to takie, w których liczba przysłoniętych próbek nie przekracza 25% ($OCC_u < 0,25$), a sekwencje z istotnymi przysłonięciami posiadają większy udział wypełnianych próbek ($OCC_u > 0,25$). Krzywa teoretyczna (wykres 4.6) ma wartość $\Delta_{p-u}\text{PSNR}$ w przybliżeniu równą 0 $[dB]$ dla takiej wartości przysłoniętych próbek.

4.8 Wyniki eksperymentalne

Wyniki części eksperymentalnej zostały przedstawione na wykresach 4.9, 4.8 i w tabelach 4.1 oraz 4.2. Sekwencje z nieznaczną liczbą przysłoniętych próbek ($OCC_u < 0,25$) posiadają ujemny średni przyrost jakości Δ_{p-u}PSNR dla większości nierównomiernych rozmieszczeń kamer. Uśredniając wyniki dla wszystkich takich sekwencji, najlepszym ustawieniem będzie równomierne rozmieszczenie kamer (rysunek 4.8). Szczegółowe wyniki przedstawiono w tabeli 4.2.

Tablica 4.2: Jakość syntezy dla różnych ustawień kamer w sekwencjach testowych (kolor zielony oznacza największe wartości) [Sta+18a]

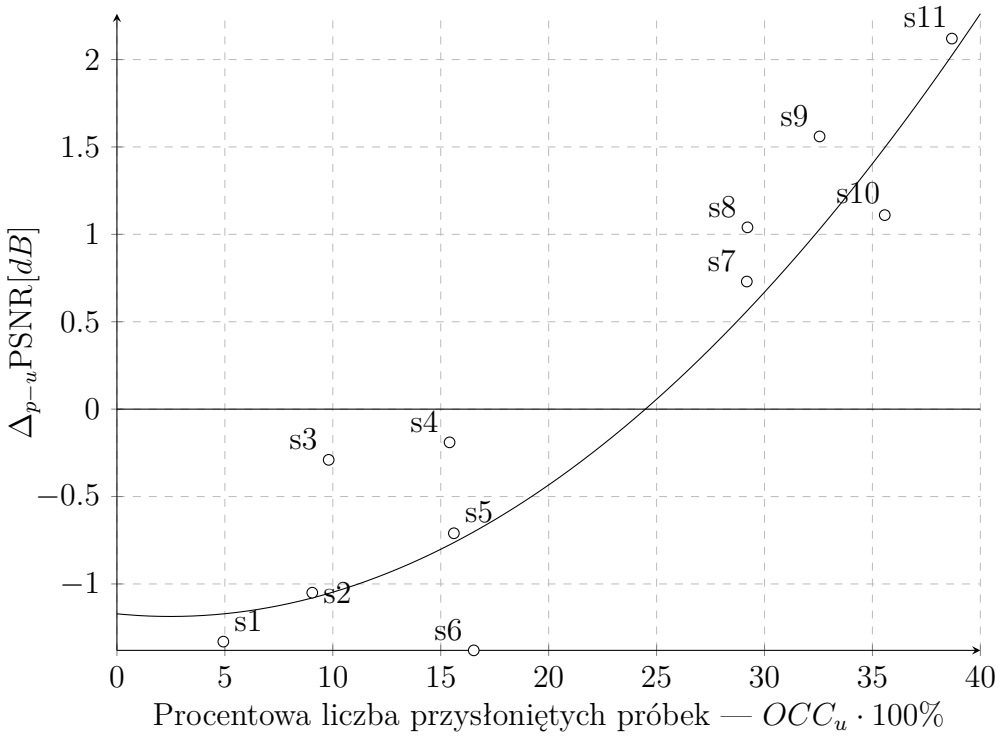
Nazwa sekwencji	Średni PSNR dla syntezowanych widoków									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
BBB Rabbit Arc [Kov+15]	26,81	26,78	26,93	27,10	27,44	27,74	27,90	28,09	28,28	**28,43**
BBB Butterfly Arc [Kov+15]	26,69	27,52	30,12	31,10	30,95	31,20	31,55	31,94	31,83	**32,15**
Dog [TFF08]	24,44	24,58	25,01	25,48	25,52	25,67	**26,12**	25,60	25,37	25,77
BBB Rabbit Linear [Kov+15]	21,55	22,66	23,29	23,46	23,61	23,75	**24,31**	23,57	23,92	23,65
Pantomime [TFF08]	19,85	22,13	24,17	26,03	26,53	26,35	26,58	26,41	**26,79**	26,74
BBB Butterfly Linear [Kov+15]	22,51	25,71	27,68	27,92	28,36	28,81	29,28	29,40	**29,48**	29,31
BBB Flowers Linear [Kov+15]	21,41	23,06	24,02	24,42	24,32	24,43	**24,45**	24,28	24,01	23,69
San Miguel [Goo+14]	23,74	24,47	24,74	**25,14**	24,83	25,09	24,83	24,40	24,39	24,10
Champagne [TFF08]	17,43	18,91	18,85	**19,21**	18,71	19,02	18,40	17,80	18,14	17,65
Bee [Sen+14]	19,16	20,68	20,95	**21,15**	21,14	21,01	20,82	20,67	20,43	20,04
BBB Flowers Arc [Kov+15]	20,03	21,03	**21,33**	21,23	20,95	20,33	20,55	19,46	19,33	19,11
Średni PSNR	22,15	23,41	24,28	24,75	24,76	24,85	24,98	24,69	24,72	24,60



Rysunek 4.8: Średni zysk jakości dla różnych odległości bazowych [Sta+18a]

W przypadku sekwencji z dużą liczbą przysłoniętych próbek ($OCC_u > 0,25$) ustawianie kamer w pary pozwala na uzyskanie lepszej jakości map głębi, a tym samym — lepszej jakości wirtualnych widoków. Uśredniając wyniki dla wszystkich takich sekwencji, największy zysk otrzymano dla znormalizowanej odległości pomiędzy optycznymi środkami kamer w parze w odległości równej 0,4 (rysunek 4.8).

W systemach wielokamerowych charakteryzujących się mniejszą odległością pomiędzy kamerami w parze zaobserwowano spadek jakości. Wynika to prawdopodobnie ze zbyt małej dokładności mapy głębi. Spadek jakości dla systemu z większą odległością pomiędzy optycznymi środkami w parze kamer prawdopodobnie powoduje wzrost liczby przysłoniętych próbek w syntezowanych widokach oraz spadek jakości.



Rysunek 4.9: Krzywa zysku parowania kamer w funkcji procentowego udziału przysłonięć. Punkty s_i reprezentują sekwencje wymienione z tabeli 4.1 [Sta+18a]

Na wykresie 4.9 pokazano eksperymentalnie obliczone zyski ustawienia kamer w pary Δ_{p-u}PSNR dla wielowidokowych sekwencji testowych wraz z krzywą teoretyczną, co potwierdza poprawność przedstawionego modelu teoretycznego. Krzywa, która jest rezultatem części teoretycznej, została zestawiona z wynikami eksperymentalnymi reprezentowanymi przez punkty s_i. Sekwencje te są reprezentatywne dla przyszłych potencjalnych zastosowań w systemach swobodnego punktu widzenia.

W związku z tym można stwierdzić, że krzywa przedstawiona na rysunku 4.9 stanowi przybliżony szacunek zysków wynikających z ustawienia kamer w pary w systemach swobodnego punktu widzenia.

4.9 Podsumowanie

W niniejszym rozdziale założono model systemu wielokamerowego z przysłonięciami oraz wyprowadzono przybliżony model jakości syntezowanego widoku. Założenia te pozwoliły na wyprowadzenie teoretycznych matematycznych wzorów głównych zjawisk wpływających na jakość syntezowanego widoku. W rozdziałach 4.4 oraz 4.5 rozpatrzono odpowiednio wpływ odległości pomiędzy kamerami i przysłonięć w scenie na jakość wirtualnego widoku. Pozwoliło to na uzyskanie teoretycznych wyników prezentujących, jak liczba przysłonięć w scenie oraz zmiana odległości pomiędzy kamerami wpływa na średnią jakość syntezowanych widoków. Wyniki modelu, o którym mowa, zostały zaprezentowane za pomocą krzywej przedstawiającej zysk ze stosowania par kamer w funkcji procentowego udziału przysłoniętych próbek w scenie (rozdział 4.6, wykres 4.6). Model ten porównano z wynikami eksperymentu, który został wyznaczony z wykorzystaniem wielowidokowych sekwencji testowych (rozdział 4.7). Krzywą teoretyczną i eksperymentalna została przedstawiona na wykresie 4.9. Uzyskane wyniki pozwoliły na potwierdzenie poprawności modelowania teoretycznego i wyciągnięcie wniosków.

Z punktu widzenia praktycznych zastosowań przeprowadzone badania wskazują, że umieszczanie kamer w parach daje zysk jakości w przypadku, kiedy w scenie występuje wiele przysłonięć. Zaobserwowana zależność, poparta wcześniejszym rozważaniem teoretycznym, ma istotne znaczenie, ponieważ powoduje zmianę średniej jakości wirtualnych widoków do 3 dB dla metryki PSNR. Wykazano, że pary kamer należy stosować, kiedy procentowa liczba przysłoniętych próbek obrazu w syntezowanych widokach dla kamer rozstawionych równomiernie przekroczy około 25%.

Przedstawione w niniejszym rozdziale badania dotyczące optymalizacji ustawienia kamer zostały opublikowane we współautorskim artykule [Sta+18a]. Główny udział autora dotyczy opracowania zależności metryki PSNR od liczby przysłonięć oraz opracowania i przeprowadzenia części eksperymentalnej. W badaniach przedstawionych w rozdziale 4.4 autor nie miał wyłącznego udziału — zostały one zawarte w pracy, dla kompletnej prezentacji wyników i wniosków, będącej osiągnięciem zespołu.

5 Wpływ kompresji na jakość wirtualnego widoku

5.1 Opis problemu

W tym rozdziale przedstawiono badania eksperymentalne mające na celu określenie wpływu stratnej kompresji widoków realizowanej przed procesem estymacji map głębi na jakość swobodnej nawigacji. Badania te mogą dotyczyć łączy A oraz C, scharakteryzowanych w rozdziale 2.2.

Podejście, które często wykorzystywano w starszych pracach dotyczących systemów swobodnego punktu widzenia, to uniknięcie potencjalnego wpływu zakłóceń kodowania stratnego i wykorzystanie danych bezstratnych [Fuj+06][Dom+14a] [Fis+20][Duf17]. Wielokamerowy system akwizycji musi zatem rejestrować obrazy nie poddane kompresji lub zakodowane bezstratne, co było jednym z założeń budowy wielu tego typu systemów (rozdział 7.2.2). W praktycznych systemach realizacja transmisji w formacie bezstratnym jest trudna, w szczególności łącza C dostarczającego dane do terminala użytkownika, przy zachowaniu racjonalnych prędkości bitowych, zwykle pomiędzy 5 a 50 Mbit/s [Boy+21][Fis+20]. Łącze to może być realizowane zgodnie ze schematem C2 (rysunek 2.2, strona 22). Wtedy wszystkie lub wybrane widoki są dostarczane do terminala użytkownika. Taka transmisja może nie obejmować map głębi, jednak wówczas do ich estymacji wykorzystane są obrazy poddane kompresji stratnej. Realizacja wiąże się z dodatkowymi obliczeniami, które powinny być zrealizowane w czasie rzeczywistym w terminalu użytkownika. Jednak pozwala ona na zmniejszenie wymaganej przepływności, ponieważ transmisja map głębi nie jest potrzebna. Istnieje wiele naukowych badań przedstawiających odpowiedni udział liczby bitów niosących informację o widokach do bitów niosących informację o mapie głębi. Może on wynosić np. około 85/15 [KWD14][MFW07][CVO11], 5/1 [Feh04], być dobrany na podstawie nieliniowego modelowania w zależności np. od techniki kompresji czy jej stopnia [AOG20][DAOG21][AO23][Kli12]. W każdym wypadku brak transmisji map głębi sprawia, że wymagana przepływność do terminala użytkownika jest mniejsza, ale musi on realizować dodatkowe obliczenia w czasie rzeczywistym — estymację map głębi.

W literaturze można znaleźć algorytmy estymacji głębi (rozdział 2.5) oraz ich realizacje działające w czasie rzeczywistym (np. [SJ15],[DTL19], jednak prezentowane rozwiązania mają wiele ograniczeń i dodatkowych wymagań, jak również jakość

estymowanej przez nie mapy głębi zazwyczaj jest znacznie gorsza w porównaniu z bardziej złożonymi algorytmami [Sta+13]. Można przewidywać, że stale rosnąca wydajność sprzętu oraz rozwój implementacji w najbliższym czasie skrócą czas estymacji mapy głębi przy zachowaniu wysokiej jakości. Rozwiązaniem umożliwiającym zmniejszenie złożoności obliczeniowej po stronie dekodera jest również wykorzystanie profilu GA (ang. Geometry Absent) w technice MIV, opracowanej przez ekspertów grupy roboczej ISO/IEC MPEG [Boy+21]. Strumień danych w tym profilu może zawierać metadane wykorzystywane przez algorytm estymacji głębi w dekoderze. Dane te pozwalają na zmniejszenie obszaru przeszukiwań i znacząco zmniejszają złożoność obliczeniową. Rozwinięcie tego profilu, zwiększające ilość metadanych, które można przesłać w strumieniu, to profil DSDE (ang. Decoder-Site Depth Estimation), który jest opracowywany w ramach kolejnej edycjach normy MIV [Mie+23a].

Zastosowanie stratnie skompresowanych widoków w estymacji map głębi ma istotną zaletę — zapewnia niższy koszt systemu FTV. W takim przypadku system akwizycji (łącze A) może składać się z prostszych, czyli np. nieprofesjonalnych, konsumenckich kamer (w przeciwieństwie do systemów bez kompresji wizji, np. [Dom+15d]), tańszego okablowania (niższe prędkości bitowe), mniej wydajnych jednostek zbierających i przetwarzających dane. Rozwiązanie to może też zapewniać mniejszą prędkość bitową przesyłaną przez łącze C, gdyż transmitując jedynie widoki do terminala użytkownika, niepotrzebnie nie przesyłamy mapy głębi, która może zostać wyznaczona w odbiorniku [Gar+22].

Cel przedstawionych badań to udowodnienie, że sekwencja wielowidokowa może być skompresowana przed procesem estymacji map głębi, a także zbadanie wpływu tej kompresji na jakość wirtualnego widoku. Autor przeprowadza eksperymenty naukowe za pomocą wielowidokowych sekwencji testowych.

5.2 Opis eksperymentów

W celu oceny efektywności różnych koderów wizji w systemach z estymacją głębi po stronie dekodera przeprowadzono eksperymenty dla różnych wariantów kodowania, które zostały przedstawione w tabeli 5.1. We wszystkich badaniach wirtualne widoki są generowane na podstawie stratnie skompresowanej sekwencji wielowidokowej. Ocena jakości odbywa się z użyciem metryki PSNR lub IV-PSNR przedstawionych w rozdziałach 3.2 oraz 3.3.

Pierwszy eksperyment nazwano **eksperymentem wstępnym**. Został on przeprowadzony na grupie wielowidokowych sekwencji dla kompresji wybranych widoków koderami pojedynczego widoku, takimi jak MPEG-2, AVC oraz HEVC. Wybrane źródłowe widoki są kodowane oddzielnie i używane jako dane wejściowe dla oprogramowania estymacji głębi oraz syntezy wirtualnych widoków. Ocena jakości jest dokonywana z wykorzystaniem metryki PSNR, wyznaczonej na podstawie wirtualnych widoków względem widoków sekwencji niewykorzystane w procesie estymacji map głębi i syntezy wirtualnych widoków.

Druga grupa badań, opisana w tym rozdziale na potrzeby rozprawy, została nazwana **eksperymentami zasadniczymi**. Wykorzystuje ona nowsze techniki kompresji oraz nową metodę oceny jakości widoków syntezowanych za pomocą metryki IV-PSNR. Pierwszy z eksperymentów zasadniczych to **eksperyment ogólny**. Ma on podobny przebieg do eksperymentu wstępnego, jednak zastosowano w nim kilka modyfikacji. Uwzględniono wnioski wynikające z wyników eksperymentu wstępnego oraz zastosowano nowsze narzędzia estymacji głębi i syntezy wirtualnych widoków. Drugi z eksperymentów zasadniczych to **eksperyment MIV**. Wykorzystuje technikę kodowania opisaną w normie ISO/IEC dla kompresji wszechogarniających treści wizyjnych MPEG Immersive Video (MIV).

Tablica 5.1: Porównanie eksperymentu wstępnego oraz zasadniczych: ogólnego i MIV

Nazwa eksperymentu		Sekwencje	Ocena jakości
Eksperyment wstępny		Wielowidokowe, tablica 5.2	PSNR
Eksperymenty zasadnicze	ogólny	Wielowidokowe, tablica 5.2	PSNR, IV-PSNR
	MIV	MIV, tablica 5.10	PSNR, IV-PSNR

Tablica 5.2: Eksperyment wstępny oraz ogólny, wielowidokowe sekwencje testowe

Nazwa sekwencji	Wykorzystane widoki	Ustawienie kamer
BBB Flowers[Kov+15]	30, 32,... 90	Liniowe
BBB Flowers*[Kov+15]	6, 7,... 36	Łukowe
BBB Butterfly[Kov+15]	30, 32,... 90	Liniowe
Pantomime[TFF08]	0, 2,... 60	Liniowe
BBB Butterfly[Kov+15]	6, 7,... 36	Łukowe
Dog[TFF08]	0, 2,... 60	Liniowe
BBB Rabbit[Kov+15]	0, 3,... 90	Liniowe
BBB Rabbit[Kov+15]	6, 7,... 36	Łukowe

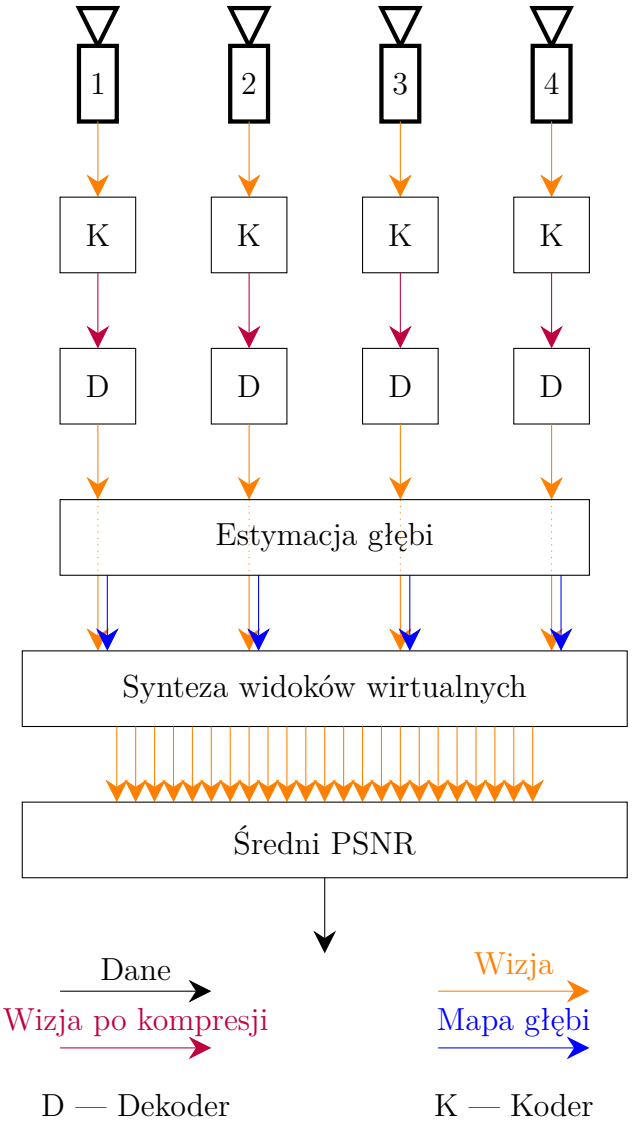
5.3 Eksperyment wstępny

5.3.1 Cel eksperymentu wstępnego

Wyniki eksperymentu wstępnego zostały opublikowane w 2016 roku [Dzi+16a]. Większość ówczesnych prac, które dotyczyły systemów swobodnej nawigacji, wykorzystywała bezstratne dane. Dlatego celem przeprowadzonego eksperymentu jest sprawdzenie, jak kompresja stratna przed procesem estymacji map głębi wpływa na jakość wirtualnego widoku. Ma on udowodnić, że możliwe jest uzyskanie postrzeganej wysokiej jakości w systemach swobodnej nawigacji, które wykorzystują stratną kompresję. Podczas testów odniesieniem była jakość uzyskiwana w systemach bez stratnej kompresji. Sprawdzono działanie popularnych technik kodowania oraz przeanalizowano, jak zmienia się jakość syntezowanego widoku uzyskanego z różnych typów zakodowanych ramek.

5.3.2 Opis eksperymentu wstępnego

Potok przetwarzania wielowidokowej wizji, który został wykorzystany w eksperymencie wstępnym, jest przedstawiony na rysunku 5.1. Wybrane widoki wejściowe są osobno kodowane za pomocą różnych technik. Wizja po dekodowaniu jest użyta do estymacji map głębi oraz syntezy wirtualnych widoków. Widoki te są syntezowane zgodnie z parametrami rzeczywistych widoków, które nie zostały wykorzystane. Umożliwia to wyznaczenie jakości systemu za pomocą średniej wartość metryki PSNR dla takich par widoków.



Rysunek 5.1: Eksperyment wstępny oraz ogólny, schemat eksperymentu kompresji danych

Rysunek 5.2: Eksperyment wstępny oraz ogólny: 31 widoków sekwencji wykorzystanych do estymacji map głębi (kolor niebieski) lub wyznaczenia metryk oceny jakości (kolor pomarańczowy)

W eksperymencie wstępnym wykorzystano sekwencje wielowidokowe z dostępnymi 31 widokami. Dla każdej sekwencji założono użycie czterech widoków do estymacji map głębi i syntezy wirtualnych widoków, a pozostałe 27 widoków źródłowych (rys. 5.2) zostało wykorzystanych jako obraz odniesienia dla wyznaczania wartości metryki PSNR. Jakość systemu swobodnej nawigacji została określona przez średnią wartość dla wszystkich wirtualnych widoków. W eksperymencie wstępnym użyto 8 sekwencji wielowidokowych: 6 sztucznych i 2 naturalnych, które zarejestrowano przez system wielokamerowy [Tan+12]. Wykorzystane sekwencje oraz widoki zostały przedstawione w tabeli 5.2. Sekwencja, oznaczona w tej tabeli gwiazdką (*), została pominięta w eksperymencie wstępnym z uwagi na bardzo niską jakość syntezowanych wirtualnych widoków. Dotyczyło to także przetwarzania bez kompresji, na podstawie oryginalnych danych, co jest rezultatem niepoprawne działającego oprogramowania. Dla każdej sekwencji założono, że widoki użyte przez oprogramowanie do estymacji map głębi, będą rozmieszczone równomiernie, co przedstawia 5.2. Dla sekwencji BBB Butterfly o ustawieniu łukowym wykorzystane widoki to 6, 19, 32 i 45. Jest to ustawienie sugerowane również przez grupy badawcze ISO/IEC MPEG dla tej sekwencji [TFF08]. Wybór widoków dla innych sekwencji w sposób, który zapewnia porównywalne różnice w skrajnych widokach, został zaprezentowany w tabeli 5.2.

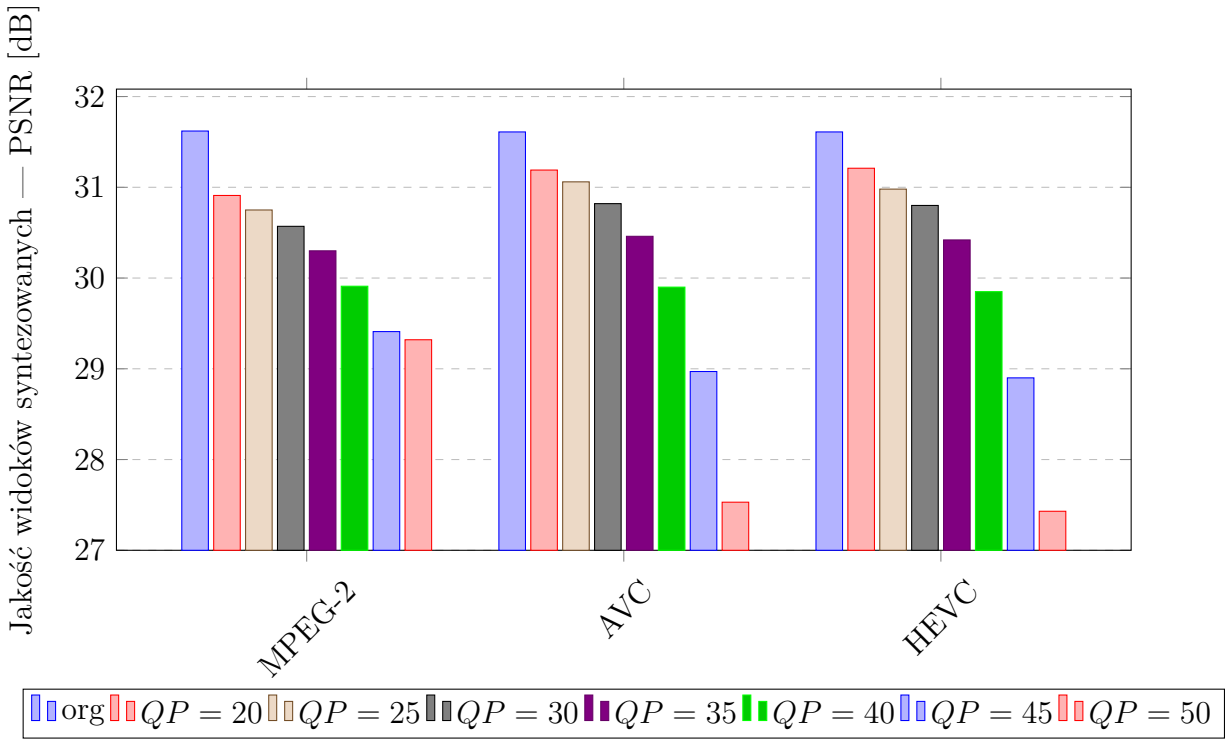
Eksperyment przeprowadzono dla niezależnej kompresji każdego z widoków z wykorzystaniem technik MPEG-2, AVC i HEVC. Długość grupy obrazów (GOP) dla każdego kodera wynosiła 13. Układ ramek w takiej grupie był następujący:
I BB P BB P BB P BB P BB P.
Podczas eksperymentu wykorzystano powszechnie dostępne zoptymalizowane implementacje technik kompresji pojedynczego widoku: dla MPEG2 — mpeg2video z pakietu FFmpeg [wFFmpeg], dla AVC — x264 [wX264], a dla HEVC — x265 [wX265]. W wszystkich koderach zastosowano zestaw ustawień do pracy szybkiej (ang. fast preset) w celu przybliżenia konfiguracji, która może zostać użyta w urządzeniach o małej mocy obliczeniowej. Zastosowano zamkniętą grupę obrazów (ang. GOP - group of pictures) bez możliwości predykcji z kolejnej grupy obrazów.

Eksperymenty przeprowadzono dla 8 różnych wariantów kodowania, na które składa się 7 wartości parametru kwantyzacji i widoki oryginalne bez kodowania stratnego. Umożliwiło to uzyskanie średniej wartości metryki PSNR obrazów syntezowanych, która była jakością odniesienia. Dla każdej sekwencji zakodowano 100 ramek obrazów i użyto ich do estymacji głębi i syntezy wirtualnych widoków. Wykorzystano

narzędzia rekomendowane przez grupy robocze ISO/IEC MPEG [Sta+13][Dzi+16c]. Jakość syntezowanych widoków mierzono poprzez oszacowanie średniej wartości metryki oceny jakości PSNR dla 27 syntezowanych wirtualnych widoków względem widoków odniesienia. Podejście to zapewnia pomiar jakości w całym systemie swobodnej nawigacji, czyli wszystkich etapów przetwarzania (kodowania, estymacji głębi, syntezy wirtualnych widoków).

5.3.3 Wyniki eksperymentu wstępnego

W tabelach 5.3, 5.4, 5.5 na stronie 69 przedstawiono wartości uzyskane podczas eksperymentu wstępnego. Każda tabela zawiera średnią wartość metryki PSNR syntezowanych widoków dla badanych wartości współczynnika kwantyzacji QP oraz wszystkich sekwencji testowych. Dla techniki kompresji MPEG2, zamiast QP od 20 do 50, wybrano wartości parametru kwantyzacji tak, aby zapewnić podobny rozkład jakości dla wszystkich koderów. Dokładne wartości parametru Q znajdują się w tabeli 5.3.



Rysunek 5.3: Średnia wartość PSNR wirtualnych widoków dla różnych wartości parametru QP w koderach oraz dla sekwencji oryginalnej (org)

Tablica 5.3: Eksperyment wstępny: średnie wartości PSNR widoku syntezowanego dla kodera MPEG2

Nazwa sekwencji	Ustawienie kamer	PSNR [dB]	PSNR [dB] dla różnych Q						
		bez kompresji	2	4	6	10	18	30	51
BBB Butterfly	Łukowe	36,9	35,0	34,8	34,5	34,2	33,8	33,3	33,2
BBB Butterfly	Liniowe	35,7	34,7	34,5	34,3	34,1	33,8	33,3	33,2
Dog	Liniowe	30,0	29,9	29,8	29,7	29,6	29,5	29,4	29,3
BBB Flowers	Liniowe	27,5	26,8	26,6	26,5	26,3	25,9	25,6	25,4
Pantomime	Liniowe	30,3	29,9	30,0	29,9	29,8	29,6	29,4	29,4
BBB Rabbit	Łukowe	31,2	30,9	30,5	30,2	29,8	29,0	28,3	28,2
BBB Rabbit	Liniowe	29,8	29,6	29,4	29,3	28,9	28,3	27,7	27,6

Tablica 5.4: Eksperyment wstępny: średnie wartości PSNR widoku syntezowanego dla kodera AVC

Nazwa sekwencji	Ustawienie kamer	PSNR [dB]	PSNR [dB] dla różnych QP						
		bez kompresji	20	25	30	35	40	45	50
BBB Butterfly	Łukowe	36,9	36,3	36,1	35,7	35,1	34,2	33,0	30,6
BBB Butterfly	Liniowe	35,7	35,4	35,2	34,9	34,4	33,7	32,5	30,4
Dog	Liniowe	30,0	29,5	29,5	29,4	29,2	28,8	28,1	26,9
BBB Flowers	Liniowe	27,5	26,8	26,6	26,5	26,2	25,8	25,3	24,6
Pantomime	Liniowe	30,3	30,0	30,0	29,9	29,9	29,5	28,7	27,6
BBB Rabbit	Łukowe	31,2	30,8	30,6	30,2	29,7	29,0	27,8	26,4
BBB Rabbit	Liniowe	29,8	29,6	29,4	29,2	28,8	28,2	27,3	26,2

Tablica 5.5: Eksperyment wstępny: średnie wartości PSNR widoku syntezowanego dla kodera HEVC

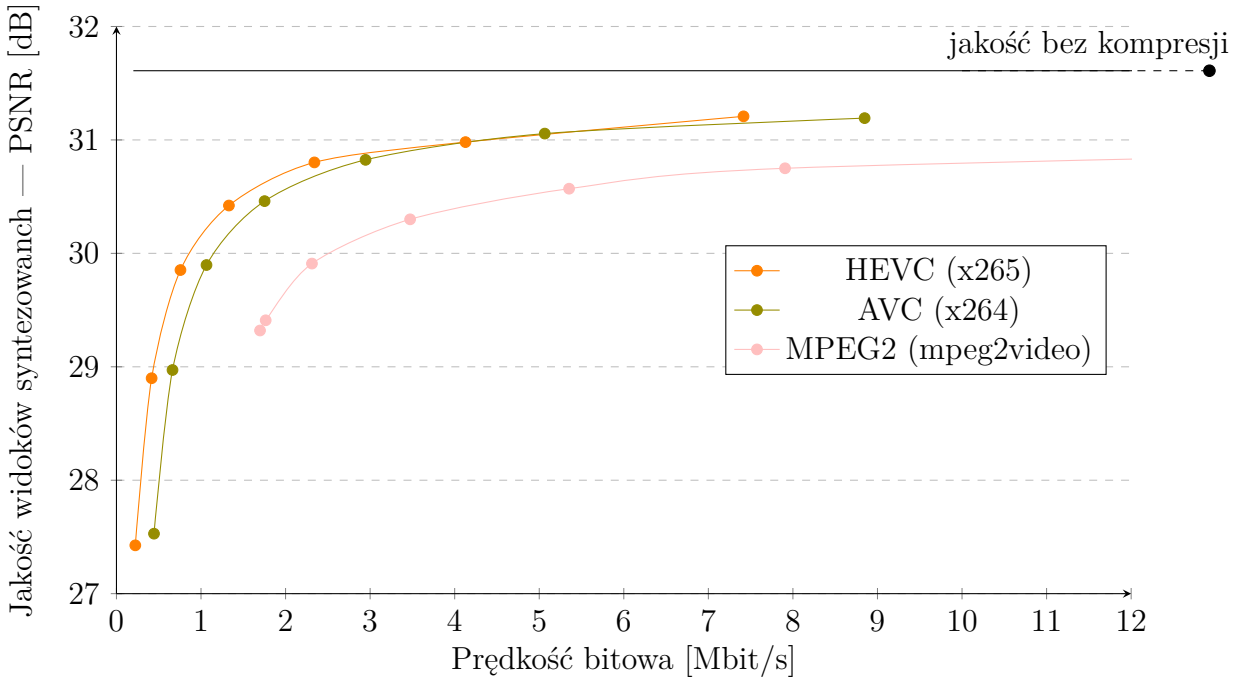
Nazwa sekwencji	Ustawienie kamer	PSNR [dB]	PSNR [dB] dla różnych QP						
		bez kompresji	20	25	30	35	40	45	50
BBB Butterfly	Łukowe	36,9	35,8	35,4	35,1	34,7	33,9	33,1	30,3
BBB Butterfly	Liniowe	35,7	35,4	35,1	34,8	33,9	33,2	31,4	30,1
Dog	Liniowe	30,0	29,5	29,5	29,4	29,2	28,8	28,1	26,9
BBB Flowers	Liniowe	27,5	26,7	26,2	26,2	25,9	25,5	25,1	24,4
Pantomime	Liniowe	30,3	30,0	29,9	29,9	29,7	29,3	28,9	28,1
BBB Rabbit	Łukowe	31,2	31,2	31,1	30,8	30,4	29,5	28,1	26,2
BBB Rabbit	Liniowe	29,8	29,8	29,7	29,5	29,2	28,8	27,5	25,9

Tablica 5.6: Eksperyment wstępny: prędkości bitowe dla wykorzystanych koderów i parametrów kwantyzacji uśrednione dla wszystkich sekwencji

QP	Prędkość bitowa [Mbit/s]			
	dla koderów			bez kompresji
	MPEG2	AVC	HEVC	stratnej
20	16,8	8,8	7,4	
25	7,9	5,1	4,1	
30	5,4	2,9	2,3	
35	3,5	1,7	1,3	1343
40	2,3	1,1	0,8	
45	1,8	0,7	0,4	
50	1,7	0,4	0,2	

Rysunek 5.3 przedstawia wartości PSNR z tabel 5.3, 5.4, 5.5, uśrednione dla wszystkich sekwencji. Prezentowany rozkład jakości przy zmianie wartości QP ma typowy i oczekiwany wykres, mimo że miara została policzona po estymacji głębi oraz po syntezie wirtualnych widoków — dla małej wartości QP jakość jest wysoka, a przy wzroście tej wartości jakość maleje.

Na rysunku 5.3 można zauważyć, iż różnica między jakością wirtualnego obrazu uzyskaną dla danych oryginalnych, czyli bez kompresji stratnej, a jakością wirtualnego obrazu uzyskaną po syntezie z wykorzystaniem danych po kompresji wizji dla $QP = 20$ i 25 jest niewielka i wynosi nieco ponad 0,5 dB. Jest to koszt obniżenia prędkości bitowej kilkaset razy: z ponad 1000 Mbit/s (uśredniona wartość dla wszystkich testowanych sekwencji, zakładając próbki ośmiobitowe, format YC_bC_r 4:2:0 oraz szybkość ramkową 25 kl/s) do kilku Mbit/s (dla wszystkich przesyłanych widoków), co zostało przedstawione w tabeli 5.6.



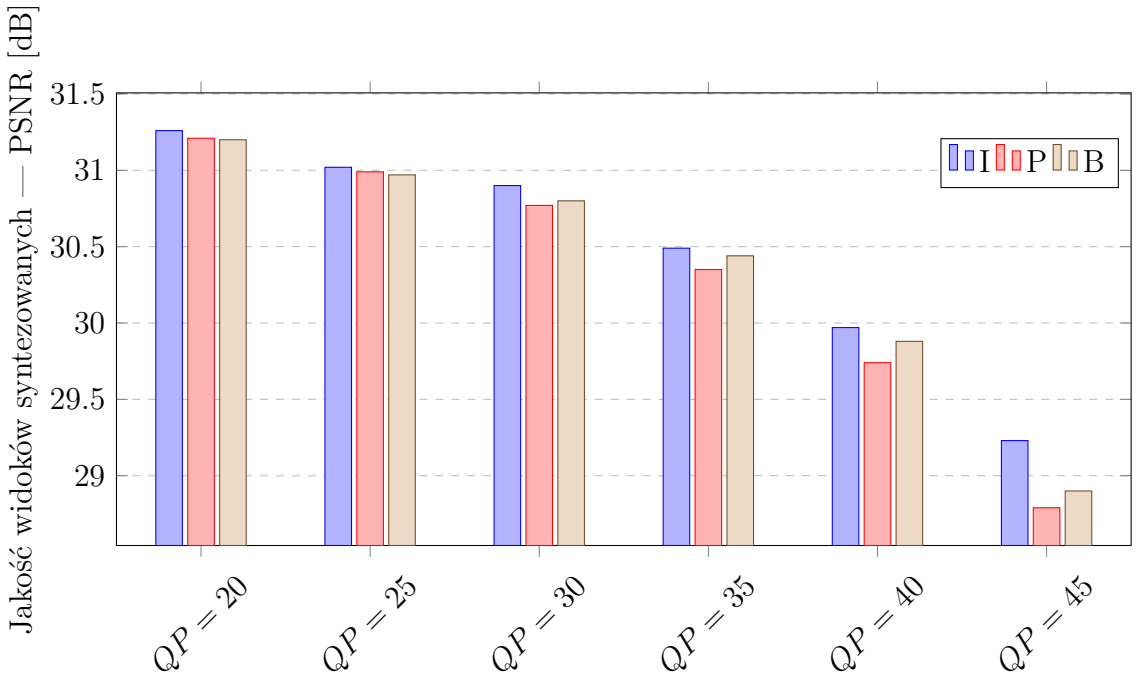
Rysunek 5.4: Eksperyment wstępny: krzywe RD dla koderów MPEG-2, AVC, HEVC

Na rysunku 5.4 przedstawiono krzywe RD dla wszystkich analizowanych koderów. Przerywana linia pozioma reprezentuje jakość widoków syntezowanych przy użyciu nieskompresowanych widoków wejściowych. Jak już wspomniano, we wszystkich eksperymentach zastosowano taki sam układ ramek:

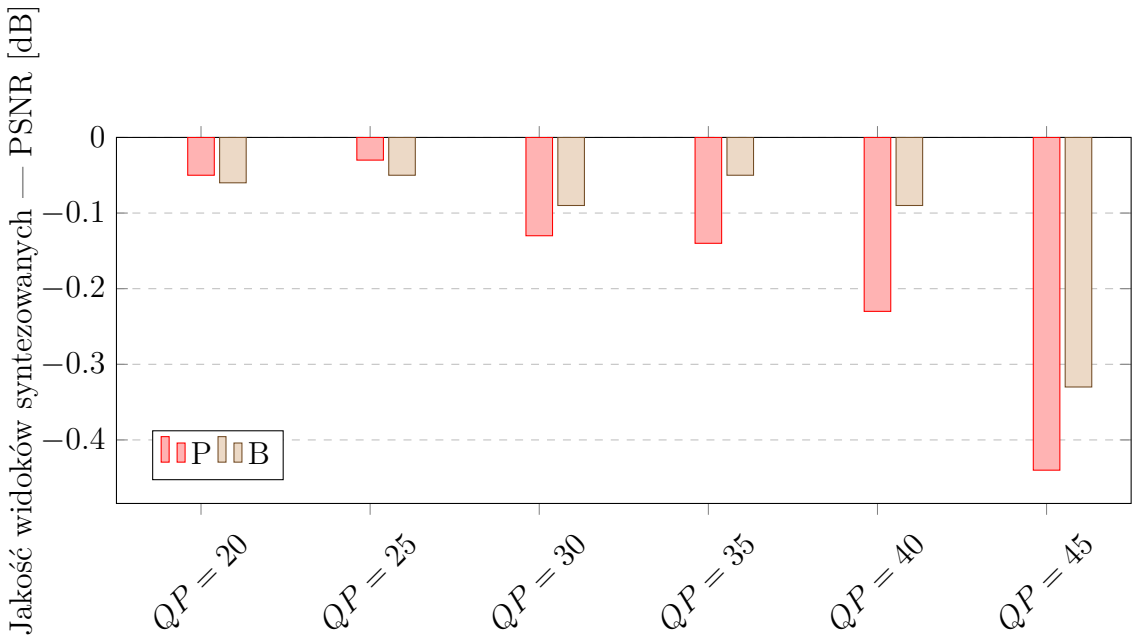
I BB P BB P BB P BB P BB P.

W związku z tym można również przeanalizować wpływ zastosowania różnych typów ramek na jakość syntezowanych widoków. Podczas badań wykorzystano domyślny algorytm alokacji bitów na ramki I,P,B. Dla wszystkich używanych koderów wyniki były podobne, dlatego przedstawiono dokładne rezultaty dla najnowszej wykorzystanej techniki, czyli HEVC. Na rysunku 5.5 przedstawiono średnie PSNR (uśrednione dla wszystkich sekwencji) dla ramek I, P i B.

Największe wartości metryki PSNR uzyskano w ramkach typu I. Jakość ramek z predykcją międzyobrazową jest niewiele gorsza. Informacje o jej stracie w ramkach typu P i B w porównaniu z ramkami typu I przedstawiono na rys. 5.6.

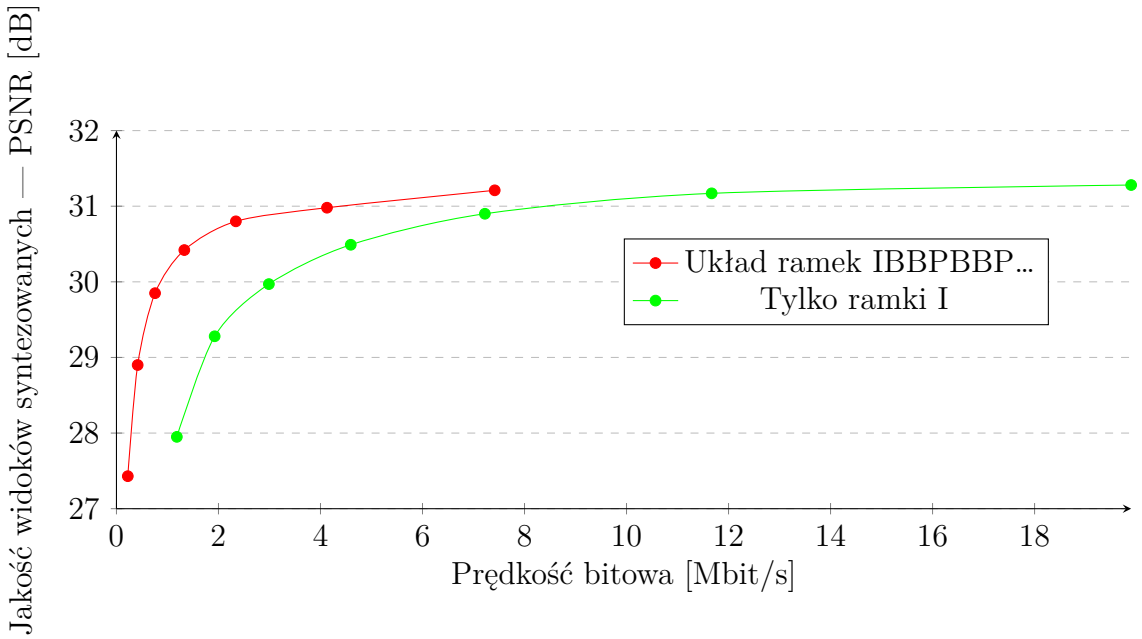


Rysunek 5.5: Eksperyment wstępny: średnia wartość PSNR dla ramek typu I, P, B — koder HEVC



Rysunek 5.6: Eksperyment wstępny: różnica PSNR dla ramek typu P, B w porównaniu do ramek typu I — koder HEVC

Przedstawione dane wyraźnie pokazują, że jakość wirtualnego widoku zsyntezowanego z wykorzystaniem ramek kodowanych wewnątrzobrazowo (typu I), jest większa niż jakość wirtualnego widoku zsyntezowanego z wykorzystaniem ramek z predykcją międzyobrazową. Dla małych wartości QP różnica PSNR dla ramek typu P i B jest bardzo niewielka (tylko $0,04\ dB$ dla $QP = 20$ i 25 oraz $0,07\ dB$ dla $QP = 30$). W przypadku większej wartości parametru QP różnica między ramkami kodowanymi wewnątrzobrazowo i międzyobrazowo zaczyna być większa. Można również dostrzec, że dla wyższych wartości współczynnika kwantowania, ramki typu B wykazują mniejszą różnicę względem ramek I. Są one widoczne, ale nie przekraczają wartości 0,5 dB. Poza tym dla małych wartości współczynnika kwantyzacji ($QP = 20$ oraz $QP = 25$) zmiana jakości pomiędzy ramkami I,P oraz B nie przekracza wartości 0,1 dB. W eksperymencie nie ingerowano w algorytm alokacji bitów w ramkach I,P,B.



Rysunek 5.7: Eksperyment wstępny: krzywe RD dla kodera HEVC z różnym układem ramek

Na rysunku 5.7 porównano krzywe RD dla kodera HEVC wyłącznie w trybie ramek typu I oraz z układem ramek analizowanym podczas eksperymentu. Wyniki potwierdzają, że dla stałego kroku kwantowania ramki typu I wprowadzają mniejsze zniekształcenia niż ramki z predykcją międzyobrazową. Odbywa się to jednak kosztem znacznie większej prędkości bitowej.

5.3.4 Wnioski

W eksperymencie wstępnym przetestowano wpływ stratnej kompresji widoków wejściowych na jakość wirtualnego widoku. Celem badań było sprawdzenie, czy można stosować kompresję stratną przed estymacją map głębi w systemach o postrzeganej wysokiej jakości. Przeprowadzony eksperyment wykazał, że zastosowanie kompresji z małą wartością współczynnika QP nie obniża znacząco jakości syntezowanych wirtualnych widoków. Dla wszystkich testowanych koderów (MPEG2, AVC i HEVC) różnica między jakością wirtualnego widoku syntezowanego przy użyciu nieskompresowanej i skompresowanej wizji nie była większa niż 1 dB, nawet dla wartości parametru kwantyzacji $QP = 30$. Co więcej, dla mniejszych wartości QP różnica była mniejsza, np. dla $QP = 20$ wynosiła około 0,5 dB (zarówno dla koderów wizji AVC, jak i HEVC). Jakość syntezy wirtualnego widoku była najlepsza dla kodera HEVC. Poza tym wykazano, że najwyższa jakość tego widoku została uzyskana ze zdekodowanych ramek typu I, ale różnice w jakości widoku syntezowanego dla ramek P i B są pomijanie małe, tzn. mniejsze niż 0,1 dB dla małych wartości współczynnika kwantyzacji ($QP = 20$, $QP = 25$). Wyniki wskazują, że nagrana sekwencja wielowidokowa może zostać skompresowana. Zapewni to kilkusetkrotne zmniejszenie wymaganej prędkości bitowej oraz zmianę jakości wirtualnego widoku nie większą niż 0,5 dB przy założeniu, że jakość przesyłanych widoków ma mały współczynnik kwantyzacji. Wykazano, że taka kwantyzacja nie obniża jakości w ramkach z predykcją międzyobrazową. W związku z tym kamery wielowidokowych systemów wizyjnych nie muszą zapewniać bezstratnie zakodowanego strumienia wizyjnego. Pozwala to obniżyć koszty systemu akwizycji, powiększając liczbę kamer, które mogą zostać wykorzystywane w badanych systemach.

5.4 Eksperymenty zasadnicze

5.4.1 Cel eksperymentów zasadniczych

Wnioski wyciągnięte z eksperymentu wstępnego pozwoliły na przeprowadzenie dwóch eksperymentów zasadniczych. Ich wyniki zostały opublikowane w 2022 roku [Grz+22], czyli 6 lat po eksperymencie wstępnym. Celem badań było wykorzystanie powstałej techniki kompresji wizji wszechkierunkowej MPEG Immersive Video (MIV) i analiza wpływu jakości widoku przesyłanego na wirtualny widok. Zastosowano w nich nowsze techniki kompresji stratnej, estymacji głębi i syntezy wirtualnych

widoków oraz zrezygnowano (eksperyment MIV) z wariantu bez kodowania stratnego. Podczas eksperymentu wstępnego wykazano że strata jakości w ramkach typu P i B jest pomijanie mała dla wysokiej jakości przesyłanego obrazu, dlatego w kolejnych badaniach nie ingerowano w układ ramek wykorzystywany przez kodery (ang. GOP — group of pictures). W odróżnieniu od eksperymentu wstępnego do pomiaru jakości wirtualnego widoku wykorzystano również metrykę IV-PSNR, która jest nową miarą dostosowaną do widoków syntezowanych [Dzi+22]. Pierwszy eksperyment przeprowadzono z użyciem wielowidokowych sekwencji testowych, drugi natomiast wykorzystywał technikę kodowania reprezentacji wszechkierunkowej MIV w trybie GA (ang. Geometry Absent) i został przeprowadzony zgodnie z warunkami testów MIV CTC [00620].

5.4.2 Opis eksperymentu ogólnego

Przebieg eksperymentu ogólnego jest podobny do przebiegu eksperymentu wstępnego. Wykorzystano te same widoki oraz sekwencje wielowidokowe przedstawione w tabeli 5.2 na stronie 65. Schemat blokowy przetwarzania wizji został przedstawiony na rysunku 5.1, strona 66 i jest również taki sam jak dla eksperymentu wstępnego. Zmianie natomiast uległa:

- Liczba ramek, które zostały zakodowane.

  Założono, że eksperyment ma być zgodny z ogólnymi warunkami testów MIV CTC [00620]. Będą one również zachowane dla kolejnego eksperymentu zasadniczego, co pozwoli na porównanie wyników. Dla każdej sekwencji testowej przetworzono 17 kolejnych ramek.

- Zastosowane techniki estymacji map głębi oraz syntezy widoków wirtualnych.

  Zastosowano dwa nowsze, niezależne algorytmy przetwarzania wielowidokowej wizji. Do estymacji głębi wykorzystano oprogramowanie IVDE [Rog+19], które również użyto w eksperymencie MIV, zgodnie z MIV CTC [00620]. Do syntezy wirtualnego widoku zastosowano zaawansowany algorytm opisany w pracy [Dzi+19], którego zaletą jest szybka, zoptymalizowana implementacja [SD20].

- Wykorzystana nowa metryka oceny jakości.

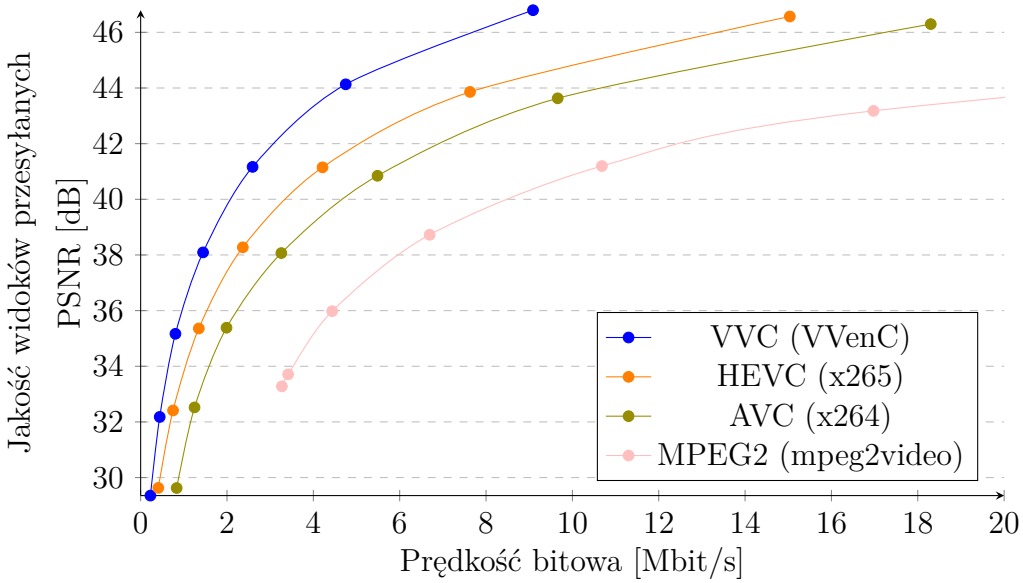
  Do oceny jakości wirtualnego widoku wykorzystano metrykę IV-PSNR [Dzi+22], która jest dostosowana do obrazów syntezowanych. Nie była ona znana podczas przeprowadzania eksperymentu wstępnego.

- Rodzaje technik kompresji, które zostały wykorzystane oraz ich konfiguracja.

  Wykorzystano szybką implementację HEVC: x265 [wX265], szybką implementację AVC: x264 [wX264], koder MPEG-2 wspierany przez oprogramowanie FFmpeg 4.4.1 [wFFmpeg]. Zastosowano również najnowszą technikę kompresji pojedynczego widoku i użyto algorytmu VVC w zoptymalizowanej implementacji VVenC [Wie+21]. Nie ingerowano w układ ramek oraz strukturę GOP, używaną przez kodery.

5.4.3 Wyniki eksperymetu ogólnego

Na wykresach 5.8 oraz 5.9 zebrano wyniki eksperymentu nazwanego eksperymentem ogólnym. Wykres 5.8 opisuje zależność pomiędzy całkowitą prędkością bitową potrzebną do transmisji wszystkich (czterech) widoków wejściowych a ich jakością wyrażoną przez średnią wartość PSNR.



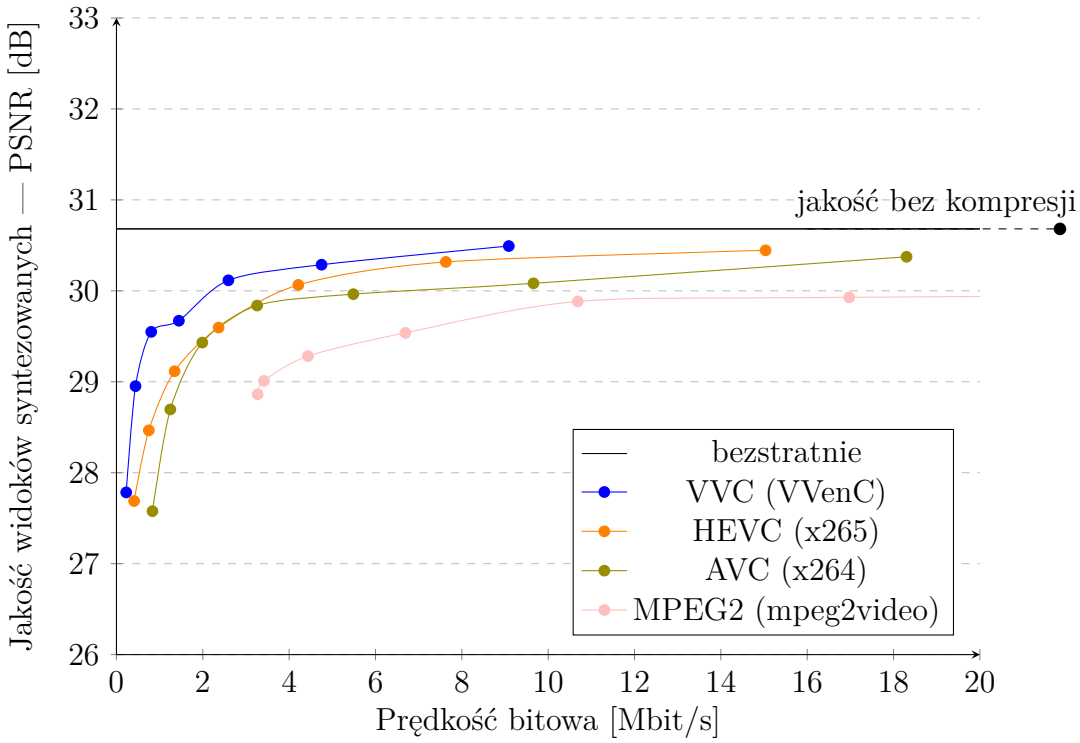
Rysunek 5.8: Eksperyment ogólny: krzywe RD dla przesyłanych widoków uśrednione dla wszystkich sekwencji

Na rysunku 5.9 przedstawiono jakość wirtualnych widoków, które są syntezowane ze zdekodowanych danych oraz map głębi, wyznaczonych na ich podstawie. Wyznaczono je z wykorzystaniem metryk: PSNR oraz IV-PSNR (rysunki 5.9a oraz 5.9b).

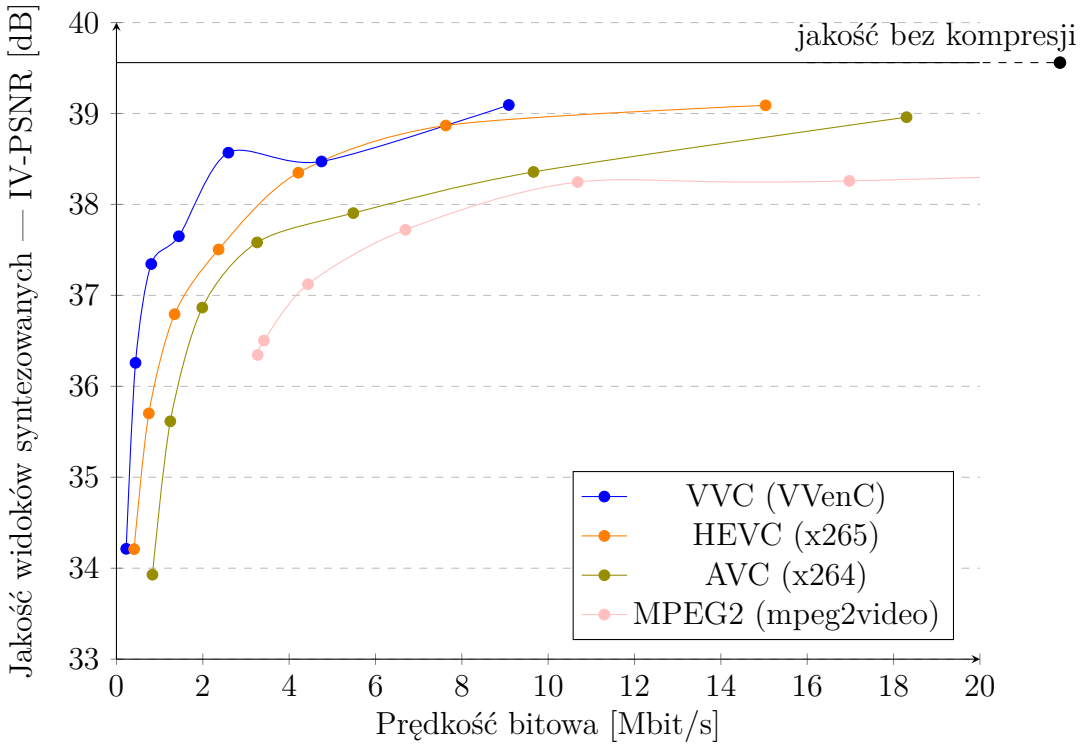
Należy zaznaczyć, że prędkości bitowe prezentowane na wykresach 5.8, 5.9a oraz 5.9b są dokładnie takie same, gdyż odpowiadają transmisji tych samych czterech widoków wielowidokowej sekwencji testowej. Zostały one poddane kodowaniu z wykorzystaniem różnego stopnia kwantyzacji oraz za pomocą różnych technik kompresji i są zgodne z oczekiwaniami. Nowsze i bardziej zaawansowane techniki kodowania wizji wykazują lepszą jakość od starszych. Jest to widoczne zarówno pod względem jakości zdekodowanych widoków, jak i jakości syntezowanych wirtualnych widoków, dla metryki PSNR oraz IV-PSNR.

Jakość widoków przesyłanych oraz syntezowanych, mierzone za pomocą średniej wartość metryk PSNR lub IV-PSNR, pozwalają na wykreślenie zależności jakości widoku przesyłanego do syntezowanego. Uzyskane krzywe zostały przedstawione na wykresie 5.10. Są to cztery niezależne linie nakreślone tym samym kolorem, które odpowiadają wykorzystanym technikom kompresji stratnej. W zestawieniu została pominięta sekwencja Rabbit Arc ze względu na duże i niemonotoniczne zmiany wartości metryk. Jest to spowodowane licznymi zakłóceniami, które były generowane przez wykorzystany algorytm estymacji głębi.

Jak można zauważyć, mimo zastosowania koderów, które znacznie różnią się efektywnością kompresji, wszystkie krzywe na wykresie 5.10 mają zbliżony przebieg, który głównie zależy od testowanej sekwencji.

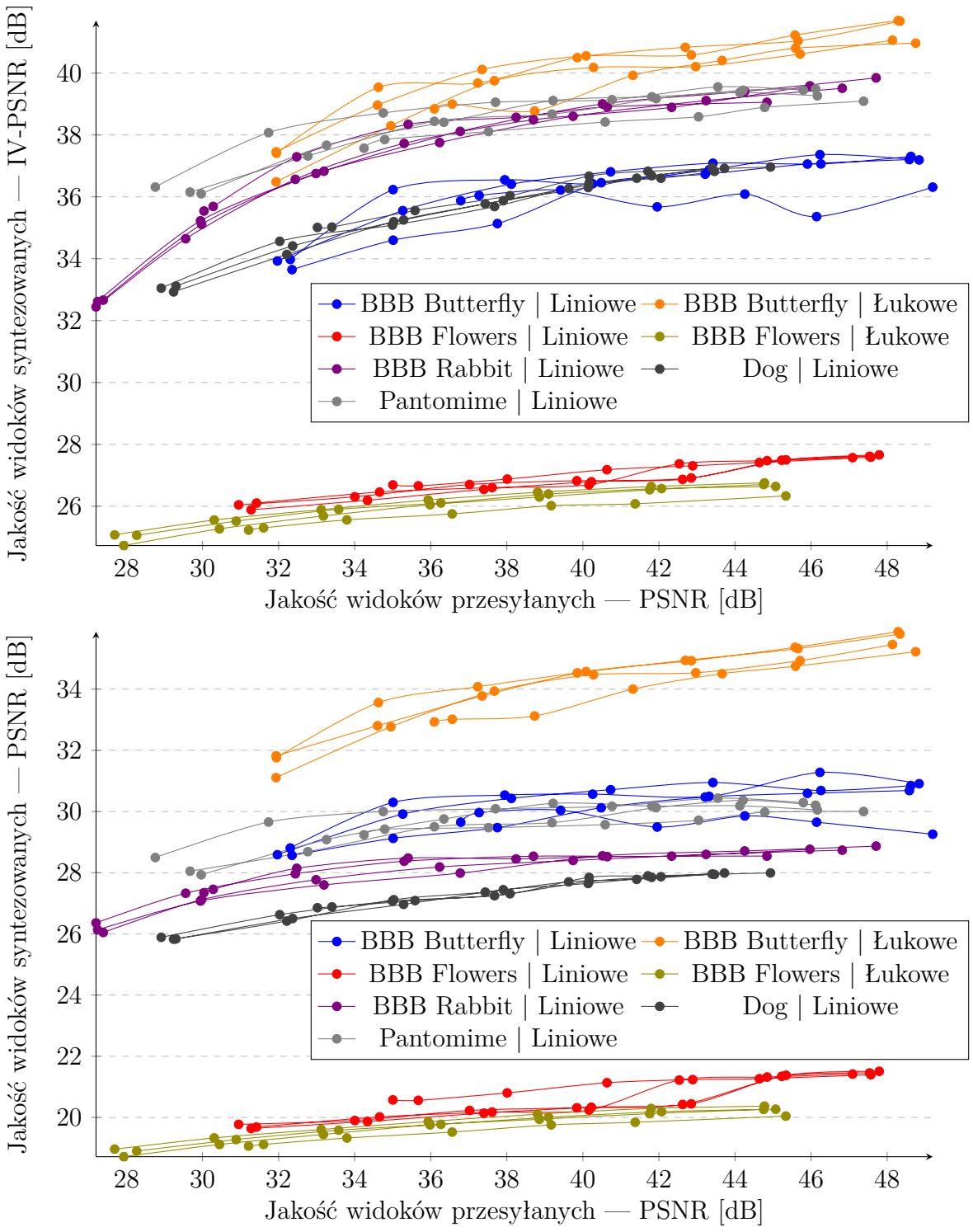


(a) Wyniki z wykorzystaniem metryki PSNR



(b) Wyniki z wykorzystaniem metryki IV-PSNR

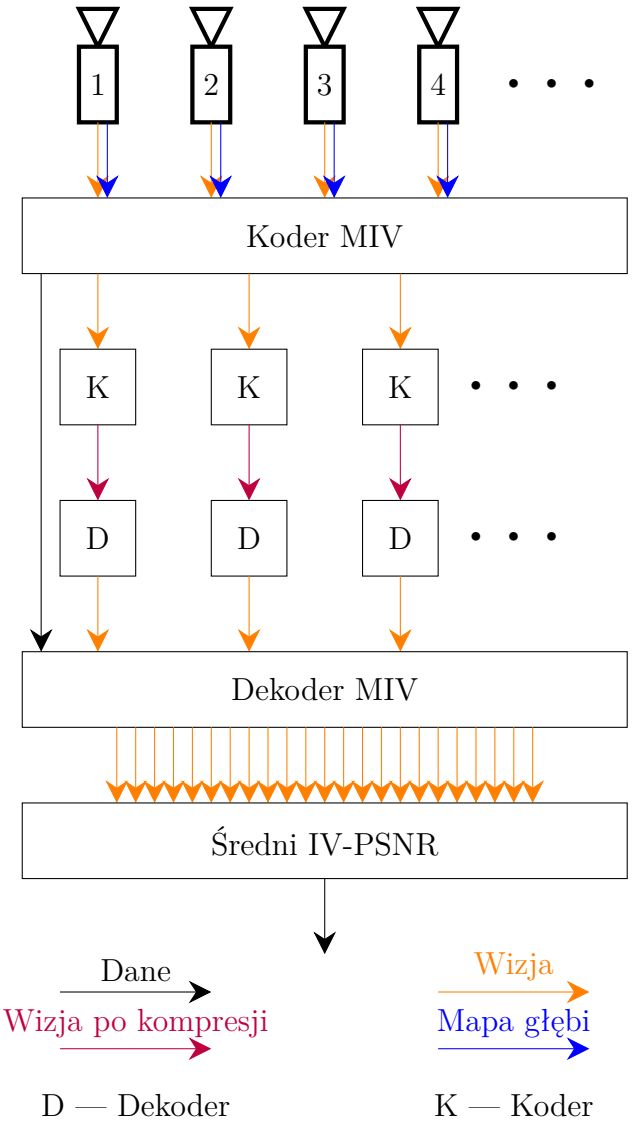
Rysunek 5.9: Eksperyment ogólny: krzywe RD widoków syntezowanych uśrednione dla wszystkich sekwencji



Rysunek 5.10: Eksperyment ogólny: zależność pomiędzy jakością widoku przesyłanego (PSNR) a jakością widoku syntezowanego (PSNR lub IV-PSNR) dla każdej sekwencji cztery krzywe odpowiadające wykorzystanym technikom kodowania

5.4.4 Opis eksperymentu MIV

Schemat blokowy przetwarzania, który użyto w eksperymencie MIV, został przedstawiony na rysunku 5.11. Ponieważ norma MIV (MPEG Immersive Video) jest niezależna od techniki kompresji wizji, do kodowania i dekodowania widoków wyrenderowanych przez koder MIV (wykorzystano profil MIV Geometry Absent [Mie+22]) mógł zostać wykorzystany każdy koder i dekoder wizji (bloki K oraz D na rysunku 5.11).



Rysunek 5.11: Schemat eksperymentu MIV.

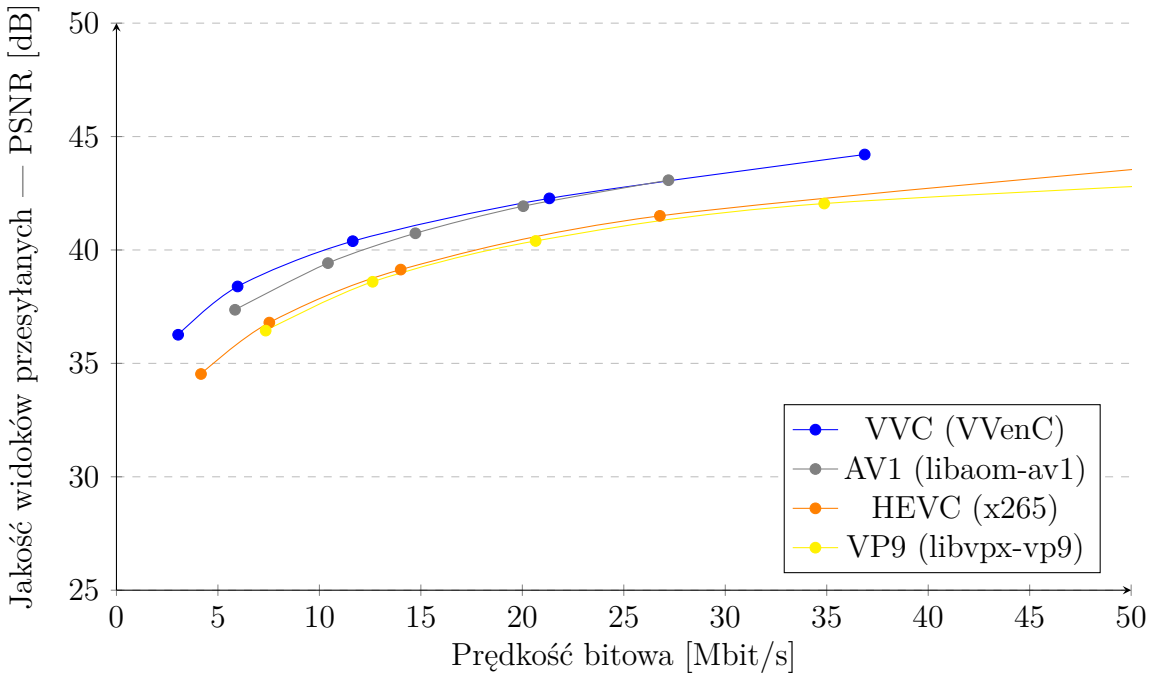
Eksperyment przeprowadzono zgodnie z wypracowanymi przez grupę ISO/IEC MPEG warunkami przeprowadzania eksperymentów (MIV Common Test Conditions — MIV CTC [00620]). Określają one zestaw 15 zróżnicowanych sekwencji (tabela 5.10) oraz definiują cały tor przetwarzania danych i oceny jakości. Wykorzystano sekwencje dookólne, perspektywiczne, treści generowane komputerowo oraz sekwencje naturalne zarejestrowane przez rzeczywiste systemy wielokamerowe. Zgodnie z MIV CTC dla każdej sekwencji testowej zakodowano 17 ramek.

W eksperymencie oceniono jakość wirtualnych widoków dla czterech różnych koderów wizyjnych, w tym dwóch koderów opracowanych przez ISO/IEC MPEG: VVC w zoptymalizowanej implementacji VVenC [Wie+21] i szybkiej implementacji HEVC: x265 [wX265], a także dwóch koderów opracowanych przez konsorcja firm: AV1 i VP9. Wykorzystano implementacje wspierane przez oprogramowanie FFmpeg 4.4.1 — libaom-av1 oraz libvpx-vp9 [wFFmpeg]. Wirtualne widoki (syntezowane w miejscu wejściowych) są wyznaczane przez dekoder MIV. Obiektywna jakość mierzona pomiędzy widokami wejściowymi a syntezowanymi była wyznaczana za pomocą metryki IV-PSNR [Dzi+22][Dzi20]. Została ona obliczona dla wszystkich widoków wejściowych i jest przedstawiona jako średnia wartość dla wszystkich widoków oraz 17 ramek obrazu.

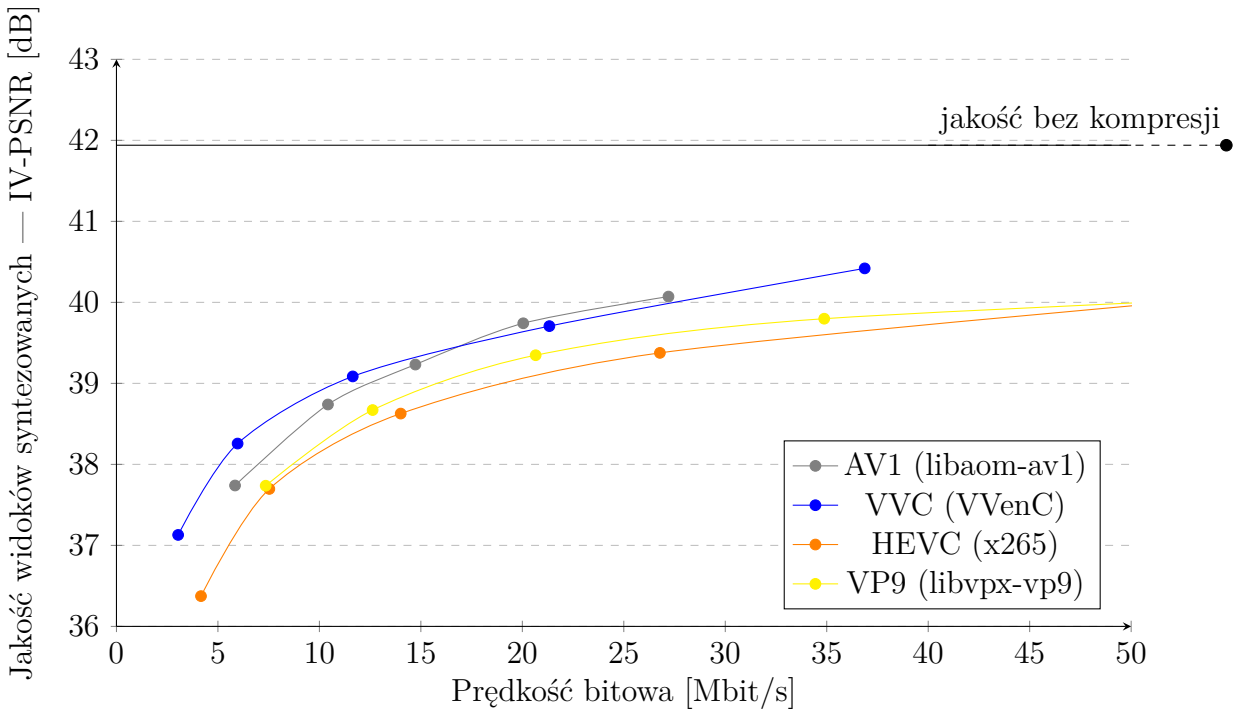
5.4.5 Wyniki eksprymentu MIV

Na wykresach 5.12 oraz 5.13 przedstawiono średnie wyniki jakości czterech testowanych koderów wizji. Na rysunku 5.12 jakość zaprezentowano w postaci krzywych PSNR dla dekodowanej wizji, która jest generowana przez koder MIV (zgodnie ze schematem 5.11) w formie atlasów. Na wykresie 5.13 przedstawiono zależność pomiędzy prędkością bitową atlasów przygotowanych przez koder MIV i kodowanych różnymi technikami kompresji a średnią jakością syntezowanych widoków (metryka IV-PSNR uśredniona dla 15 sekwencji, 17 ramek i wszystkich syntezowanych widoków wejściowych).

Należy zaznaczyć, że prędkości bitowe prezentowane na rysunkach 5.12 oraz 5.13 są dokładnie takie same, gdyż odpowiadają przepływnościom, wynikającym z przesyłanej wizji.



Rysunek 5.12: Eksperyment MIV: krzywe RD dla zdekodowanych atlasów przygotowanych przez koder MIV

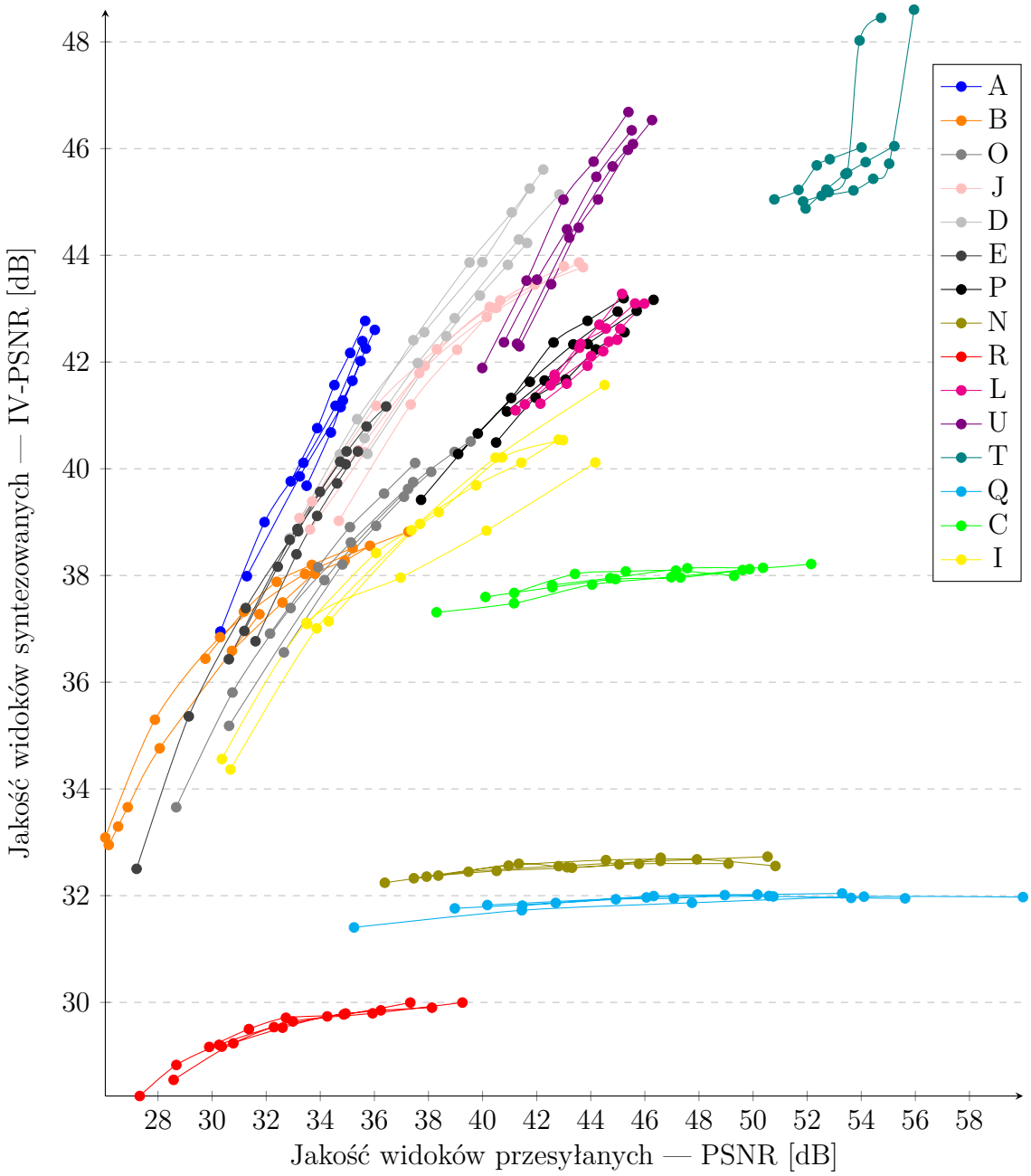


Rysunek 5.13: Eksperyment MIV: krzywe RD dla widoków syntezowanych

Porównując wyniki uzyskane przy użyciu różnych koderów, można sformułować kilka ogólnych obserwacji. Po pierwsze, potwierdzono, że MIV jest techniką niezależną od techniki kodowania, a krzywe RD dla wszystkich koderów wyglądają podobnie. Po drugie, kodery VVC i AV1 w stosowanych szybkich implementacjach wypadają podobnie zarówno pod względem jakości dekodowanych atlasów (widoków przygotowanych przez koder MIV), jak i syntezowanych widoków wirtualnych. Z drugiej strony ciekawsze wydają się wyniki dla koderów HEVC i VP9. Pod względem jakości dekodowanych atlasów oba kodery zapewniają podobne prędkości bitowe, co przedstawiono na wykresie 5.12. Jednak porównując IV-PSNR syntezowanych wirtualnych widoków, można stwierdzić niewielką, ale zauważalną lepszą jakość dla techniki VP9, co przedstawiono na wykresie 5.13.

Możliwość oceny jakości na dwóch etapach przetwarzania danych wizyjnych (przed syntezą widoku — wykres 5.12 i po syntezie widoku — wykres 5.13) pozwala na wykreślenie zależności pomiędzy tymi dwiema metrykami jakości, co zostało przedstawione na wykresie 5.14. Krzywe zostały wykreślone niezależnie dla każdej sekwencji testowej (rozróżnialne kolorem) i każdego testowanego kodera wizji (nierozróżnialne; cztery krzywe dla każdej sekwencji). Prezentują one zależność pomiędzy średnią wartością PSNR dekodowanych atlasów a średnią wartością IV-PSNR syntezowanych widoków dla każdej sekwencji (oznaczono innym kolorem) i techniki kompresji.

Jak można zauważyć, na rysunku 5.14 wykresy krzywych związanych z daną sekwencją mają zbliżony przebieg. Większe różnice pomiędzy krzywymi oraz brak zmiany jakości syntezowanego widoku można zauważyć dla sekwencji Q (ChessPieces), N (Chess), C (Hijack) i R (Group). Oznacza to, że dla tych sekwencji jakość syntezowanych widoków słabo zależy od jakości dekodowanego atlasu. Znaczne zwiększenie całkowitej prędkości bitowej prawie nie wpływa na jakość wirtualnego widoku, czyli jakość dostarczaną użytkownikowi systemu swobodnej nawigacji. Jest to spowodowane licznymi zakłóceniami w wirtualnych widokach.



Rysunek 5.14: Eksperyment MIV: zależność pomiędzy jakością widoku przesyłanego (PSNR) a widokiem syntezowanym (IV-PSNR) dla sekwencji z tablicy 5.10

5.5 Model zależności przesyłanych widoków do wirtualnego

W celu analizy rezultatów uzyskanych w eksperymentach zasadniczych: ogólnym oraz MIV przeanalizowano podobieństwo uzyskanych wyników. Zależność jakości widoków syntezowanych od przesyłanych jest podobna dla różnych technik kompresji, dlatego przeanalizowano wyniki ze względu na występowanie zależności, do której wykorzystano współczynnik korelacji liniowej Pearsona [Wei12]. Współczynnik ten wyznaczono w celu określenia liniowej zależności między jakością widoku przesyłanego (wyrażonego wartością metryki PSNR) a jakością widoku wirtualnego (mierzonego wartością metryki IV-PSNR). Jego wartości dla wyników z eksperymentu ogólnego oraz MIV przedstawiono odpowiednio w tabelach 5.8, oraz 5.7.

Tablica 5.7: Eksperyment MIV: współczynnik korelacji liniowej Pearsona pomiędzy jakością widoku przesyłanego (PSNR) a jakością widoku wirtualnego (IV-PSNR), identyfikator sekwencji w tabeli 5.10

Identyfikator	VVC	HEVC	AV1	VP9
A	1,00	1,00	1,00	1,00
B	1,00	0,95	0,99	0,97
O	0,99	0,99	0,99	1,00
J	0,97	0,99	0,99	0,98
D	1,00	1,00	1,00	1,00
E	1,00	0,99	1,00	1,00
P	1,00	1,00	0,99	0,99
N	0,95	0,97	**0,88**	**0,67**
R	0,92	0,96	0,95	0,95
L	0,99	1,00	1,00	1,00
U	0,99	1,00	1,00	0,99
T	**0,79**	0,99	**0,88**	0,97
Q	0,95	0,92	0,95	**0,73**
C	0,96	0,99	**0,66**	0,98
I	0,98	0,97	0,99	0,99

Tablica 5.8: Eksperyment ogólny: współczynnik korelacji liniowej Pearsona pomiędzy jakością widoku przesyłanego (PSNR) a jakością widoku wirtualnego (IV-PSNR), identyfikator sekwencji w tabeli 5.9

Identyfikator	VVC	HEVC	AVC	MPEG2
FL	0,98	0,99	0,97	0,99
FŁ	0,98	0,98	0,97	0,99
BL	**0,83**	0,96	0,90	**-0,03**
PA	**0,88**	0,96	0,91	0,99
BŁ	0,93	0,94	0,92	0,96
DO	0,98	0,99	0,99	0,97
RL	0,94	0,94	0,93	0,90
RŁ	**0,86**	0,98	0,98	0,92

Rezultaty wyróżnione przez pogrubienie mają wartość niższą niż $0,90$. Najgorszy wynik uzyskano dla sekwencji o identyfikatorze BL kodowanej koderem MPEG2. Odbiega on od wszystkich pozostałych i świadczy o braku zależności pomiędzy jakością widoku przesyłanego a syntezowanego, który ma stałą wartość. Pozostałe wartości korelacji liniowej mają bardzo wysoką wartość dla większości sekwencji i technik kompresji. Świadczą one o tym, że pomiędzy jakością widoku przesyłanego i wirtualnego występuje zależność liniowa. Pozwala to na zaproponowanie modelu tego podobieństwa. W tym celu, dla każdej sekwencji testowej wyznaczono krzywą liniowej regresji, uśrednioną dla wszystkich technik kompresji, która stanowi model liniowy. Nie jest on zawsze zgodny z kształtem uzyskanych krzywych, ale został wstępnie wykorzystany w celu sprawdzenia wielkości popełnianego błędu. Pozwoliło to na przybliżone, dla danej sekwencji i każdego kodera, określenie jakości widoków syntezowanych na podstawie jakości widoków przesyłanych, zgodnie z poniższym wzorem:

$$\text{IV-PSNR}_p \approx f_{AG}(\text{PSNR}_p) = a \cdot \text{PSNR}_p + b, \tag{5.1}$$

gdzie IV-PSNR_p oraz PSNR_p to zmierzone wartości metryk oceny jakości odpowiednio widoku wirtualnego oraz przesyłanego, a f_{AG} jest zaproponowanym liniowym modelem tej zależności.

Tablica 5.9: Eksperyment ogólny: wielowidokowe sekwencje testowe oraz parametry linii trendu a i b

Identyfikator	Nazwa sekwencji	a	b
FL	BBB Flowers liniowe[Kov+15]	**0,1**	23,04
FŁ	BBB Flowers łukowe[Kov+15]	**0,1**	22,48
BL	BBB Butterfly liniowe[Kov+15]	**0,15**	29,92
PA	Pantomime[TFF08]	**0,16**	32,29
BŁ	BBB Butterfly łukowe[Kov+15]	0,22	30,89
DO	Dog[TFF08]	0,25	26,31
RL	BBB Rabbit liniowe[Kov+15]	0,32	25,71
RŁ	BBB Rabbit łukowe[Kov+15]	0,34	23,98

Tablica 5.10: Eksperyment MIV: sekwencje testowe MIV CTC oraz parametry a i b modelu liniowego

Identyfikator	Sekwencja	Liczba widoków	Ustawienie kamer	a	b
Q	ChessPieces	10	Inne	**0,02**	31,06
N	Chess	10	Inne	**0,03**	31,43
C	Hijack	10	Inne	**0,06**	35,35
R	Group	21	Inne	**0,13**	25,15
P	Carpark	9	Liniowe	0,43	23,25
I	Mirror	15	Inne	0,44	21,67
L	Fencing	10	Łukowe	0,45	22,32
J	Kitchen	25	Inne	0,50	22,73
T	Hall	9	Liniowe	0,55	16,47
B	Museum	24	Inne	0,55	19,27
O	Fan	15	Inne	0,61	16,80
D	Painter	16	Inne	0,67	16,97
U	Street	9	Liniowe	0,83	8,70
E	Frog	13	Inne	0,90	8,88
A	ClassroomVideo	15	Inne	0,99	7,08

W tabelach 5.9 i 5.10 przedstawiono parametry a oraz b zaproponowanego modelu. Niewielkie wartości parametru a (poniżej wartości 0,2) wyróżniono kolorem czerwonym. Dla wszystkich takich sekwencji zmiana jakości dekodowanych atlasów w niewielki sposób wpływa na jakość wirtualnych widoków. Jest to spowodowane pojawieniem się silnych zakłóceń oraz błędów syntezy, co przedstawiono na rysunkach: 5.15 dla przykładowych sekwencji z eksperymentu ogólnego oraz 5.16 dla sekwencji z eksperymentu MIV.

Rysunek 5.15: Eksperyment ogólny: widoki wirtualne z zakłóceniami

Rysunek 5.16: Eksperyment MIV: widoki wirtualne z zakłóceniami

Dla każdej sekwencji (s) i techniki kompresji (k) — tabele 5.11 oraz 5.12 — wyznaczono błąd zaproponowanego modelu liniowego, zgodnie z wzorem:

$$\varepsilon(s,k,p) = \text{IV-PSNR}_p - f_{AG}(\text{PSNR}_p), \tag{5.2}$$

gdzie p oznacza punkt testowy, dla którego dokonano pomiaru wartości PSNR widoku przesyłanego (PSNR_p) oraz zmierzonej wartości IV-PSNR widoku syntezowanego (IV-PSNR_p). W tabelach 5.11 oraz 5.12 przedstawiono maksymalny oraz średni błąd bezwzględny (odpowiednio ε_m oraz ε_{sr}) wyznaczony dla liniowego modelu, zgodnie z wzorami:

$$\varepsilon_m(s,k) = \max_p |\varepsilon(s,k,p)|, \tag{5.3}$$

$$\varepsilon_{sr}(s,k) = \frac{\sum_p |\varepsilon(s,k,p)|}{P}, \tag{5.4}$$

gdzie P to liczebność zbioru pomiarów.

Tablica 5.11: Eksperyment ogólny: błąd [dB] maksymalny ε_m i średni ε_{sr} przybliżonej krzywej liniowej, identyfikator sekwencji w tabeli 5.9

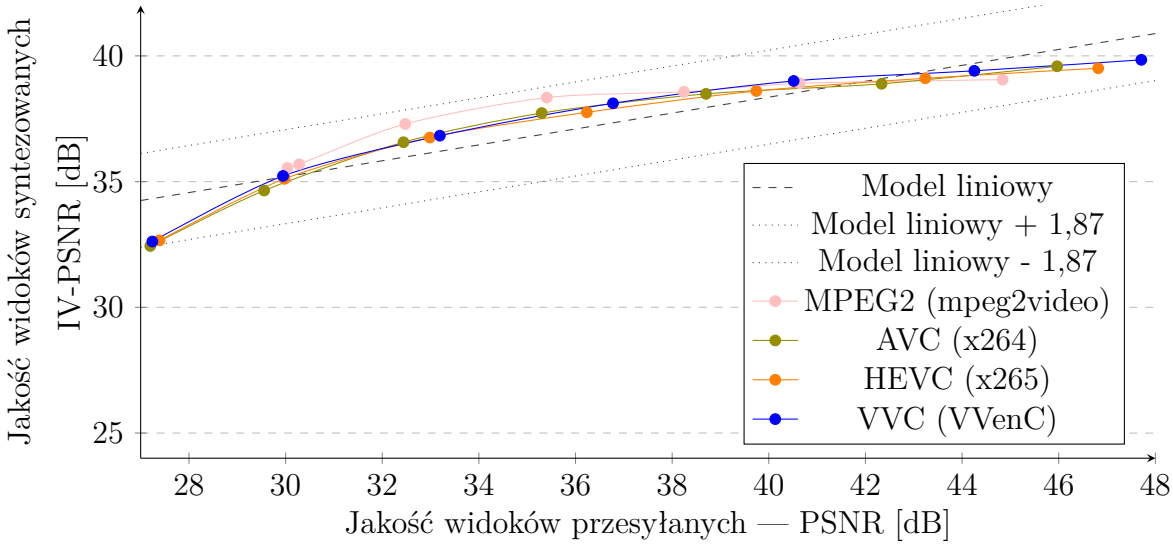
Identyfikator		technika kompresji (k)							
		VVC		HEVC		AVC		MPEG2	
		$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$
sekwencje (s)	FL	0,29	0,08	0,27	0,14	0,23	0,11	0,27	0,15
	FŁ	0,31	0,19	0,40	0,14	0,23	0,15	0,45	0,24
	BL	**1,02**	**0,59**	**1,17**	0,46	0,83	0,48	**1,54**	**0,70**
	PA	0,90	**0,54**	0,85	0,30	0,95	0,39	0,73	0,35
	BŁ	0,96	0,48	0,93	0,44	**1,50**	**0,55**	0,75	0,29
	DO	0,46	0,20	0,49	0,16	0,66	0,21	0,53	0,37
	RL	**1,71**	**0,70**	**1,71**	**0,66**	**1,87**	**0,73**	**1,43**	**0,78**
	RŁ	**2,32**	**1,20**	**1,11**	**0,68**	**1,59**	0,71	**2,04**	**1,23**

Tablica 5.12: Eksperyment MIV: błąd [dB] maksymalny ε_m i średni ε_{sr} przybliżonej krzywej liniowej, identyfikator sekwencji w tabeli 5.10

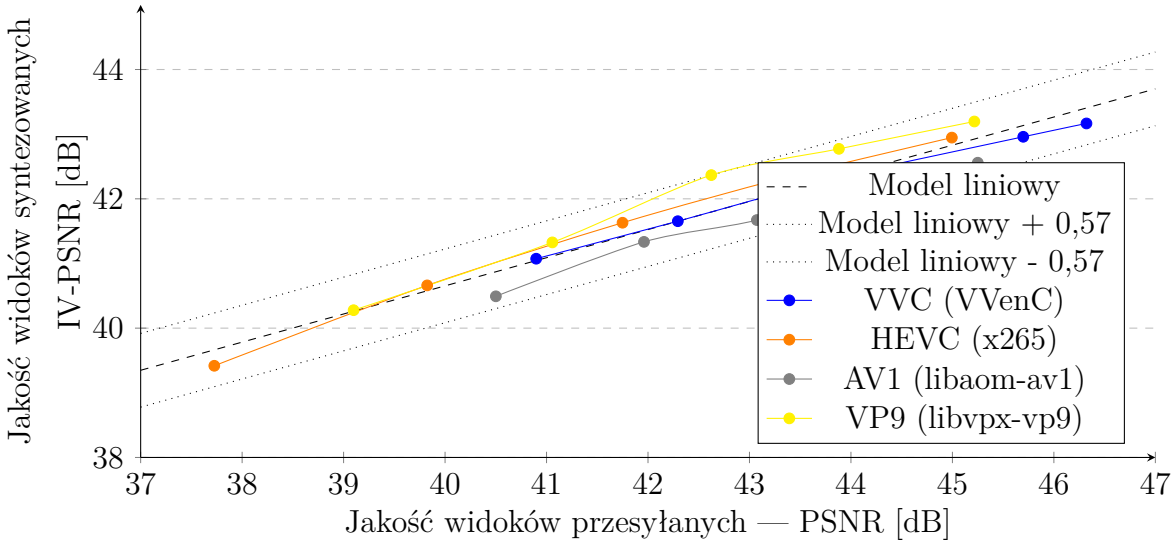
Identyfikator		technika kompresji (k)							
		VVC		HEVC		AV1		VP9	
		$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$	$\varepsilon_m(s,k)$	$\varepsilon_{sr}(s,k)$
sekwencje (s)	A	0,27	0,11	0,36	0,14	0,50	0,29	0,45	0,31
	B	0,82	0,49	**1,09**	**0,66**	0,57	0,28	0,78	0,53
	O	0,59	0,33	0,76	0,31	0,30	0,14	0,55	0,42
	J	0,62	0,40	0,46	0,25	0,90	0,27	0,57	0,33
	D	0,41	0,21	0,43	0,31	0,53	0,43	0,53	0,40
	E	0,37	0,20	0,77	0,40	0,44	0,24	0,22	0,17
	P	0,24	0,09	0,25	0,18	0,38	0,30	0,57	0,30
	N	0,08	0,03	0,12	0,04	0,11	0,05	0,17	0,09
	R	0,34	0,18	0,48	0,18	0,24	0,11	0,28	0,15
	L	0,17	0,10	0,50	0,22	0,28	0,25	0,25	0,14
	U	0,47	0,26	0,24	0,11	0,59	0,33	0,81	0,49
	T	**1,31**	**0,91**	0,85	0,45	**1,84**	**0,89**	0,60	0,34
	Q	0,11	0,06	0,28	0,11	0,08	0,05	0,10	0,06
	C	0,10	0,08	0,20	0,09	0,23	0,13	0,06	0,04
	I	**1,09**	0,47	**1,19**	**0,67**	0,47	0,21	0,63	0,42

Jak można zauważyć, błąd maksymalny rzadko przekracza wartość 1 dB, a błąd średni dla większości przedstawionych wyników ma wartość mniejszą niż 0,5 dB. Takie wyniki zostały wyróżnione poprzez pogrubienie. Zaproponowany liniowy model przybliża krzywe dla różnych technik kompresji z dużą dokładnością — w szerokim zakresie jakości popełniany błąd zazwyczaj jest mniejszy niż 1 dB.

Zależności metryki IV-PSNR od wartości PSNR z rozróżnieniem wszystkich badanych technik kompresji wraz z zaproponowanym liniowym modelem oraz liniami przesuniętymi o maksymalną wartość błędu przedstawiono na rysunkach 5.17 oraz 5.18 dla wybranych sekwencji testowych. Jak wynika z tabeli 5.11, błąd dla sekwencji BBB Rabbit (RL, RŁ) jest największy i przekracza 1 dB. Warto zwrócić uwagę, że na wykresie 5.17 wartość tego błędu wynika z kształtu krzywych, który nie jest liniowy, a wykresy kodowania dla wszystkich technik mają bardzo podobny przebieg. Zaproponowany model można by łatwo zmodyfikować, wykorzystując inne modele niż liniowy, co w tym przypadku zmniejszyłoby wartość błędu. Dla większości sekwencji zaproponowany model liniowy przybliża jakość z niedużym błędem (przykładowo dla sekwencji Carpark, przedstawionej na wykresie 5.18). Wynosi on zazwyczaj mniej niż średnio 0,5 dB, a maksymalna jego wartość nie przekracza 1 dB.



Rysunek 5.17: Eksperyment ogólny: zależność pomiędzy jakością widoku przesyłanego (PSNR) a jakością widoku syntezowanego (IV-PSNR) wraz z modelem liniowym oraz krzywymi przesuniętymi o ε_m dla sekwencji Rabbit Linear



Rysunek 5.18: Eksperyment MIV: zależność pomiędzy jakością widoku przesyłanego (PSNR) a jakością widoku syntezowanego (IV-PSNR) wraz z modelem liniowym oraz krzywymi przesuniętymi o ε_m dla sekwencji P — Carpark

5.6 Wnioski

W rozdziale przedstawiono analizę wpływu kompresji stratnej wprowadzanej przez różne kodery wizyjne na proces estymacji map głębi oraz syntezy wirtualnego widoku. Przeprowadzone badania pozwoliły na poczynienie istotnych obserwacji.

Uzyskane wyniki wykazały, że wysokiej jakości kompresja stratna wizji nie wpływa znacząco na jakość syntezowanych wirtualnych widoków. W eksperymentach wstępnym oraz ogólnym dla wszystkich badanych koderów różnica między jakością wirtualnego obrazu syntezowanego przy użyciu widoków bez kompresji stratnej a obrazami poddanymi kompresji nie była większa niż 1 dB — zarówno dla metryki PSNR, jak i metryki dostosowanej do obrazów syntezowanych IV-PSNR. Ponadto dla bardzo małych wartości współczynnika kwantyzacji różnica była mniejsza, np. dla $QP = 20$ wynosiła około $0,5\ dB$ (poza techniką kodowania MPEG-2). Wykazano również, że metryka jakości wirtualnych widoków ma najwyższą wartość dla ramek typu I, ale strata jakości w ramkach P i B jest pomijanie mała dla niewielkich wartości współczynnika kwantyzacji i w przedstawionych badaniach nie przekraczała wartości $0,1\ dB$ dla $QP = 20$ oraz $QP = 25$.

W eksperymencie przeprowadzonym z wykorzystaniem techniki MPEG Immersive Video (MIV) różnica jakości jest większa, ale nie przekracza 2 dB. Wynika to z bar-

dziej zaawansowanych metod przetwarzania w koderze i dekoderze (przygotowanie przesyłanych widoków — atlasów) oraz bardziej rozbudowanego zestawu badanych sekwencji, uwzględniających chociażby różne ustawienia oraz rodzaje kamer. W niektórych z wirtualnych widoków również wystąpiły liczne zakłócenia i błędy syntezy, co obniżyło uśrednioną wartość wyniku.

Zaprezentowane wyniki wykazują, że zarówno w przypadku wykorzystania techniki ISO/IEC, dotyczącej kompresji wizji wszechogarniającej (MIV), jak i w ogólnym przypadku jakość wirtualnych widoków zależy od jakości widoku przesłanego. Zastosowana technika kompresji w niewielkim stopniu wpływa na jakość wirtualnych widoków, a tym samym na postrzeganą jakość usługi przez odbiorcę systemu.

Na podstawie poczynionych obserwacji w pracy zaproponowano model zależności pomiędzy zmianą jakości dekodowanych widoków wejściowych (mierzoną za pomocą metryki oceny jakości PSNR) a zmianą jakości wirtualnych widoków (mierzoną za pomocą metryki oceny jakości IV-PSNR) syntezowanych na ich podstawie. Pozwala on oszacować, jaka będzie postrzegana jakość usług w systemach wirtualnej nawigacji, niezależnie od stosowanej techniki kompresji, na podstawie jakości przesyłanego widoku. W opracowanym modelu średni błąd jest zazwyczaj mniejszy niż 0,5 dB, a maksymalny — mniejszy niż 1 dB. Mimo prostej liniowej definicji stanowi on dość dobre przybliżenie zależności pomiędzy jakościami widoków w szerokim zakresie jakości. Należy zauważyć, że większe wartości błędu dla niektórych sekwencji wynikają z nieliniowego kształtu krzywych jakości wirtualnego widoku względem jakości widoku przesyłanego. Błąd ten byłby mniejszy w przypadku wykorzystania innego modelu niż linowy. Jednakże, dla większości badanych sekwencji, zaproponowany model zapewnia małą wartość błędu średniego i maksymalnego.

Wykazano również, że współczynnik kierunkowy a opracowanego liniowego modelu zależności pomiędzy jakością przesyłanych widoków a jakością wirtualnych widoków, wskazuje na poprawną syntezę. Duża wartość tego parametru oznacza, że spadek jakości w przesyłanym widoku silnie wpływa na jakość obrazu. Jest to oczekiwana zależność, która zazwyczaj oznacza syntezowany widok bez zakłóceń. Mała wartość parametru a, której odpowiada linia modelu przebiegająca płasko, wskazuje, iż zmiana jakości przesyłanego widoku nie wpływa na jakość systemu. Oznacza to, że wykorzystane techniki estymacji map głębi oraz syntezy wirtualnych widoków z różnych przyczyn nie funkcjonują prawidłowo, a mapy głębi oraz syntezowane

widoki posiadają liczne zakłócenia. Powodem mogą być m.in. niepoprawne parametry sekwencji, odbicia w scenie lub błędy w oprogramowaniu.

Należy zaznaczyć, że wszystkie badania przedstawione w tym rozdziale dotyczyły wykorzystania stratnie skompresowanej wizji w systemach swobodnego punktu widzenia. Wnioski i obserwacje dowodzą, że estymacja głębi po kompresji jest możliwa i może zapewniać postrzeganą wysoką jakość. Jakość ta zależy od jakości przesyłanego widoku, a w niewielkim stopniu od zastosowanej techniki kompresji.

6 Analiza wpływu opóźnień na postrzeganą jakość usługi w wyświetlaczach nagłownych

6.1 Wstęp

Celem badań przedstawionych w tej części rozprawy doktorskiej jest sprawdzenie wpływu opóźnień transmisji na ludzką percepcję i doświadczaną jakość usług w systemach z wyświetlaczem nagłownym (ang. HMD — Head-mounted display). Aby zapewnić wysoką jakości usługi oraz swobodnego przemieszczania się w scenie wirtualnej, prezentowany widok powinien zmieniać się zgodnie z rzeczywistą pozycją i kierunkiem głowy użytkownika. Innymi słowy, opóźnienie reakcji pomiędzy zmianą rzeczywistej pozycji i/lub kierunku obrotu głowy użytkownika a wyświetleniem zaktualizowanego widoku wirtualnego T powinno być możliwie małe.

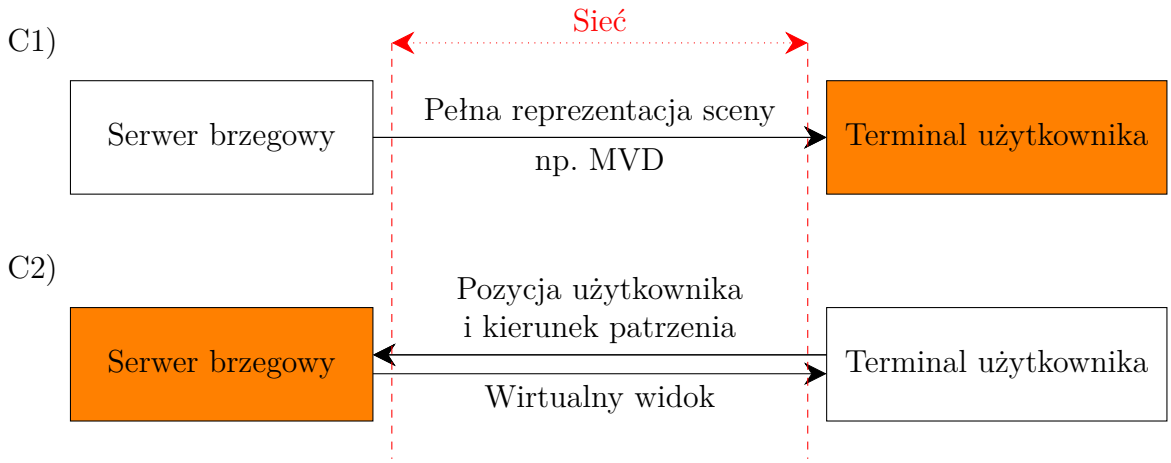
W pracy przeanalizowano różne scenariusze transmisji, w których występują składniki sumarycznego czasu T oraz przeprowadzono testy subiektywne, analizujące ich wpływ na subiektywną jakość usług.

6.2 Scenariusze dostarczania danych dedykowanych do terminala użytkownika

Na rysunku 6.1 przedstawiono dwa skrajne sposoby dostarczania danych do terminala użytkownika przez łącze C, zgodnie z systematyką przedstawioną w rozdziale 2.2. W rozwiązaniu **C1** zaprezentowanym na rysunku 6.1 terminal ten otrzymuje pełną reprezentację sceny (rozdział 2.3), która nie zależy od pozycji i kierunku patrzenia. W tym przypadku synteza wirtualnych widoków odbywa się wyłącznie w terminalu użytkownika. Zaletą takiego rozwiązania jest krótki czas przygotowywania wirtualnego widoku — wszystkie dane są dostępne i stale aktualizowane przez serwer, a opóźnienie T wynika jedynie z czasu T_u przetwarzania danych przez terminal użytkownika:

$$T = T_u. \tag{6.1}$$

Niezależnie od szczegółowej realizacji, pełna reprezentacja sceny wymaga transmisji dużej ilości danych do terminala. Znaczna część z nich nie jest wykorzystywana, co jest nieefektywne. Ponadto, przeprowadzanie syntezy w terminalu użytkownika



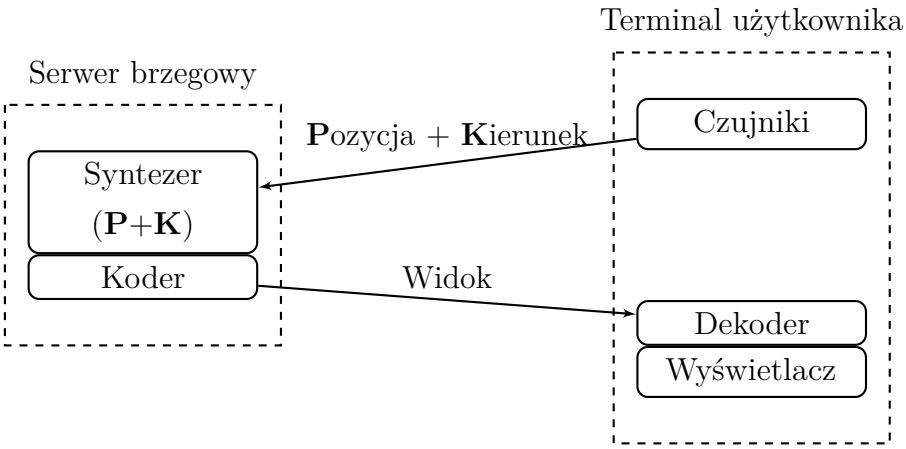
Rysunek 6.1: Podstawowe sposoby wymiany danych pomiędzy terminalem użytkownika a serwerem (syntezę widoku wirtualnego symbolizuje kolor pomarańczowy)

na podstawie pełnej reprezentacji sceny wiąże się z dwiema wadami. Po pierwsze, terminal musi zapewniać moc obliczeniową, niezbędną na wykonanie syntezy. Po drugie, obciążenie obliczeniowe wpływa negatywnie na czas pracy terminali mobilnych zasilanych baterią.

Drugie skrajne rozwiązanie — **C2** przedstawiono również na rysunku 6.1. Polega ono na transmisji informacji o pozycji użytkownika i kierunku patrzenia do serwera brzegowego oraz przygotowaniu wirtualnego, dedykowanego widoku. Takie rozwiązanie znacząco zmniejsza ilość obliczeń wykonywanych przez terminal użytkownika, ale zwiększa czas opóźnienia T potrzebny na otrzymanie wirtualnego widoku. Opóźnienie to jest sumą czasów transmisji informacji o pozycji i kierunku patrzenia (łącze w górę T_{Nup}), transmisji danych z serwera do terminala użytkownika (łącze w dół T_{Ndown}) oraz czasów przetwarzania danych przez serwer i urządzenie użytkownika (T_s oraz T_u):

$$T = T_{Nup} + T_s + T_{Ndown} + T_u. \tag{6.2}$$

Podczas badań przedstawionych w dalszej części rozprawy skupiono się na drugim sposobie, polegającym na przesyłaniu danych dedykowanych dla użytkownika (C2). W tym rozwiązaniu jest możliwa optymalizacja składników opóźnienia T. Główne składniki tego opóźnienia to zazwyczaj czas transmisji danych przez sieć, czyli T_{Nup} oraz T_{Ndown}, będące odpowiednio czasem transmisji informacji o pozycji i kierunku



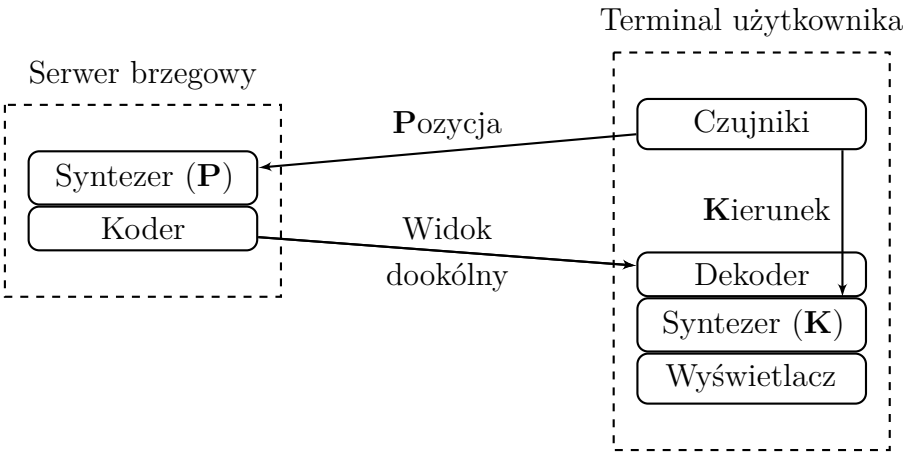
Rysunek 6.2: Scenariusz C2.1: bez syntezera w terminalu użytkownika

patrzenia oraz czasem transmisji wirtualnego widoku. Przy założeniu, że czasy przetwarzania danych przez serwer oraz terminal użytkownika są pomijalnie małe, można wyszczególnić następujące scenariusze:

- scenariusz C2.1, w którym występuje opóźnienie transmisji pozycji i kierunku patrzenia użytkownika,
- scenariusz C2.2, w którym występuje jedynie opóźnienie transmisji pozycji użytkownika,
- scenariusz C2.3, w którym opóźnienia transmisji pozycji i kierunku nie występują.

Scenariusze te zostały opisane poniżej i przedstawione odpowiednio na rysunkach 6.2, 6.3 oraz 6.4.

W scenariuszu **C2.1** serwer przygotowuje wirtualny widok, który musi zostać jedynie zdekodowany oraz wyświetlony w wyświetlaczu HMD, co zostało przedstawione na rysunku 6.2. Blok Syntezer **(P+K)** realizuje syntezę widoku z prawidłową pozycją użytkownika w scenie oraz poprawnym kierunkiem patrzenia (takie dane są przesyłane do serwera). W tym scenariuszu ilość obliczeń realizowanych przez terminal użytkownika jest mniejsza niż w scenariuszach C2.2 i C2.3. Ponadto, możliwe jest wykorzystanie algorytmów predykcji pozycji i kierunku patrzenia użytkownika w celu minimalizacji wpływu opóźnienia [Fue+19][Ala+17][Bao+16][Kim+18]. W scenariuszu C2.1 czas prezentacji poprawnej pozycji wirtualnego widoku T_p oraz opóźnienie



Rysunek 6.3: Scenariusz C2.2: uproszczony syntezer w terminalu użytkownika

kierunku patrzenia T_k będą takie same i równe:

$$T = T_p = T_k, \tag{6.3}$$

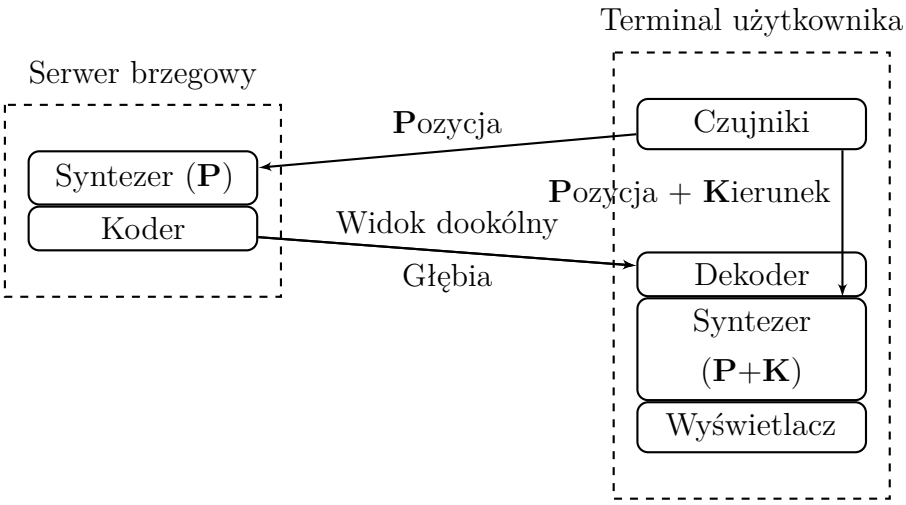
$$T = T_{Nup} + T_s + T_{Ndown} + T_u \tag{6.4}$$

W scenariuszu **C2.2** (rysunek 6.3) do serwera dociera jedynie informacja o pozycji użytkownika w scenie bez informacji o kierunku patrzenia. Serwer wysyła wirtualny widok dookólny w poprawnej pozycji użytkownika **(P)**. Taki widok jest przetwarzany po stronie terminala użytkownika w celu uzyskania widoku wyświetlanego przez hełm nagłowny. W tym scenariuszu synteza sprowadza się jedynie do projekcji widoku z odpowiednim kierunkiem, co zostało zaprezentowane na rysunku przez blok Syntezer **(K)**. Czas opóźnienia prezentacji poprawnej pozycji wirtualnego widoku T_p będzie taki sam jak w scenariuszu C.2.1. Czas opóźnienia kierunku patrzenia będzie natomiast pomijalnie mały:

$$T_p = T_{Nup} + T_s + T_{Ndown} + T_u, \tag{6.5}$$

$$T_k \approx 0. \tag{6.6}$$

Przykładem rozbudowy oraz modyfikacji scenariusza C2.2 może być sytuacja, w której serwer odbiera również informacje o kierunku patrzenia użytkownika z hełmem HMD i wykorzystuje dane o obrocie w celu przesłania widoku w bardziej wydajny sposób. Przykładowo serwer może zastosować poprawę jakości obrazu w kierunku zgodnym z kierunkiem patrzenia użytkownika. Można to zrealizować poprzez



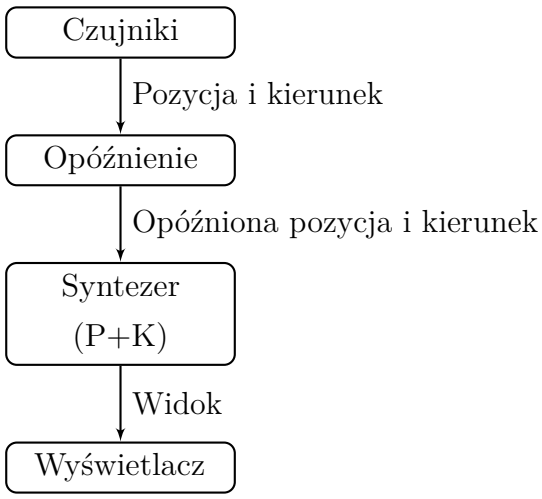
Rysunek 6.4: Scenariusz C2.3: pełen syntezer w terminalu użytkownika

adaptacyjną kwantyzację lub adaptacyjnie dobieraną rozdzielczość, a nawet prędkość ramkową (np. przy użyciu rozwiązań [Dí+18], [Sre+16], [Cor+17], MPEG-DASH [HS16] lub MPEG-OMAF [Run00]). Taka optymalizacja ma jedną istotną wadę: może ona spowodować, że przy dynamicznych i dużych ruchach oraz obrotach użytkownika, a także przy dużych opóźnieniach transmisji, jakość prezentowanej wizji może się pogorszyć w stosunku do niezmodyfikowanego scenariusza C2.2. Jednakże czasy opóźnienia T_p oraz T_k nie ulegną zmianie.

Scenariusz **C2.3** został zaprezentowany na rysunku 6.4. Serwer wysyła widok dookólny (związany z przesłaną lokalizacją użytkownika) oraz odpowiadającą mu mapę głębi (bądź dowolną inną przestrzenną reprezentację sceny), a użytkownik dokonuje syntezy widoku poprzez reprojekcję przesyłanej wizji do aktualnej pozycji. W takim przypadku opóźnienia związane z przemieszczaniem się T_p oraz kierunkiem patrzenia T_k zależą jedynie od czasu przetwarzania użytkownika i są pomijalnie małe:

$$T_p = T_k \approx 0. \tag{6.7}$$

Zaletą tego rozwiązania jest mniejsza ilość przesyłanych danych. Główna jego wada to natomiast obciążony terminal, który musi mieć dużą moc obliczeniową, niezbędną do pełnej syntezy zarówno kierunku patrzenia, jak i pozycji użytkownika. Poza tym synteza wirtualnego widoku jest wykonywana dwa razy, co może powodować powstawanie dodatkowych wad obrazu wnikających z podwójnej syntezy, np. pęknięć (tj. pionowych obszarów bez próbek), będących skutkiem zmiany rozdzielczości



Rysunek 6.5: Schemat blokowy potoku przetwarzania danych podczas eksperymentu

przestrzennej i — przede wszystkim — pojawiania się obszarów przysłoniętych, dla których nie było danych. Istnieje wiele zaawansowanych algorytmów, które mogą zostać wykorzystane do poprawnego wirtualnego widoku (np. [KEV18] [ZL16] [LZ17]), jednak w scenariuszu C2.3 muszą być one wykonane także w syntezie realizowanej przez terminal użytkownika, co podnosi złożoność i wymaganą moc obliczeniową. Warto zauważyć, że transmisję zgodnie z tym scenariuszem można również rozpatrywać jako modyfikację rozwiązania C1 przedstawionego na rysunku 6.1. W scenariuszu C2.3, tak samo jak w rozwiązaniu C1, opóźnienie T wynika jedynie z czasu przetwarzania przez terminal użytkownika. Dane o pozycji i kierunku syntezowanego widoku pozwalają ograniczyć ilość przesyłanych informacji, a terminal użytkownika musi mieć moc obliczeniową, niezbędną do pełnej syntezy (P+K) wirtualnego widoku.

6.3 Badania z wykorzystaniem wyświetlacza nagłownego

Celem przedstawionego eksperymentu jest oszacowanie, jak wartość opóźnienia T wpływa na postrzeganą przez użytkowników jakość. Eksperyment przeprowadzono w formie testów subiektywnych dla dwóch scenariuszy, w których występują opóźnienia pozycji i/lub kierunku patrzenia użytkownika. Są to rozwiązania zapewniające transmisje dedykowanych danych do terminala oraz niewymagające pełnej syntezy. Zostały one przedstawione jako scenariusz C2.1 oraz scenariusz C2.2.

W eksperymencie wprowadzono modyfikacje w schemacie każdego z wyżej przedstawionych scenariuszy, aby uprościć implementację, pozwalającą na sprawne przeprowadzanie subiektywnych badań. Po pierwsze, aby ocenić tylko efekty związane z opóźnieniem, a nie samą jakością wizji, zastosowano system bez kompresji, czyli pominięto bloki kodera i dekodera. Po za tym opóźnienia transmisji, które występują w rzeczywistym systemie, były symulowane za pomocą jednego bloku opóźnienia, który został przedstawiony na rysunku 6.5 . Blok ten może zmodyfikować parametry wirtualnego widoku niezależnie dla kierunku patrzenia i pozycji użytkownika. Testowane scenariusze C2.1 oraz C2.2 wymagają jedynie odpowiedniej modyfikacji ustawień bloku opóźnienia i mogą być realizowane przez to samo oprogramowanie.

Opis eksperymentów:

Osoby poddane testom dokonywały ocen jakości swobodnej nawigacji z wykorzystaniem wyświetlacza nagłownego. Wyświetlane widoki wirtualne były syntezowane zgodnie z aktualną pozycją i kierunkiem hełmu HMD. W eksperymencie wykorzystano jedną ramkę wielowidokowych sekwencji testowych i z powodu możliwości swobodnej nawigacji jest ona nazywana sekwencją testową. W testach użyto pierwszą klatkę z trzech dookólnych sekwencji. Dwie z nich (ClassroomVideo i Museum) są rekomendowane przez ISO/IEC MPEG do badań 3DoF+ [Jun+18], a jedna (Poznań-People360) pochodzi z zestawu badawczego, zaproponowanego przez Politechnikę Poznańską i ETRI [Sta+18b]. Sekwencja Museum wstępnie została przetworzona przed eksperymentem — widoki półsferyczne połączono w jeden dookólny widok, a rozmiar ramki został zmniejszony do 2048×1024 w celu skrócenia czasu przetwarzania i zagwarantowania płynności prezentowanych treści.

Podczas testów wykorzystano urządzenie Oculus Rift CV1 [wOcu]. Zawartość została wyrenderowana przez dedykowane oprogramowanie z wykorzystaniem karty graficznej nVidia GTX 970. Prędkość odświeżania obrazu wynosiła 90 klatek na sekundę dla każdego oka użytkownika, co ograniczyło możliwe zmiany kroku opóźnienia do 11 ms. Odpowiada to opóźnieniu parametrów widoku, czyli pozycji i kierunku prezentacji wizji w HMD o pojedynczą klatkę. Czas, związany z przechwytywaniem i przetwarzaniem danych przez czujniki i oprogramowanie Oculus, jest pomijalnie krótki. Wyświetlacz HMD podłączono bezpośrednio do wyjścia HDMI karty graficznej oraz posiada wyświetlacz typu OLED (ang. Organic Light-Emitting Diode). Opóźnienie wyświetlania (w znaczeniu 80% czasu odpowiedzi — ang. 80% response time) w tego typu ekranach jest mniejsze niż 1 ms nawet dla dużych ekranów tele-

wizyjnych [wOLED]. Urządzenie Oculus Rift już w wersji prototypowej zapewniało czas reakcji poniżej 1 ms [wArstechnica][KM16].

Tablica 6.1: Opóźnienia prezentowane podczas pojedynczej sesji; kolejność losowano dla każdego uczestnika niezależnie

Kolejność prezentacji	T_u	T_k	Scenariusz
1	0ms	0ms	Odniesienie
2	110ms	110ms	Kotwica
Losowa kolejność	0ms	0ms	Ukryte odniesienie
	11ms	11ms	C2.1
	22ms	22ms	C2.1
	33ms	33ms	C2.1
	44ms	44ms	C2.1
	55ms	55ms	C2.1
	0ms	44ms	C2.2
	0ms	88ms	C2.2
	0ms	132ms	C2.2
	0ms	176ms	C2.2
	0ms	220ms	C2.2

W testach badano dwa warianty opóźnienia: opóźnienie syntezowania zarówno kierunku patrzenia, jak i pozycji użytkownika (co odpowiada scenariuszowi C2.1, rysunek 6.2) oraz opóźnienie jedynie pozycji użytkownika w scenie (bez opóźnienia kierunku patrzenia scenariusz C2.2, rysunek 6.3). Wartości opóźnień, które są wykorzystane podczas testów subiektywnych, zostały ustalone przed przedstawionym eksperymentem i opierały się na testach wstępnych, przeprowadzonych przez grupę ekspertów. Subiektywne testy przeprowadzono zgodnie z założeniami metody oceny bezwzględnej jakości z ukrytym odniesieniem (ACR-HR — rozdział 3.2), kierując się wytycznymi zawartymi w normach ITU-T Rec. P.910 [Bibc] oraz ITU-R Rec. BT500 [Biba]. W trakcie każdej sesji testów uczestnicy oceniali jakość swobodnej nawigacji związanej tylko z jedną sekwencją. Na początku uczestnicy zostali poddani treningowi za pomocą dwóch wersji sekwencji: referencyjnej, czyli bez opóźnienia, oraz kotwicy, czyli sekwencji z dużą wartością opóźnienia wynoszącą 110 ms. Wartość ta odnosiła się zarówno do kierunku patrzenia, jak i pozycji. Kolejne sekwencje

były ocenianie przez użytkowników i prezentowane dla każdego uczestnika w losowej kolejności, co zostało przedstawione w tabeli 6.1.

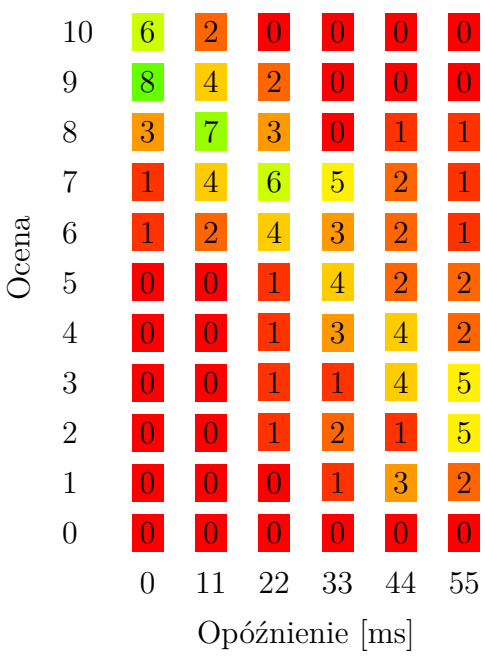
Każdą sekwencję, poza odniesieniem oraz kotwicą, pokazano w 11 wersjach opóźnienia, na które składała się jedna sekwencja bez opóźnień (ukryte odniesienie) oraz pięć wariantów opóźnienia dla każdego z dwóch testowanych scenariuszy. Uczestnicy zostali poproszeni o podanie wyniku w 11-punktowej skali MOS (ang. Mean Opinion Score): od 0 (bardzo złe odczucia subiektywne) do 10 (doskonałe odczucia subiektywne). Poruszali się oni w obrębie kwadratu 2 na 2 metry. W celu poprawy rzetelności badania usunięto wyniki, które uznano za błędne. Wartość MOS zadeklarowana przez uczestnika, który znacznie lepiej ocenił (2 lub więcej w 11-punktowej skali) sekwencję z większym opóźnieniem, była odrzucana. Dla wszystkich testowanych kategorii oraz wirtualnych sekwencji błędne pomiary były usuwane niezależnie. Dla każdej z badanych sekwencji uzyskano co najmniej 14 ważnych ocen, co zostało przedstawione w tabeli 6.2.

Tablica 6.2: Liczba ważnych ocen

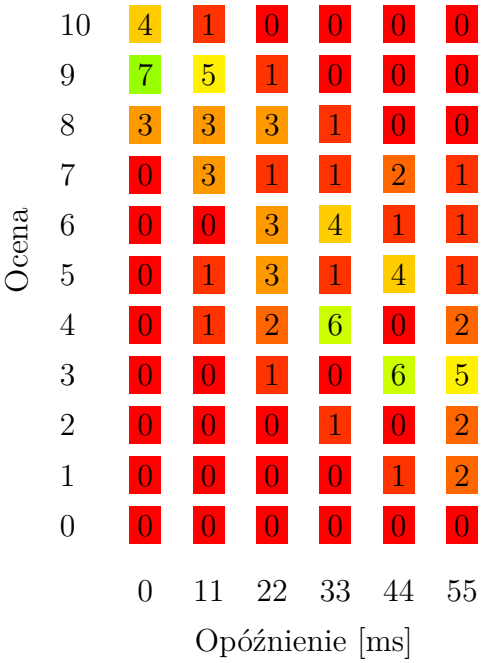
Scenariusz / Sekwencja	C2.1	C2.2
ClassroomVideo	19	23
TechnicolorMuseum	14	16
Poznań-People360	16	14

6.4 Wyniki

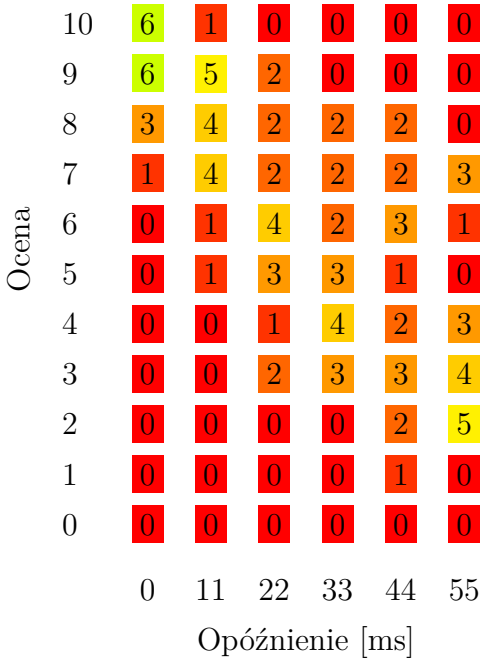
Na rysunku 6.6 przedstawiono rozkład ocen przeprowadzonych testów subiektywnych dla scenariusza C2.1, czyli systemów bez syntezy w terminalu użytkownika. Podczas eksperymentu dane do wyświetlacza nagłownego zostały wyliczone z opóźnieniem informacji o pozycji oraz kierunku patrzenia. Na wykresach kolor oznacza liczbę uczestników, którzy dokonali konkretnej oceny dla danego opóźnienia. Na rysunku 6.7 przedstawiono wykresy dla trzech badanych sekwencji, na których przedstawiono uśrednioną wartość MOS. Przedział ufności, który został również przedstawiony, wynosi 95%. Wyniki na każdy z wykresów dla opóźnienia 0 ms mają wysoką jakość (MOS ≥ 9) i monotonicznie maleją, a niska jakość (MOS ≤ 3) została osiągnięta dla opóźnienia około 55 ms.



(a) ClassroomVideo



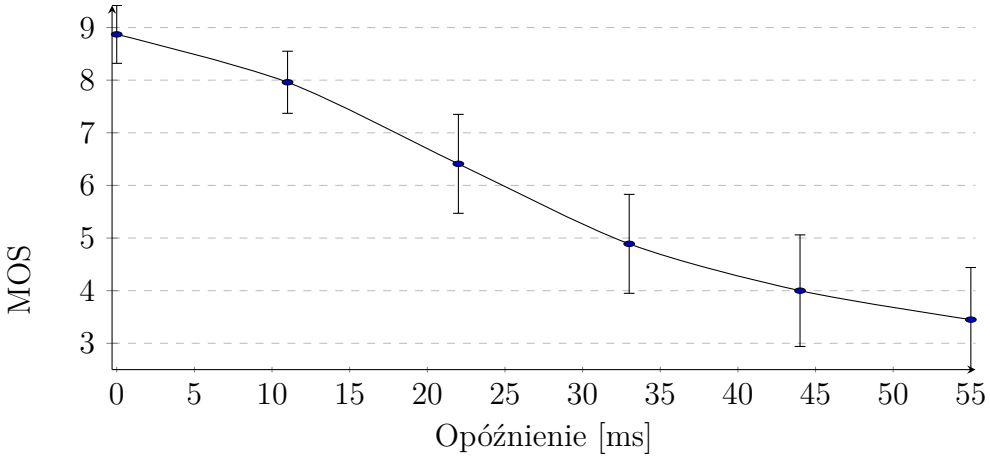
(b) TechnicolorMuseum



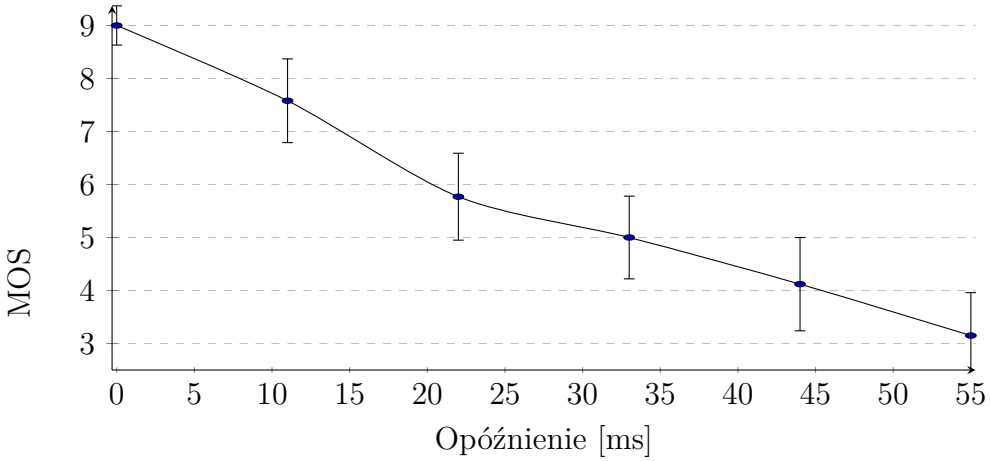
(c) Poznań-People360

(d) Skala kolorów odpowiadająca liczbie wystąpień danej oceny

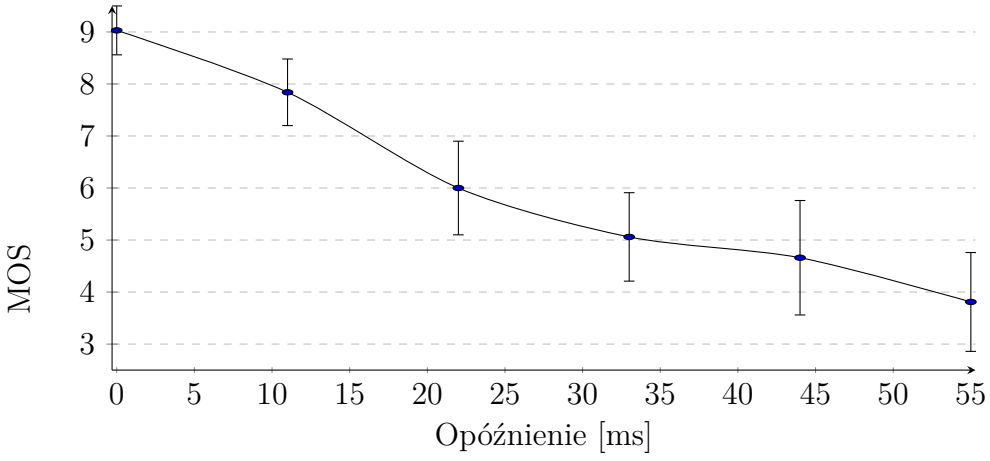
Rysunek 6.6: Rozkład ocen podczas testów subiektywnych. Scenariusz C2.1 — opóźniona pozycja i kierunek patrzenia



(a) ClassroomVideo



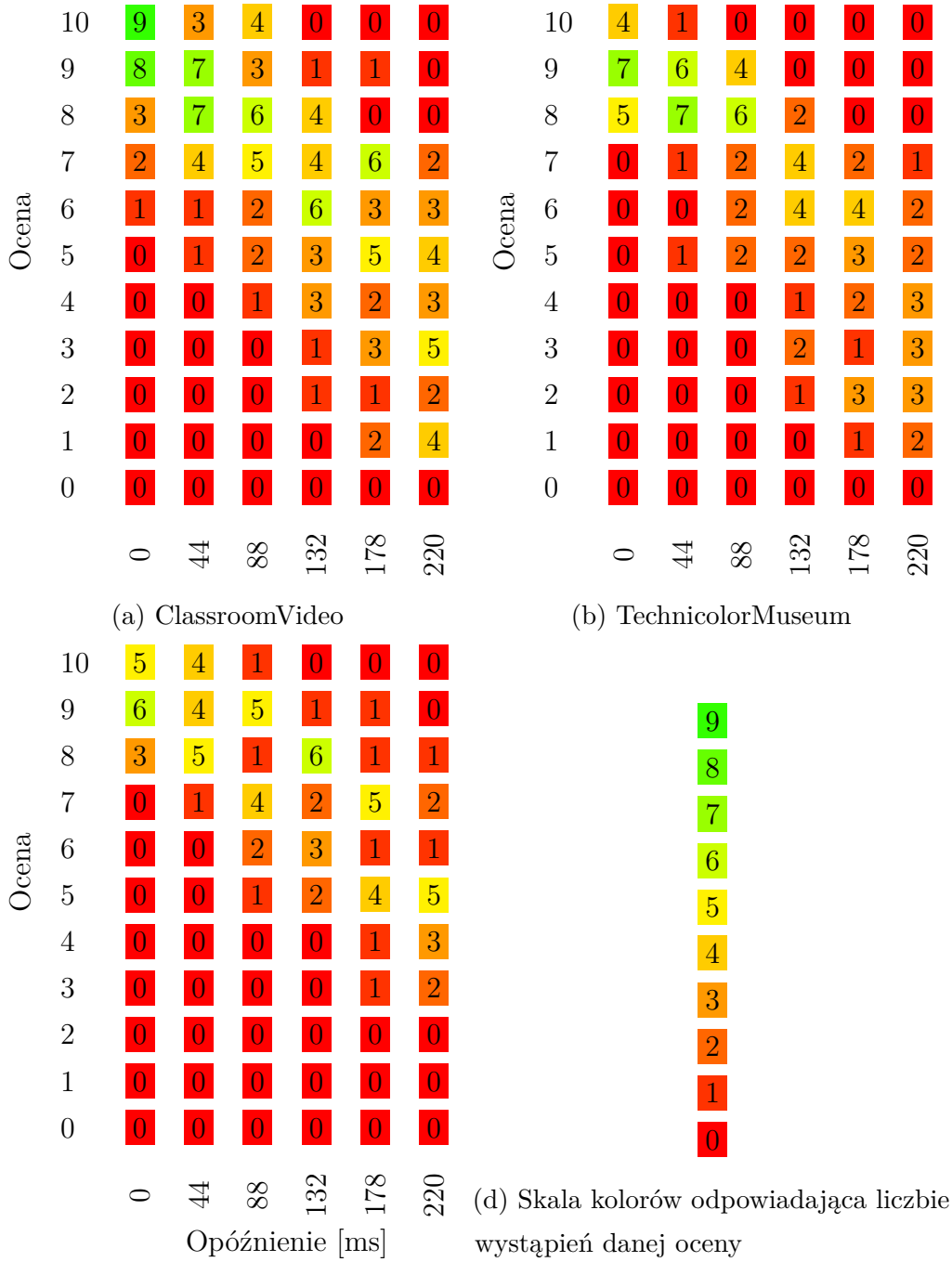
(b) TechnicolorMuseum



(c) Poznań-People360

Rysunek 6.7: Średni wynik oceny dla Scenariusza C2.1 z uwzględnieniem 95% poziomu ufności

Liczba uczestników, których oceny nie odrzucono i jest ona składową wyników zarówno dla scenariusza C2.1, jak i C2.2, została przedstawiona w tabeli 6.2. Na rysunkach 6.8 i 6.9 zaprezentowano analogiczne wykresy dla drugiego badanego scenariusza, czyli C2.2. Był on realizowany przez opóźnienie jedynie informacji o pozycji użytkownika. Rysunek 6.8 przedstawia mapę wyników jakości odczuwanej przez uczestników subiektywnych testów. Na rysunku 6.9 zostały przedstawione wykresy z uśrednionymi wartościami dla wszystkich ważnych pomiarów. Krzywa, prezentująca rezultaty uzyskane dla sekwencji Poznań-People360, posiada nieco większe wartości oceny MOS niż wyniki pozostałych sekwencji. Prawdopodobnie jest to związane ze stosunkowo dużą odległością obiektów od użytkownika. Wykresy mają podobny przebieg do wyników uzyskanych dla scenariusza C2.1, jednakże wartości czasu są znacznie większe, gdyż niska jakość (MOS ≤ 3) jest osiągana dla opóźnienia większego niż 220 ms.



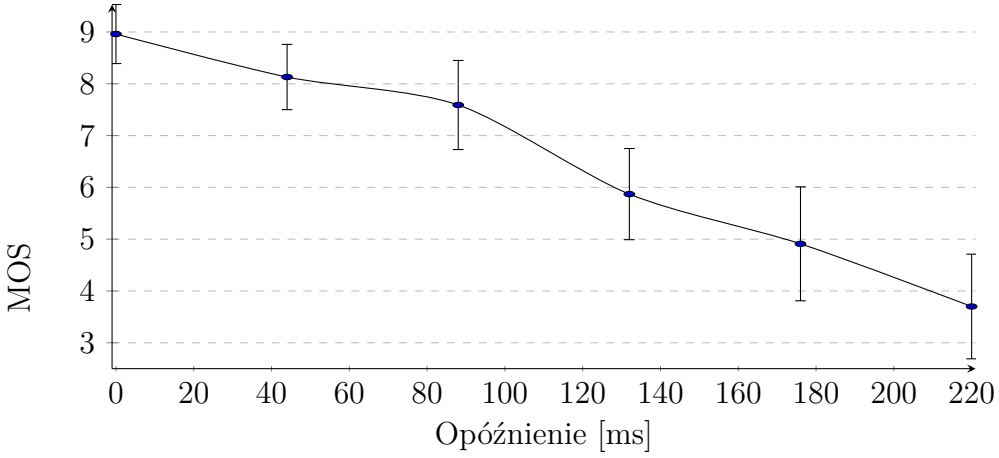
(a) ClassroomVideo

(b) TechnicolorMuseum

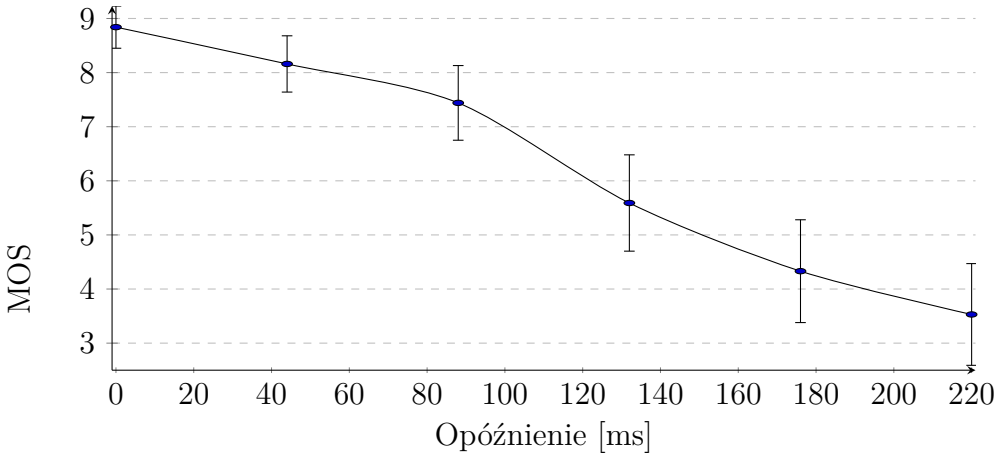
(c) Poznań-People360

(d) Skala kolorów odpowiadająca liczbie wystąpień danej oceny

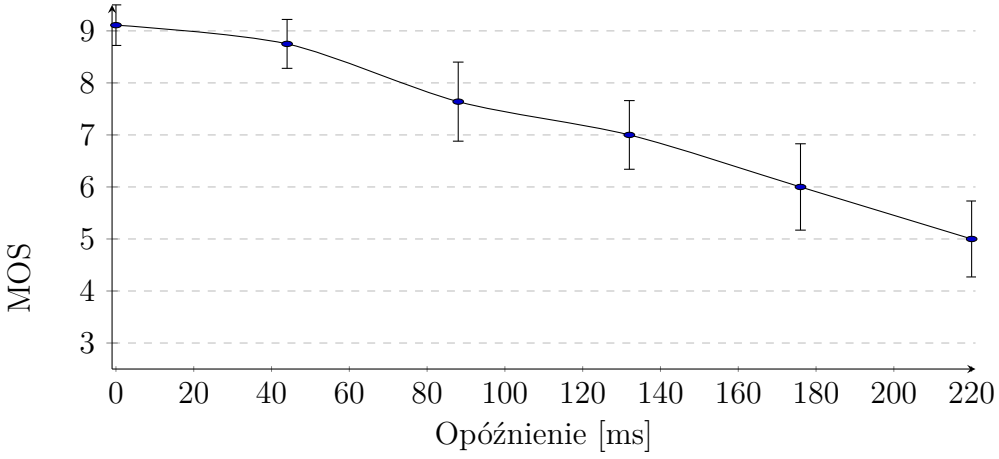
Rysunek 6.8: Rozkład ocen podczas testów subiektywnych. Scenariusz C2.2 — opóźniona pozycja



(a) ClassroomVideo



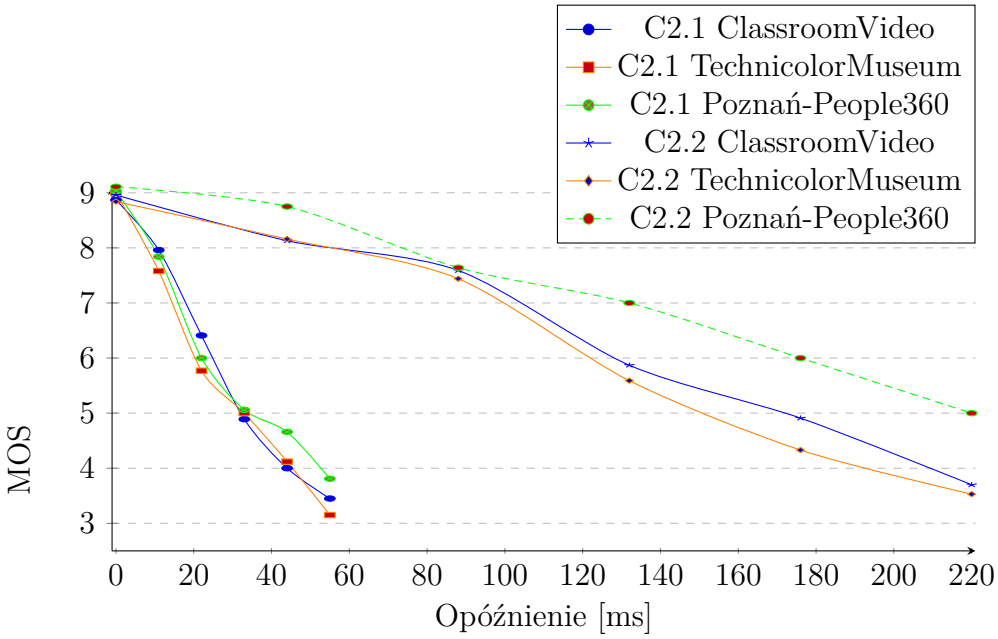
(b) TechnicolorMuseum



(c) Poznań-People360

Rysunek 6.9: Średni wynik opinii dla scenariusza C2.2 z uwzględnieniem 95% poziomu ufności

Na wykresie 6.10 zaprezentowano średnie wyniki oceny uczestników dla wszystkich badanych sekwencji i scenariuszy. Przedział ufności został pominięty w celu zachowania przejrzystości prezentowanych danych. Wszystkie wyrysowane krzywe dla badanego scenariusza mają zbliżony przebieg, a ich różnica nie jest większa niż 1 w 11-stopniowej skali MOS (z wyjątkiem C2.2 Poznań-People360). Można zatem założyć, że wyniki dla przebadanych scenariuszy mają uniwersalny charakter, który pozwala na sformułowanie ogólnych wniosków. Ludzka percepcja jest znacznie bardziej wrażliwa na opóźnienie kierunku patrzenia.



Rysunek 6.10: Średni wynik testów subiektywnych dla scenariusza C2.1 oraz C2.2

6.5 Podsumowanie

W rozdziale przedstawiono autorskie wyniki dotyczące wpływu opóźnienia transmisji dedykowanych danych do użytkownika z wyświetlaczem nagłownym. Celem eksperymentu było ocenienie subiektywnej jakości swobodnej nawigacji w wirtualnej scenie. Zidentyfikowano możliwe rozwiązania techniczne i na ich podstawie zaimplementowano symulator opóźnienia pozycji i kierunku patrzenia użytkownika w scenie, który został wykorzystany podczas testów subiektywnych.

Przeprowadzone eksperymenty doprowadziły do ciekawych wniosków. Po pierwsze, ludzka percepcja lepiej toleruje opóźnioną pozycję syntezowanego widoku, a jest znacznie wrażliwsza na opóźniony kierunek patrzenia użytkownika. Na przykład, aby osiągnąć dobrą jakość (wynik MOS równy 8), opóźnienie około 44 ms jest dopuszczalne w przypadku opóźnienia informacji o pozycji użytkownika - rysunek 6.10. Dla systemów, w których zarówno informacja o kierunku patrzenia, jak i pozycji jest opóźniona, wartość ta nie może być większa niż 11 ms. Po drugie, zwiększając opóźnienie, miara MOS zmniejsza się szybciej dla opóźnionej informacji o pozycji i kierunku patrzenia użytkownika. W przypadku scenariusza C2.1 z opóźnioną informacją o pozycji i kierunku patrzenia, dla opóźnienia dwa razy większego niż MOS 8, czyli 22 ms, miara ta wynosiła 6. W scenariuszu C2.2, w którym opóźniona była jedynie informacja o pozycji użytkownika, wartość MOS nadal przekraczała 7 dla opóźnienia 88 ms.

Podsumowując, przeprowadzone badania pokazały, iż transmisja syntezowanego widoku dookólnego do użytkownika sprawia, że subiektywna jakość wirtualnego widoku przy opóźnieniu 44 ms jest porównywalna z opóźnieniem 11 ms dla systemów zgodnych ze scenariuszem C2.1, w których synteza wirtualnego widoku jest realizowana tylko przez serwer. Proste przetwarzanie wizji przez użytkownika nie wymaga dużej mocy obliczeniowej, a może znacząco zwiększyć czas niezbędny na przygotowanie i transmisję danych przez serwer brzegowy. Należy również zauważyć, że scenariusze C2.3 lub C1, z pełną syntezą widoku, które bardziej obciążają terminal użytkownika, mogą zapewnić wysoką jakość usługi nawet wtedy, gdy opóźnienie jest znacznie większe niż 44 ms.

7 Eksperymentalny system swobodnej nawigacji

7.1 Wstęp

W tym rozdziale został przedstawiony eksperymentalny system swobodnej nawigacji. Składa się on z dwóch elementów: nowego wielokamerowego systemu akwizycji oraz serwera swobodnej nawigacji. System wielokamerowy jest kompletnym systemem, pozwalającym na rejestrowanie danych rzeczywistych, które po dodatkowych obliczeniach (niewykonywanych w czasie rzeczywistym, m.in. estymacji map głębi) są wykorzystywane przez serwer swobodnej nawigacji. Zapewnia on usługę swobodnego punktu widzenia, czyli przygotowuje wirtualne widoki, które są przesyłane przez sieć do terminali użytkowników.

Rzeczywiste sekwencje są najczęściej rejestrowane systemem wielokamerowym, który działa synchronicznie. Takie systemy nie są powszechnie dostępne, a ich budowa wymaga często specjalnych kamer oraz infrastruktury pozwalającej na ich synchronizację i zarządzanie [Fuj+06][FT04][Jia+18][Dom+14a][Guo+24]. Zarejestrowanie i przygotowanie kompletnej rzeczywistej sekwencji jest złożonym i skomplikowanym procesem, który wymaga również estymacji wewnętrznych i zewnętrznych parametrów. Proces ten, nazywany kalibracją systemu [Zha99][Dom10][Cyg02], jest bardzo wymagający i skomplikowany. Sprawia to, że wyznaczone parametry kamer oraz systemu są często obarczone błędem, który uniemożliwia estymację dokładnych map głębi oraz syntezę wirtualnych widoków o postrzeganej wysokiej jakości. Badania podjęte w niniejszej rozprawie związane są z systemami swobodnego punktu widzenia i ukierunkowanymi na wykorzystanie danych rzeczywistych, dlatego został w nich wykorzystany zbiór sekwencji testowych (rozdział 3.1), również zawierający takie dane. W rozdziale 7.2 autor zbudował nowy, praktyczny system wielokamerowy, który pozwolił na nagranie m.in. dodatkowej sekwencji wykorzystanej w pracy (Poznań Volleyball [Dom+18a]), a do jego budowy posłużyły tanie, konsumenckie kamery. Został on przedstawiony w rozdziale 7.2.

Przygotowaną kompletną sekwencję, czyli taką, która poza widokami z kamer zawiera m.in. parametry wewnętrzne i zewnętrzne, można dostarczyć do terminala użytkownika za pomocą serwera swobodnej nawigacji. Pełna architektura systemu umożliwia transmisje danych z wydarzeń rejestrowanych na żywo i została przedstawiona w rozdziale 2.2. Implementacja takiego serwera wymaga stworzenia dla

każdego elementu dedykowanego oprogramowania, które wykorzystuje odpowiednie techniki m.in transmisji, kompresji (rozdział 2.3), estymacji map głębi (rozdział 2.5) i syntezy widoku wirtualnego (rozdział 2.6). Implementacje algorytmów muszą działać również w czasie rzeczywistym, czyli z prędkością ramkową dostarczaną do terminala użytkownika. Poza tym kamery systemu muszą udostępniać rejestrowane dane do serwera reprezentacji na żywo. W ramach niniejszej pracy doktorskiej, w rozdziale 7.3, została przedstawiona uproszczona implementacja serwera swobodnej nawigacji.

7.2 System wielokamerowy

7.2.1 Założenia realizacji systemu wielokamerowego

Realizacja systemu wielokamerowego związana jest z m.in. wyborem kamer, ich liczbą, jednoczesnym zarządzaniem nimi oraz synchronizacją momentów akwizycji ramek.

Zagadnienie synchronizacji szczegółowo omówiono w rozdziale 2.8, z którego wynika, że system wielokamerowy wymaga synchronizacji sprzętowej. Taka synchronizacja może wykorzystywać karty sieciowe, wspierające precyzyjne protokoły czasu lub dedykowaną infrastrukturę synchronizacyjną. Specjalne przemysłowe albo profesjonalne kamery studyjne często umożliwiają taką synchronizację [Fuj+06][FT04][Dom+14a][Guo+24]. Dlatego systemy wielokamerowe zazwyczaj korzystają z tego typu kamer, czyli takich, które posiadają wejście synchronizacyjne, np. [wTM-1400][wDC-BGH1][wXHG1]. Najczęściej zapisują one widoki rzeczywiste w formacie bezstratnym.

Jak wykazano w eksperymencie przedstawionym w rozdziale 5.3, estymacja map głębi oraz synteza wirtualnego widoku o wysokiej jakości może odbywać się na podstawie stratnie skompresowanych sekwencji wizyjnych, o ile zapewniają one wysoką jakość, czyli mały stopień kwantyzacji. Kamery systemu mogą zatem wykorzystywać algorytmy stratnej kompresji danych. Poza tym, jak wykazano w kolejnych rozdziałach, 5.4, kamera może wspierać dowolną technikę kompresji, co zwiększa liczbę potencjalnych modeli kamer, z których może zostać zbudowany system.

Podsumowując, kamery systemu powinny:

- zapewniać synchronizację momentu akwizycji ramek,
- umożliwiać zarządzanie wszystkimi kamerami,

- zapewniać wysoką jakość rejestrowanych widoków.

Dodatkowym istotnym założeniem, które zostało przyjęte przy budowie systemu wielokamerowego, jest mały koszt jego realizacji. System ma wykorzystywać kamery z kompresją stratną, czyli tanie, konsumenckie i powszechnie dostępne.

7.2.2 Przegląd systemów

Tablica 7.1: Przegląd wielokamerowych systemów akwizycji [Guo+24][Duf17]

	Rok budowy	Liczba kamer	Typ kamer	Ustawienie kamer
Nagoya University	2005	100	JAI Pulnix TMC-1400CL	Liniowe/Łukowe
Fraunhofer HHI	2008	16	Hitachi HVF31CL-S1	Liniowe
Politechnika Poznańska	2009	9	Canon XH-G1 CCD	Liniowe
Hasselt University	2012	8	Basler avA1600-50gc	Liniowe
Hasselt University	2012	16	Prosilica GC	Dowolne
Politechnika Poznańska	2013	10	Canon XH-G1 CCD	Dowolne
Shanghai Jiao Tong University	2024	12	Panasonic DC-BGH1	Łukowe

Przegląd wybranych systemów wielokamerowych został umieszczony w tabeli 7.1. Wszystkie ze zaprezentowanych korzystają z studyjnych kamer profesjonalnych lub specjalnych przemysłowych, a dane są rejestrowane w formacie bezstratnym. Większość z nich składa się z nie więcej niż kilkunastu kamer, a ich ustawienie często nie może być modyfikowane.

Wśród kamer tańszych, nieprofesjonalnych i korzystających z algorytmów kompresji stratnej, w czasie budowy systemu (2016) pojawiły się na rynku kamery sportowe GoPro. Wybrane modele kamer tego producenta (m.in. GoPro Hero3 Black oraz GoPro Hero4 Black) mają dodatkowe złącze w tylnej części kamery (ang. HERO Port). Jest ono wykorzystywane do podłączania dodatkowych akcesoriów (tzw. plecaków kamer, ang. backpack), które pozwalają m.in. na podłączenie dodatkowej baterii, stworzenie pary z dwóch kamer [wGoProDH] oraz kamery dookólnej za pomocą sześciu kamer [wOmni]. Takie rozwiązania nie są dostępne w innych kamerach sportowych. Na rynku pojawiły się również akcesoria innych firm, które pozwalają na łączenie większej ilości kamer sportowych GoPro Hero, np. w celu stworzenia kamery dookólnej składającej się z większej ilości kamer [bullet360].

7.2.3 GoPro Hero4 Black

Do budowy systemu zostały wykorzystane kamery GoPro Hero4 Black [wGoPro]. Jest to jeden z dwóch modeli dostępnych konsumenckich kamer sportowych (w czasie budowy systemu — rok 2016 — drugi model to GoPro Hero3 Black), które posiadają dodatkowe złącze (ang. HERO Port). Pozostałe parametry GoPro Hero4 Black są następujące [wGoPro]:

- sensor typu CMOS,
- liczba elementów światłoczułych 12 MPix,
- maksymalny format ramki wizji 3840×2160,
- szybkość ramkowa dla rejestrowanej wizji w formacie 1920×1080: 24, 25, 30, 48, 50, 60, 80, 90, 120,
- czułość ISO 100, ISO 200, ISO 400, ISO 800, ISO 1600, ISO 6400,
- ustawienia balansu bieli 3000K, 5500K, 6500K,
- technika kompresji wizji: MPEG-4 Part 10, H.264 (AVC),
- wbudowany mikrofon,
- technika kompresji fonii MPEG-4 AAC,
- obudowa wodoodporna oraz wstrząsoodporna,
- karta SD z pamięcią do 64 GB,
- komunikacja bezprzewodowa za pomocą Bluetooth oraz Wi-Fi,
- wyjście HDMI,
- akcesoria podłączone dedykowanym złączem i pozwalające na synchronizację, sterowanie, zasilanie.

7.2.4 Architektura systemu akwizycji

Możliwość tworzenia z kamer GoPro Hero pary [wGoProDH] stała się podstawą do budowy systemu wielokamerowego. W takiej konfiguracji jedna z kamer staje się kamerą główną (ang. master), druga — podległą (ang. slave), która odbiera wszystkie informacje o konfiguracji oraz sygnały synchronizacji linii i ramki od kamery głównej.

Popularność modelu kamery GoPro HERO sprawiła, że pojawiły się informacje o wyprowadzeniach i komunikacji przez tylne złącze (ang. HERO Port) [wMewpro]. Informacje te umożliwiły zaprojektowanie nowego systemu wielokamerowego wykorzystującego ten model kamery. Pozwala on na:

- łączenie kamer w dużej odległości od siebie,

- zarządzanie i sterowanie wszystkimi kamerami,
- zasilanie kamer,
- bezprzewodowy dostęp do zarejestrowanych przez kamery plików,
- rejestrację fonii.

Składa się on z:

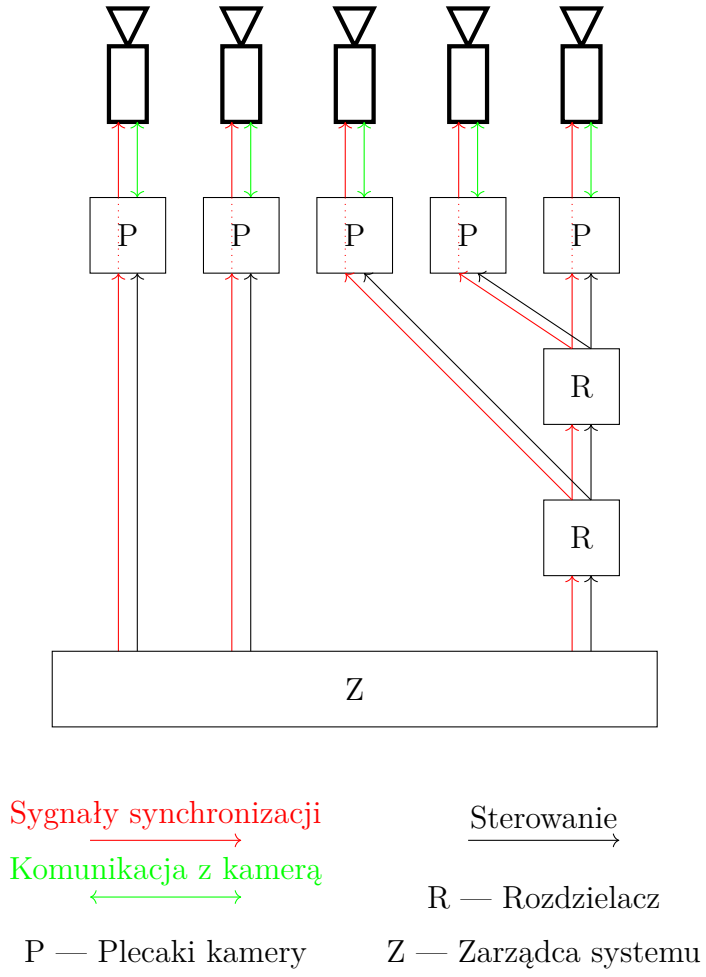
- kamer GoPro Hero4 Black,
- płytki rozszerzeń, plecaka (P),
- urządzenia będącego zarządcą systemu (Z),
- urządzeń rozdzielających sygnały sterujące i synchronizacyjne, rozdzielaczy (R),
- zasilaczy oraz okablowania sieciowego.

Schemat połączeń systemów sterującego i synchronizującego został przedstawiony na rysunku 7.1, a zdjęcia prototypów urządzeń ilustrują rysunki 7.2 oraz 7.3.

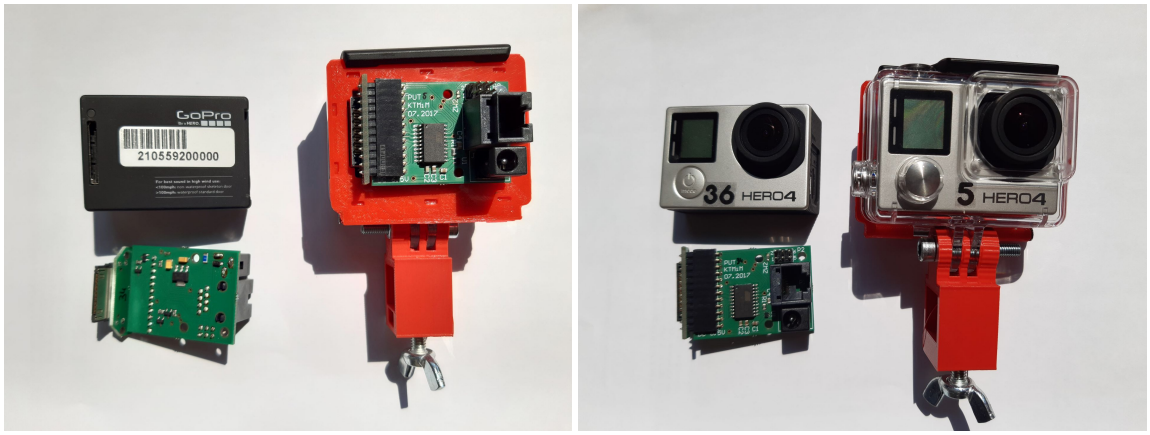
Zarządca systemu (**Z**) steruje pracą kamer oraz dostarcza dla nich sygnał synchronizacyjny. Stanowi on kamerą główną (ang. master) dla wszystkich pozostałych w systemie. Wysyła on informacje o zmianach w ustawieniach oraz sygnały synchronizacji linii i ramek w momencie uruchomienia nagrywania. Informacje te mogą być przesyłane bezpośrednio do kamer systemu albo pośrednio — z wykorzystaniem rozdzielaczy (**R**). Odpowiadają one za powielanie i retransmisje sygnałów sterujących i synchronizacyjnych. W celu zmniejszenie opóźniania czasu odpowiadającego czasowi przesyłania jednego bajtu został opracowany własny protokół komunikacji. Dzięki temu kamery systemu mogą być podłączone bezpośrednio do zarządcy systemu lub pośrednio, czyli przez dowolną liczbę rozdzielaczy, co zostało przedstawione na rysunku 7.1. Komunikacja od zarządcy do kamer jest komunikacją rozsiewczą bez sygnałów potwierdzenia. Do połączenia systemu zostały wykorzystane standardowe kable sieciowe oraz złącze RJ45, ale nie jest ona zgodna z komunikacją Ethernet ze względów na czas odpowiedzi i opracowany własny protokół transmisji. Plecaki kamery (**P**) odbierają przesyłane komunikaty oraz wpisują je do kolejki zadań, która jest realizowana niezwłocznie, kiedy jest taka możliwość, czyli kamera nie jest zajęta.

Taka sytuacja występuje dla każdej z kamer niezależnie, a dane przez nie rejestrowane są zapisywane na wewnętrznej karcie pamięci. Zdalny dostęp do nich jest możliwy przez sieć bezprzewodową. Każda z kamer posada kartę sieciową, a konfiguracja i informacje dostępowe do sieci Wi-Fi są umieszczone w specjalnym pliku.

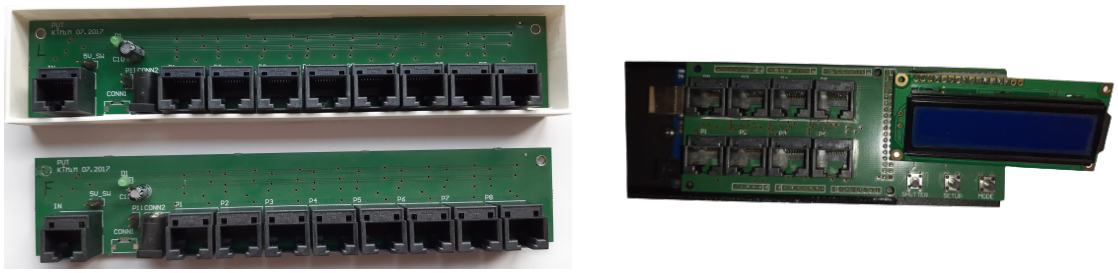
Przedstawiony w tym rozdziale system wielokamerowy składa się z 36 kamer GoPro Hero 4. Mogą one być dowolnie rozmieszczane przestrzennie, co wykorzystano w różnych praktycznych zastosowaniach opisanych poniżej.



Rysunek 7.1: Przykładowy schemat połączeń kamer GoPro



Rysunek 7.2: Kamera GoPro, dedykowany płytka tzw. plecak kamery, obudowa oraz uchwyt



Rysunek 7.3: Rozdzielacz sygnałów (R) oraz urządzenie zarządzające wszystkimi kamerami (Z)

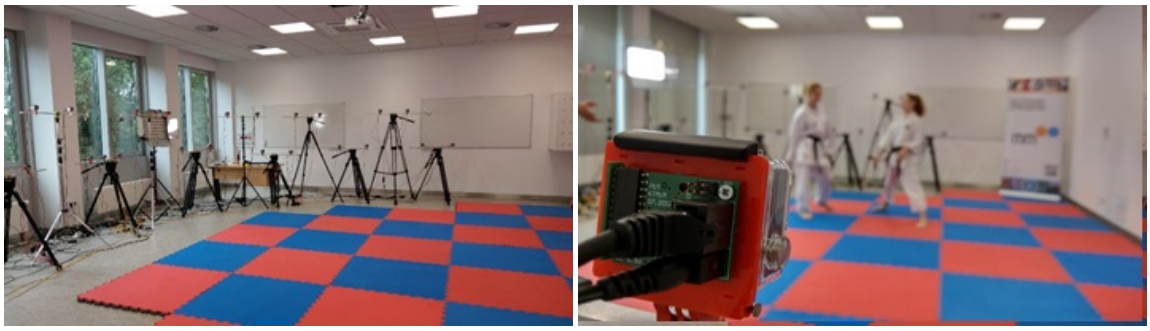
Rysunek 7.4: Kamery na hali sportowej — sekwencje — Poznań Volleyball [Dom+18a] oraz Basketball

7.2.5 Wykorzystanie systemu

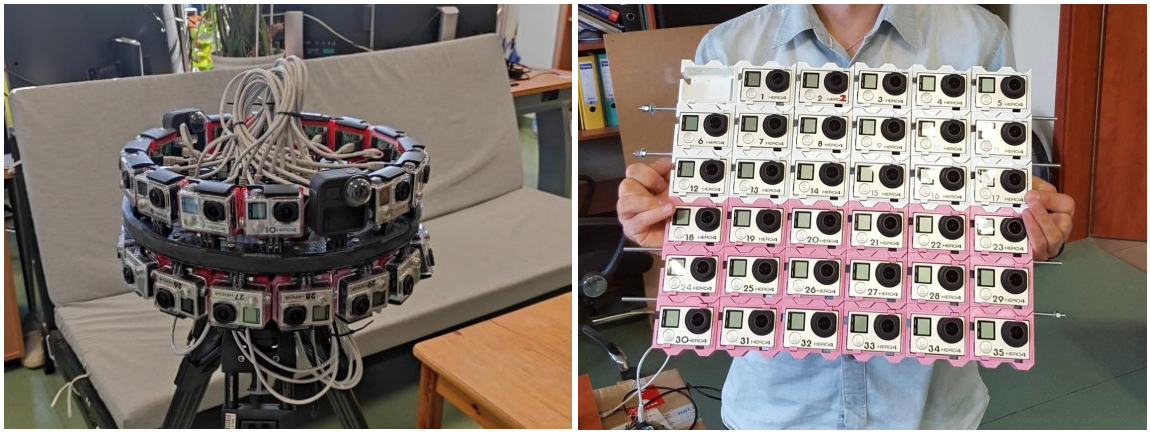
Przedstawiony system wielokamerowy został po raz pierwszy wykorzystany podczas nagrania w hali sportowej. Kamery zostały umieszczone wewnątrz stalowej ramy przykręconej do ścian pomieszczenia za pomocą zaprojektowanych uchwytów. Taka konfiguracja pozwoliła na jednoczesne nagranie sekwencji dla dwóch lub trzech systemów liniowych, ponieważ kamery były umieszczone wzdłuż dwóch lub trzech ścian budynku, tworząc literę L lub U. Uchwyty kamer oraz rama zapewniały możliwość modyfikowania ustawień oraz odporność systemu na czynniki zewnętrzne, takie jak np. uderzenia piłek. Przedstawiona sytuacja powodowałaby zmianę parametrów zewnętrznych i wymagałaby ponownego ich wyznaczania. Dzięki systemowi możliwe było nagranie sekwencji Poznań Volleyball [Dom+18a] oraz Basketball. Stalowa rama wzdłuż ścian budynku oraz kamery podczas nagrywania sekwencji zostały przedstawione na zdjęciach 7.4. Sekwencje te wykorzystano w dalszej części rozdziału przez serwer swobodnej nawigacji.

Sposób mocowania uchwytów został również przeprojektowany do ustawienia łukowego, umieszczonego na statywach na dwóch wysokościach. W badanym eksperymentalnym rozmieszczeniu kilka kamer było przymocowanych do jednego statywu za pomocą poprzecznej belki, która umożliwiała swobodną zmianę odległości pomiędzy kamerami. Powstała sekwencja jest udostępniona ekspertom grupy MPEG [Mie+23b].

System wielokamerowy został również wykorzystany w innych konfiguracjach do wytworzenia unikalnych wielokamerowych danych, m.in. w konfiguracji macierzowej i dookólnej, które zostały przedstawione na rysunku 7.6. Nagrane sekwencje są używane przez Instytut Telekomunikacji Multimedialnej Politechniki Poznańskiej.



Rysunek 7.5: Kamery w ustawieniu łukowym [Mie+23b]



Rysunek 7.6: Testowane ustawienia zestawu wielokamerowego GoPro

7.3 Serwer swobodnej nawigacji

7.3.1 Wstęp

Ogólna architektura serwera przedstawiona w rozdziale 2.2 umożliwia dostarczenie usługi swobodnego punktu widzenia z wydarzenia realizowanego na żywo. Implementacja takiego serwera wymagałaby systemu akwizycji, serwera reprezentacji oraz serwera brzegowego działającego w czasie rzeczywistym (ang. real-time). Ponieważ system wielokamerowy, który został zbudowany z konsumenckich kamer (rozdział 7.2), nie pozwala na transmisje danych z kamer na żywo, serwer swobodnej nawigacji będzie ograniczał się jedynie do serwera brzegowego, który korzysta z przygotowanych wcześniej danych. Takie założenie pozwala również na korzystanie z zaawansowanych algorytmów estymacji map głębi, które nie działają w czasie rzeczywistym, co wpływa pozytywnie na jakość syntezowanego wirtualnego widoku.

Implementacja serwera brzegowego systemu swobodnej nawigacji, który wysyła syntezowany widok do terminala użytkownika, składa się z trzech etapów. Pierwszy to odczyt właściwych danych z dysku, kolejny to synteza wirtualnego widoku, a ostatni stanowią kodowanie i transmisje danych.

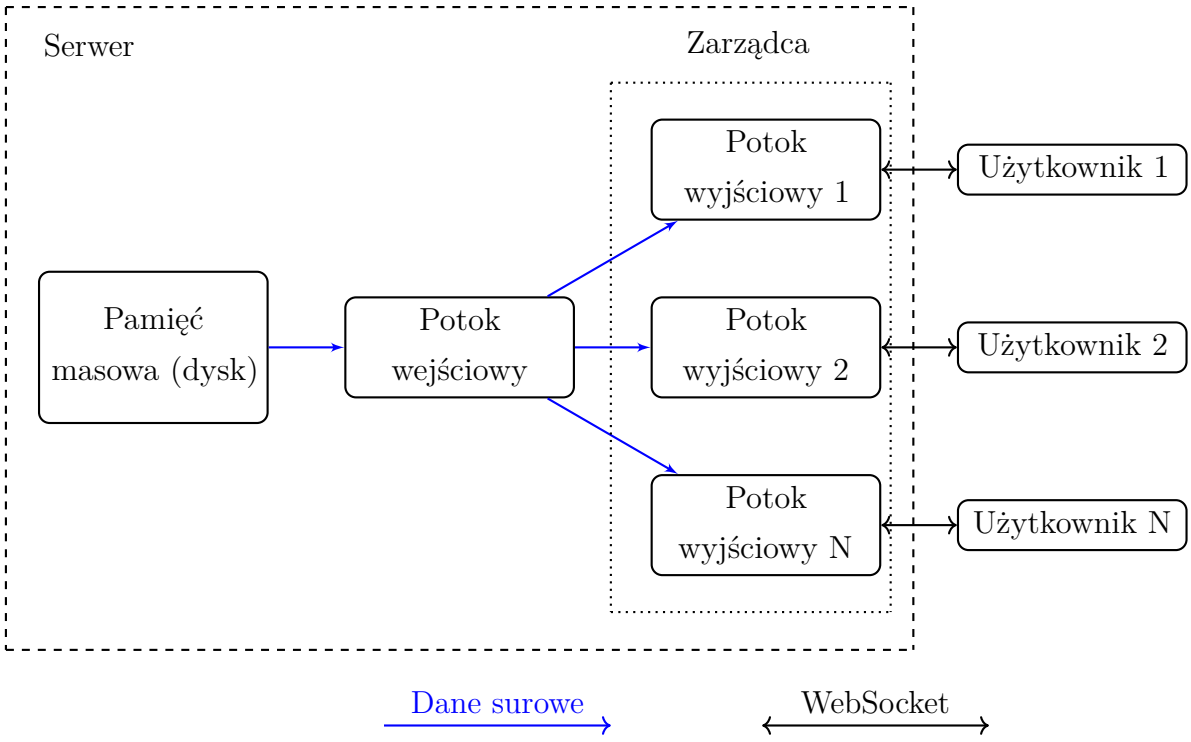
Zastosowane algorytmy syntezy wirtualnego widoku powinny zapewnić postrzeganą wysoką jakość przy stosunkowo małej złożoności obliczeniowej. Proces przetwarzania danych przez serwer wraz z czasem transmisji i dekodowania wirtualnego widoku przez użytkownika powinien być co najmniej krótszy niż czas wyświetlania ramki wizji. Implementacje algorytmów syntezy mogą używać m.in. zrównoleglania obliczeń, instrukcji wektorowych i procesora graficznego (rozdział 2.6). Kodowanie oraz transmisja wirtualnego widoku powinny wykorzystywać protokoły, oraz schematy kodowania zapewniające krótki czas dekodowania względem kodowania, czyli wykorzystywać protokoły i narzędzia stworzone do komunikacji w czasie rzeczywistym, takie jak np. RTP [CS03], WebRTC [VB16].

7.3.2 Założenia realizacji serwera swobodnej nawigacji

Realizacja serwera swobodnej nawigacji ma pozwalać na dostarczenie wirtualnego widoku do użytkownika, posługując się danymi zarejestrowanymi przez eksperymentalny system wielokamerowy. Założono wykorzystanie możliwie prostych rozwiązań i algorytmów, które mogą zapewnić wysoką jakość systemu. Implementacja serwera brzegowego powinna mieć złożoność obliczeniową, która pozwala na uruchomienie

z wykorzystaniem typowego komputera osobistego w czasie rzeczywistym. Jest konieczne, aby serwer obsłużył więcej niż jednego użytkownika, a zasoby (przepływność, zużywana moc obliczeniowa) powinny być dynamicznie przydzielane w zależności od liczby uczestników. Założono, że użytkownicy systemu będą mieli możliwość podłączenia się do serwera z wykorzystaniem różnych urządzeń bez instalacji dedykowanego oprogramowania.

7.3.3 Architektura serwera swobodnej nawigacji

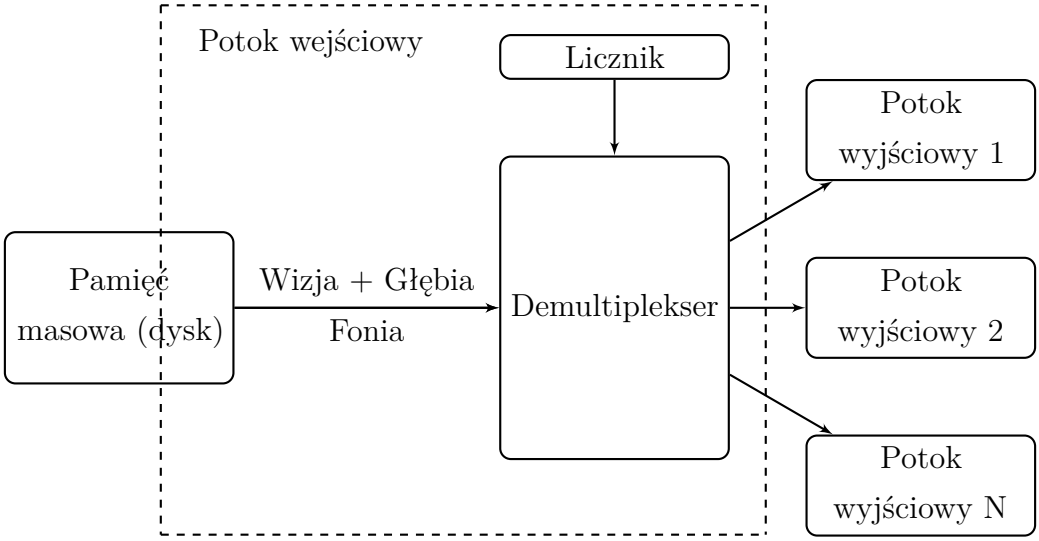


Rysunek 7.7: Architektura serwera swobodnej nawigacji

Propozycja architektury składa się z dwóch głównych części przetwarzających dane potokowo: wejściowej oraz wyjściowych. Oprogramowanie serwera jest modułowe i pozwala na jego dostosowanie do liczby użytkowników. Zaproponowana architektura została przedstawia na rysunku 7.7. Potok wejściowy jest odpowiedzialny za dostarczanie danych do wszystkich potoków wyjściowych (w formacie danych surowych, nieskompresowanych, ang. RAW), a potoki wyjściowe odbierają dane od podłączonego użytkownika i tworzą potok przetwarzający, który dostarcza wirtualny widok indywidualny. Komunikacja pomiędzy użytkownikiem a zarządcą odbywa się z wyko-

rzystaniem protokołu HTTP (pominięta na schemacie) oraz pozwala na utworzenie potoku wyjściowego dla nowego terminala użytkownika. Dane dla każdego z nich są przygotowywane przez dedykowany potok przetwarzania, a komunikacja odbywa się za pomocą protokołu WebSocket. Implementacja jest wielowątkowa i pozwala również na wielowątkowość w obrębie każdego z bloków [Sta+18d].

7.3.4 Potok wejściowy

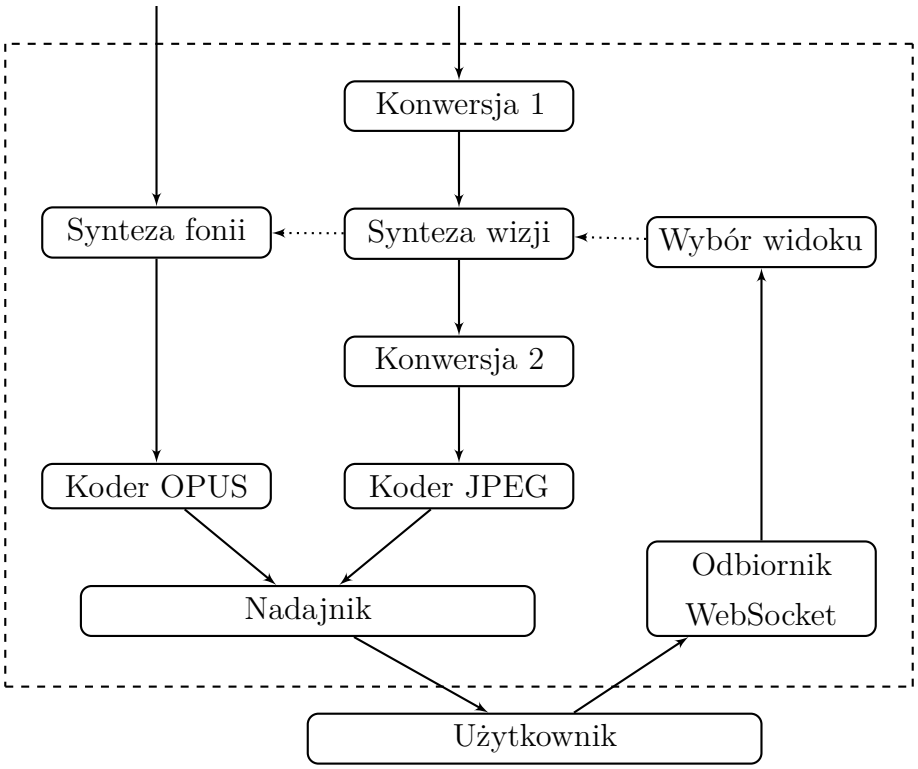


Rysunek 7.8: Budowa potoku wejściowego serwera swobodnej nawigacji

Potok wejściowy odczytuje dane źródłowe z dysku z prawidłową szybkością ramkową oraz powiela je dla potoków wyjściowych. Został on przedstawiony na rysunku 7.8. W celu umożliwienia równoległego przetwarzania dane są powielane i dostarczane niezależnie do potoków wyjściowych. Ich odczyt dotyczy widoków, które będą wykorzystywane przez syntezery w potokach wyjściowych. Pozwala to na zmniejszenie wymaganego transferu do dysku (sekwencja Poznań Volleyball [Dom+18a] składa się z 34 widoków). Odczyt odbywa się z wyprzedzeniem w celu wyeliminowania wpływu zmiennego czasu dostępu do plików. Dane są aktywowane przez precyzyjny licznik czasu i dostarczane do potoku wyjściowego. Demultiplekser rozdziela i powiela je (jeśli zachodzi taka konieczność) w wyznaczonych przez licznik chwilach. Widoki wraz z mapami głębi oraz dźwiękiem są uzupełniane o znacznik czasu oraz dodatkowe informacje, wykorzystywane przez syntezer takie jak, np.

parametry zewnętrzne i wewnętrzne. W ten sposób potok wejściowy wyznacza indywidualny kompletny zestaw danych dla każdego z potoków wyjściowych.

7.3.5 Potok wyjściowy



Rysunek 7.9: Budowa potoku wyjściowego serwera swobodnej nawigacji

Budowa potoku wyjściowego, który jest odpowiedzialny za syntezę wirtualnego widoku oraz komunikację z użytkownikiem, została przedstawiona na rysunku 7.9. Schemat przedstawia również bloki odpowiedzialne za przetwarzanie fonii, które nie są przedmiotem tej pracy.

Dane odbierane z potoku wejściowego poddawane są konwersji formatu, syntezie, kolejnej konwersji oraz kodowaniu techniką kompresji MJPEG [CSL12]. Fonia jest również syntezowana i kodowana koderem OPUS [Val+13]. Strumień danych z obydwu ścieżek jest transmitowany do użytkownika poprzez WebSocet i połączenie sieciowe. Informacja zwrotna o zmianie wirtualnego widoku jest odbierana od terminala i wpływa na miejsce syntezy kolejnej ramki obrazu.

Blok **Wybór widoku** zawiera metadane o liczbie i parametrach wszystkich kamer systemu. Znając pozycje użytkownika i kierunek patrzenia odbiorcy, wyznacza parametry zewnętrzne wirtualnego widoku oraz widoki rzeczywiste, z których będzie on syntezowany. Zaimplementowano algorytm [DSD18] pozwalający na uzyskanie wysokiej jakości przy ograniczonej liczbie wymaganych widoków, co pozwoliło na zmniejszenie złożoności obliczeniowej algorytmu syntezy oraz przepływności danych wymaganych do przesłania.

Obrazy wejściowe są przechowywane w formacie planarnym YCbCr 4:2:0 (wszystkie próbki składowych barwnych są umieszczone w jednym obszarze pamięci). Syntezer wymaga formatu przeplecionego 4:4:4 (próbki składowych barwnych są przeplecione z sobą w jednym obszarze pamięci), a koder — ponownie formatu planarnego. Konwersje te są realizowane odpowiednio przez bloki Konwersja 1 oraz Konwersja 2.

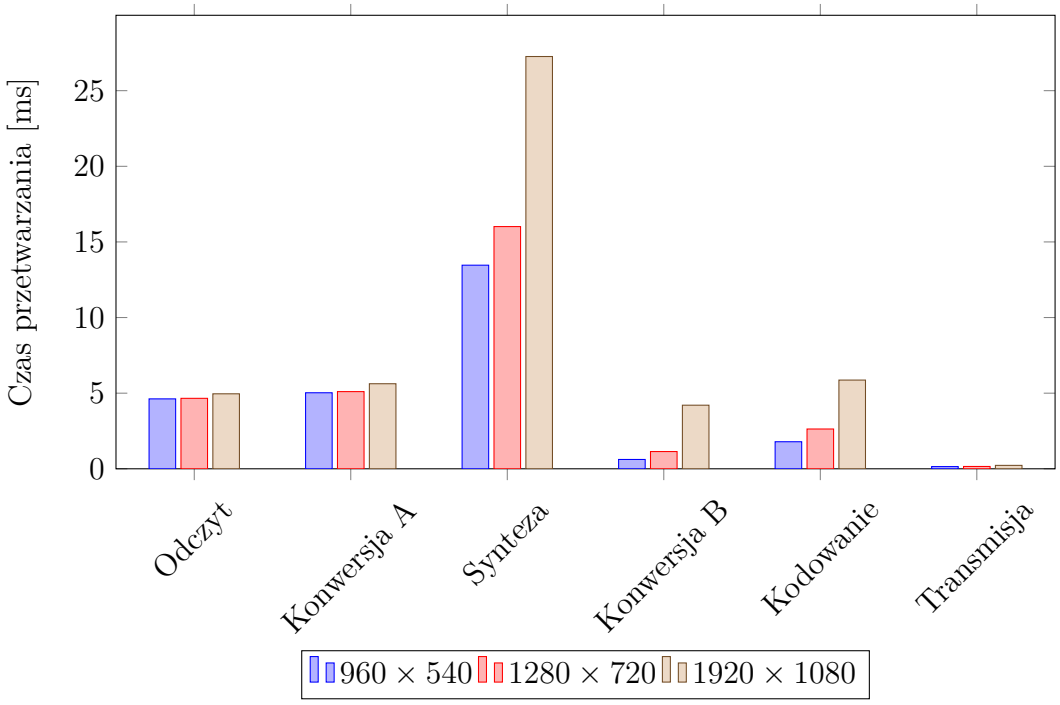
Synteza jest najbardziej złożoną operacją całego potoku przetwarzania. Zaimplementowany syntezer wymaga dwóch widoków wejściowych i towarzyszących im map głębi oraz dokonuje obliczeń na procesorze (CPU). Jest to zoptymalizowana szybka implementacja wykorzystująca sprzętowe przyspieszenia procesora (instrukcje SSIMD, AVX) oraz zrównoleglanie obliczeń [DS18].

Dane przed wysłaniem są kodowanie za pomocą kodera JPEG, który został wykorzystany do dostarczania wizji. Dokonuje on kompresji każdej z ramek niezależnie, co znacznie upraszcza kodowanie oraz dekodowanie. Strumień transmitowanych ramek jest również odporny na gubione pakiety i ramki z powodu braku predykcji, która występuje w metodach kompresji wizji. Przedstawione rozwiązanie skutkuje zwiększeniem ilości transmitowanych danych.

Do przesłania zakodowanego strumienia wizyjnego został wykorzystany transmiter WebSocket. Umożliwia on komunikację w dwie strony w trybie połączeniowym za pomocą protokołu TCP. Użytkownik za pomocą przeglądarki internetowej nawiązuje połączenia z serwerem. Całe oprogramowanie pobierane jest w formie strony internetowej i skryptów JavaScript.

Tablica 7.2: Uśredniony czas przetwarzania w serwerze

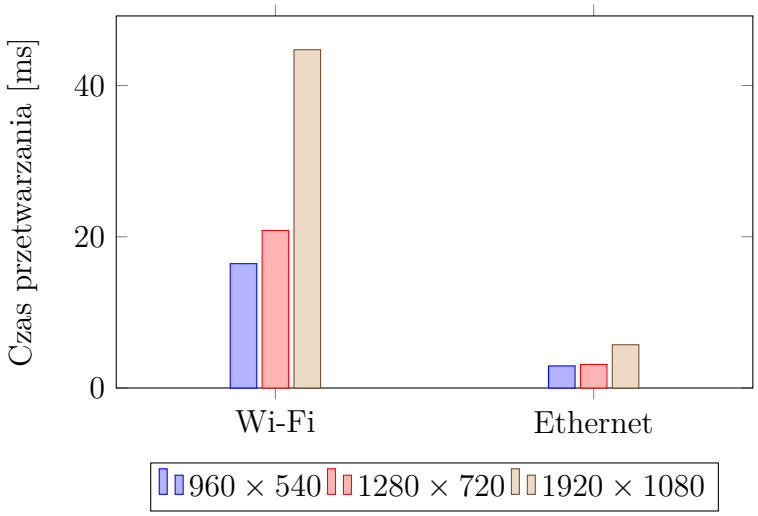
Rozmiar ramki	Średni czas przetwarzania [ms]							Maksymalna prędkość ramkowa
	Odczyt	Konwersja A	Synteza	Konwersja B	Kodowanie	Transmisja	Suma	
960 × 540	4,623	5,027	13,462	0,617	1,789	0,146	26,062	74,28
1280 × 720	4,659	5,103	16,015	1,140	2,631	0,154	30,295	62,44
1920 × 1080	4,958	5,624	27,259	4,205	5,866	0,226	50,742	36,68



Rysunek 7.10: Uśredniony czas przetwarzania w serwerze swobodnej nawigacji

7.3.6 Wyniki

Wyniki uśrednionych czasów przetwarzania danych zostały przedstawione w tabeli 7.2 i na wykresie 7.10. Do testów wykorzystano dwie wielowidokowe sekwencje Poznań Basketball oraz Poznań Volleyball [Dom+18a] o rozmiarze ramki w formacie HD i prędkości 25 ramek na sekundę. Obraz został zsyntezowany w 3 rozdzielczościach: oryginalnej (1920 × 1080), formacie HD-ready (1280 × 720) oraz formacie 1/4 HD (960 × 540). Dla każdej z nich wykonano pomiary czasów przetwarzania z wykorzystaniem procesora Intel Quad-Core 4,0 GHz, wykonanego w mikroarchitekturze Skylake.



Rysunek 7.11: Czas przesyłania danych do użytkownika

Tablica 7.3: Czas dostarczenia nowej ramki do użytkownika

Rodzaj połączenia	Rozmiar ramki	Średni czas przetwarzania [ms]	Średni czas przesyłania [ms]	Suma [ms]
Ethernet	960 × 540	26,062	2,919	28,981
	1280 × 720	30,295	3,111	33,406
	1920 × 1080	50,742	5,720	56,462
Wi-Fi	960 × 540	26,062	16,441	42,503
	1280 × 720	30,295	20,837	51,132
	1920 × 1080	50,742	44,726	95,468

Czasy transmisji danych przez sieć zostały zmierzone dla dwóch scenariuszy. Terminal użytkownika był połączony przez sieć lokalną z serwerem, korzystając z kablowego połączenia Ethernet (1000Base-T) lub połączenia bezprzewodowego Wi-Fi (IEEE 802.11n). Na wykresie 7.11 zaprezentowano uśrednione dane, przedstawiające czas transmisji pakietu w dwóch kierunkach (od serwera do terminala użytkownika oraz pakietu powrotnego, potwierdzającego odbiór danych).

W tabeli 7.3 przedstawiono sumę średnich czasów przetwarzania danych przez serwer w trzech badanych rozmiarach ramki oraz średni czas transmisji danych o położeniu i odbioru wirtualnego widoku w dwóch przedstawionych scenariuszach połączenia. Jest to uśredniony czas potrzebny na dostarczenie wirtualnego widoku do użytkownika.

7.3.7 Wnioski

W rozdziale została przedstawiona prosta realizacja serwera swobodnej nawigacji, zaimplementowanego w ramach niniejszej rozprawy doktorskiej. Dzięki zastosowaniu wydajnego wielowątkowego procesora i zoptymalozwanych technik syntezy oraz kodowania, serwer pozwala na dostarczenie wirtualnego widoku do użytkownika. Korzysta on z przygotowanych danych w postaci parametrów sceny, obrazów z kamer oraz wyliczonych map głębi. Czas reakcji systemu związany jest z rozmiarem ramki wirtualnego widoku oraz rodzajem połączenia terminalu użytkownika z serwerem. Jakość obrazu zależy od pozycji wirtualnego widoku, dokładności map głębi i algorytmu syntezy.

Przedstawiona implementacja, mimo wykorzystania prostych technik, pozwala na przygotowanie i dostarczenie nowego widoku w czasie krótszym niż czas prezentacji jednej ramki (40 ms dla 25 kl./s). W celu spełnienia tej wytycznej może zostać zastosowane również połączenie bezprzewodowe dla niskich rozdzielczości.

Zestawiając otrzymane wyniki z czasami transmisji wymaganymi dla użytkownika z wyświetlaczem nagłownym (rozdział 6), można stwierdzić, że zastosowane proste techniki i implementacje CPU nie wystarczą do zapewnienia postrzeganej wysokiej jakości dla użytkownika w hełmie HMD. Przedstawiony w rozdziale 6 (rysunek 6.2, strona 97) scenariusz C2.1, wymaga prezentacji aktualnego położenia w czasie około 11 ms. Czas syntezy, jaki jest niezbędny dla przedstawionej prostej implementacji serwera, jest większy, podobnie jak bezprzewodowa transmisja przez Wi-Fi. Niemniej modyfikując działanie serwera do realizacji według scenariusza C2.2 (rysunek 6.2), mógłby on zapewniać aktualny widok użytkownikowi korzystającemu z wyświetlacza nagłownego w czasie mniejszym niż 44 ms nawet przez Wi-Fi dla najmniejszego testowanego rozmiaru ramki (tabelka 7.3). Jednak jakość prezentowanego widoku w takim rozwiązaniu byłaby niska z uwagi na mały dostarczany rozmiar ramki, który w scenariuszu C2.2 jest widokiem dookólnym.

7.4 Podsumowanie

W rozdziale została przedstawiona architektura systemu wielokamerowego zbudowanego z kamer GoPro 4 Hero Black. Dzięki nim udało się stworzyć niedrogi, synchroniczny system nagrywania, który został wielokrotnie wykorzystany w różnych konfiguracjach do uzyskania unikalnych naturalnych sekwencji. Autor ma znaczny wkład w opracowanie koncepcji systemu, dyskusję alternatywnych wariantów, przyjęcie założeń i wytworzenie oraz testowanie prototypów, a później dedykowanych urządzeń: plecaków kamer, rozdzielaczy i zarządcy systemu. Brał również udział w nagraniach i przetwarzaniu sekwencji Poznań Volleyball [Dom+18a], Basketball oraz modyfikacji systemu do pracy w układach łukowych, macierzowych i dookólnych.

Przedstawiona implementacja serwera swobodnego punktu widzenia pozwala na wyznaczenie czasów potrzebnych na przygotowanie danych dla terminala użytkownika. Mimo zastosowania algorytmów w implementacjach CPU, prostego algorytmu kompresji oraz połączeniowego protokołu transmisji uzyskane czasy pozwalają na dostarczenie aktualnego widoku w czasie krótszym niż czas prezentacji jednej ramki nawet dla bezpołączeniowego bezprzewodowego. Implementacja mogłaby zostać również wykorzystana do dostarczenia treści do użytkownika z wyświetlaczem nagłownym. Należałoby wtedy zmienić wariant transmisji do scenariusza C2.2 z dodatkową prostą syntezą realizowaną przez urządzenie użytkownika. Niemniej przesłany mały rozmiar ramki wirtualnego widoku dookólnego, dodatkowo przetwarzany przez drugą syntezę, mocno ograniczałby jakość dostarczanej usługi.

Przedstawiona implementacja może stanowić podstawę do dalszych prac badawczych poprzez modyfikacje wykorzystanych technik oraz testowanie nowych algorytmów. Mogą one dotyczyć m.in przeniesienia części obliczeń na procesor graficzny (GPU, ang. Graphics Processing Unit), zastosowania wydajniejszych technik kodowania oraz protokołów transmisji dostosowanych do usług realizowanych w czasie rzeczywistym.Należy również zaznaczyć, że mimo wykorzystania prostych algorytmów, implementacja serwera jest możliwa i może on obsługiwać kilku użytkowników z wykorzystaniem wydajnego komputera klasy PC.

8 Podsumowanie

W pracy podjęto badania dotyczące postrzeganej jakości w systemach swobodnej nawigacji. Mają one zarówno charakter teoretyczny, eksperymentalny, jak i konstrukcyjny.

Rozważania teoretyczne dotyczyły opracowania modelu wpływu przysłonięć na jakość wirtualnego widoku (rozdział 4) oraz opracowania zależności jakości widoku wirtualnego w funkcji jakości widoków przysłanych (rozdział 5).

W rozdziałach 4, 5 oraz 6 opisano czasochłonne eksperymenty przeprowadzone przez autora z wykorzystaniem zbioru sekwencji testowych. Pozwoliły one na walidację modeli teoretycznych przedstawionych w rozdziale 4. Prace eksperymentalne przedstawione w rozdziale 5 były podstawą do zaproponowania modelu wpływu jakości przesyłanego widoku na jakość wirtualnego widoku. Ponadto, przeprowadzono testy subiektywne z wykorzystaniem wyświetlacza nagłownego w celu zbadania wpływu opóźnień w systemie swobodnego punktu widzenia. Eksperyment ten został przedstawiony w rozdziale 6.

W rozdziale 7 pracy przedstawiono nowy, kompletny system wielokamerowy oraz serwer swobodnej nawigacji, które stanowią osiągnięcie konstrukcyjne.

Badania dotyczące różnych elementów systemów wizyjnych swobodnego punktu widzenia pozwoliły na sformułowanie czterech niezależnych tez badawczych, które zostały przedstawione w rozdziale 1.3.2. W pracy zaprezentowano wyniki wspierające te tezy w następujący sposób:

Teza 1: Rejestracja złożonych scen wizualnych przy ustawieniu kamer wokół sceny parami o małej odległości bazowej umożliwia uzyskanie syntezy wirtualnych widoków o lepszej jakości, niż gdy ta sama liczba kamer rzeczywistych jest rozmieszczona równomiernie wokół sceny.

W rozdziale 4 przedstawiono model teoretyczny, który opisuje zjawiska wpływające na jakość syntezy widoków wirtualnych w zależności od złożoności sceny. Wyniki tego modelu zostały porównane z wynikami eksperymentu przeprowadzonego za pomocą wielowidokowych sekwencji testowych. Prezentowane rezultaty pozwalają na wybór ustawienia kamer w zależności od liczby przysłoniętych punktów obrazów dla kamer ustawionych równomiernie. Przez złożoną

scenę rozumiemy taką, w której ponad 25% punktów obrazu jest przysłoniętych dla kamer ustawionych równomiernie. Wykazano, że dla takich scen powinno się ustawiać kamery parami, aby uzyskać wyższą postrzeganą jakość. Przedstawione badania pokazują, że średnia jakość wirtualnych widoków mierzona metryką PSNR, może zmienić się nawet o 3 dB, w zależności od ustawienia kamer.

Teza 2: Jakość obrazów wirtualnych uzyskiwanych ze zdekodowanych widoków zależy silnie od jakości tych widoków i w stosunkowo niewielkim stopniu od zastosowanej techniki kompresji stratnej.

W rozdziale 5 przedstawiono obszerne badania dotyczące jakości widoków wirtualnych w systemach wykorzystujących stratnie skompresowane widoki do estymacji map głębi. Przedstawione wyniki w eksperymencie wstępnym udowadniają, że stratna kompresja zachowująca wysoką jakość widoków przesyłanych nie wpływa znacząco na jakość wirtualnych widoków. Eksperymenty zasadnicze wykazały, że jakość wirtualnego widoku zależy głównie od jakości przesyłanych widoków i w stosunkowo niewielkim stopniu od zastosowanej techniki kompresji. Zaproponowany liniowy model tej zależności pozwala na oszacowanie jakości wirtualnego widoku niezależnie od techniki kompresji. Średnia wartość błędu tego modelu jest zazwyczaj mniejsza niż 0,5 dB.

Teza 3: W systemach wirtualnej rzeczywistości z wyświetlaczami nagłownymi dokuczliwość opóźnień pomiędzy ruchem użytkownika a odświeżeniem ekranu jest większa dla ruchu rotacyjnego niż dla ruchu translacyjnego.

Rozdział 6 przedstawia analizę zaproponowanych scenariuszy transmisji danych do terminala użytkownika, dla których przeprowadzono testy subiektywne z wykorzystaniem wyświetlacza nagłownego. Badania pozwalają stwierdzić, że transmisja widoku dookólnego do terminala użytkownika pozwala na kilkukrotne zwiększenie czasu wymaganego do komunikacji z serwerem bez zmiany jakości subiektywnej. Porównywany system to taki, w którym tylko serwer realizuje syntezę wirtualnego widoku, a transmitowany widok jest bezpośrednio

wyświetlany w wyświetlaczu nagłownym.

Teza 4: Za pomocą konsumenckich kamer i typowego komputera osobistego można zbudować system wirtualnej rzeczywistości działający w czasie rzeczywistym, przy założeniu, że model sceny przestrzennej jest przygotowywany wcześniej.

W rozdziale 7 przedstawiono kompletny system wielokamerowy zbudowany z konsumenckich kamer oraz sprawnie działający system swobodnej nawigacji wykorzystujący wcześniej przygotowane widoki oraz mapy głębi. System ten został wielokrotnie wykorzystany m.in do nagrań sekwencji badawczych, które zwiększyły ilość naturalnych danych, udostępnionych grupom roboczym ISO/IEC MPEG. Serwer swobodnej nawigacji, mimo wykorzystania prostych algorytmów kompresji i komunikacji, pozwala na sprawną komunikację z terminalem użytkownika.

Podsumowując, przedstawione w pracy wyniki badawcze pozwalają uznać tezy za udowodnione.

Ważniejsze osiągnięcia przedstawione w pracy doktorskiej:

- Opracowanie modelu teoretycznego wpływu przysłonięć w scenie na jakość wirtualnego widoku (rozdział 4.5).
- Opracowanie i przeprowadzenie eksperymentu (rozdział 4.7), który potwierdza poprawność dwóch teoretycznych modeli wpływu przysłonięć w scenie oraz odległości pomiędzy kamerami na jakość wirtualnego widoku.
- Udowodnienie (rozdział 5.3), że zachowująca wysoką jakość kompresja widoków przed estymacją map głębi nie wpływa znacząco na jakość wirtualnego widoku, co pozwala na jej stosowanie w systemach swobodnego punktu widzenia.
- Przedstawienie liniowego modelu (rozdział 5.5) wpływu jakości przesyłanych widoków na wirtualny widok. Jakość wirtualnego widoku zależy głównie od jakości przesyłanych widoków i w stosunkowo niewielkim stopniu od zastosowanej techniki kompresji.

- Udowodnienie za pomocą testów subiektywnych (rozdział 6.3), że scenariusz transmisji ukierunkowany na minimalizację opóźnienia ruchu rotacyjnego pozwala na zachowanie takiej samej postrzeganej jakości nawet przy znacznie dłuższym czasie transmisji pomiędzy terminalem użytkownika z wyświetlaczem nagłownym a serwerem.
- Budowa systemu wielokamerowego (rozdział 7.2) oraz implementacja serwera swobodnej nawigacji (rozdział 7.3), które pozwalają na dostarczanie wirtualnych widoków do terminali kilku użytkowników.

W celu przeprowadzenia badań zawartych w rozprawie autor pracy stworzył różne moduły oprogramowania, głównie w językach C, C++ oraz Python. Stanowiło to podstawę do przeprowadzenia wielu długotrwałych i złożonych eksperymentów. Polegały one na przetwarzaniu wizji wielowidokowej (rozdział 4 oraz 5), autorskich modyfikacjach oprogramowań związanych z prezentacją wizji użytkownikowi w wyświetlaczu nagłownym (rozdział 6) oraz implementacji oprogramowania serwera swobodnej nawigacji (rozdział 7). Autor zaprojektował płytki PCB niezbędne do budowy prototypów urządzeń systemu wielokamerowego, przygotował oprogramowanie potrzebne do poprawnego działania wykorzystanych układów cyfrowych, zaprojektował obudowy i mocowania systemu wielokamerowego dla różnych jego ustawień. Ponadto autor współtworzył metrykę oceny jakości IV-PSNR [Dzi+22], która jest metryką dostosowaną dla obrazów syntezowanych wykorzystywaną m.in. przez grupy robocze MPEG działające w ramach ISO/IEC JTC1/SC29.

9 Lista publikacji naukowych autora

Publikacje w czasopismach o zasięgu międzynarodowym

- A. Dziembowski, D. Mieloch, J. Stankowski i A. Grzelka. "IV-PSNR—The Objective Quality Metric for Immersive Video Applications". W: *IEEE Transactions on Circuits and Systems for Video Technology* 32.11 (2022), s. 7575–7591. DOI: 10.1109/TCSVT.2022.3179575

- O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch i J. Samelak. "A Free-Viewpoint Television System for Horizontal Virtual Navigation". W: *IEEE Transactions on Multimedia* 20.8 (2018), s. 2182–2195. ISSN: 1520-9210. DOI: 10.1109/TMM.2018.2790162

- D. Mieloch i A. Grzelka. "Segmentation-based Method of Increasing The Depth Maps Temporal Consistency". W: *International Journal of Electronics and Telecommunications* vol. 64.No 3 (2018). DOI: 10.24425/123521

Pozostałe artykuły w czasopismach

- K. Wegner, J. Stankowski, K. Klimaszewski, O. Stankiewicz, A. Grzelka, M. Lorkiewicz, H. Żabiński i T. Grajek. "MUCHA – System Rejestracji I Przetwarzania Obrazu Przestrzennego". W: *Twierdza* 87.1 (2018), s. 48–54. ISSN: 1507-6474

- A. Dziembowski, A. Grzelka i D. Mieloch. "Zwiększanie Rozdzielczości Obrazu i Mapy Głębi w Celu Poprawy Jakości Syntezy Widoków Wirtualnych". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* 2017 (czer. 2017). DOI: 10.15199/59.2017.6.57

- D. Mieloch, A. Dziembowski i A. Grzelka. "Estymacja Głębi Dla Systemów Wielowidokowych". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* 2017 (czer. 2017). DOI: 10.15199/59.2017.6.75

- A. Dziembowski, A. Grzelka i D. Mieloch. "Wielowidokowa Synteza w Systemach Telewizji Swobodnego Punktu Widzenia". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* (czer. 2016), s. 233–236. DOI: 10.15199/59.2016.6.14

- D. Mieloch, A. Dziembowski i A. Grzelka. "Segmentacja Obrazu w Estymacji Map Głębi". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* (czer. 2016), s. 241–244. DOI: 10.15199/59.2016.6.16

- J. Borko, G. Dulnik, A. Grzelka, A. Łuczak i A. Paszkowski. "A Parametric Synthesizer of Audio Signals on FPGA". W: *Measurement Automation Monitoring* Vol. 61, No. 7 (2015), s. 367–369. ISSN: 2450-2855

- J. Borko, G. Dulnik, A. Grzelka, A. Łuczak i A. Paszkowski. "A Remote Programming Module of FPGA Boards". W: *Measurement Automation Monitoring* Vol. 61, No. 7 (2015), s. 373–375

- G. Dulnik, A. Grzelka i A. Łuczak. "A Gigabit Ethernet Interface With an Embedded Lossless Data Encoder on FPGA". W: *Measurement Automation Monitoring* Vol. 61, No. 7 (2015), s. 364–366

Artykuły w materiałach konferencji międzynarodowych indeksowanych w Web of Science oraz IEEE Xplore

- A. Grzelka, A. Dziembowski, D. Mieloch i M. Domański. "The Study of the Video Encoder Efficiency in Decoder-Side Depth Estimation Applications". W: *30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision : WSCG*. 2022, s. 248–255. DOI: 10.24132/CSRN.3201.31

- A. Grzelka, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski i M. Domański. "Impact of Video Streaming Delay on User Experience with Head-Mounted Displays". W: *2019 Picture Coding Symposium (PCS)*. 2019, s. 1–5

- M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, D. Mieloch, R. Rataiczak, O. Stankiewicz, J. Stankowski i K. Wegner. "Real-Time Virtual Navigation Provision by Simple Means". W: *2018 International Conference on Signals and Electronic Systems (ICSES)*. 2018, s. 69–73

- J. Stankowski, A. Grzelka, D. Mieloch i K. Wegner. "Processing Pipeline for Real-Time Remote Delivery of Virtual View in FTV Systems". W: *2018*

International Conference on Signals and Electronic Systems (ICSES). 2018, s. 118–123

- D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz i M. Domański. "Temporal Enhancement of Graph-Based Depth Estimation Method". W: *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2017, s. 1–4

- A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz i M. Domański. "Enhancing View Synthesis With Image and Depth Map Upsampling". W: *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2017, s. 1–4

- D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz i M. Domański. "Graph-Based Multiview Depth Estimation Using Segmentation". W: *2017 IEEE International Conference on Multimedia and Expo (ICME)*. 2017, s. 217–222. DOI: 10.1109/ICME.2017.8019532

- M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, K. Klimaszewski, D. Mieloch, R. Ratajczak, O. Stankiewicz, J. Siast, J. Stankowski i K. Wegner. "Demonstration of a Simple Free Viewpoint Television System". W: *2017 IEEE International Conference on Image Processing (ICIP)*. 2017, s. 4589–4591

- M. Domański, A. Dziembowski, A. Grzelka i D. Mieloch. "Optimization of Camera Positions for Free-Navigation Applications". W: *2016 International Conference on Signals and Electronic Systems (ICSES)*. 2016, s. 118–123. DOI: 10.1109/ICSES.2016.7593833

- M. Domański, M. Bartkowiak, A. Dziembowski, T. Grajek, A. Grzelka, A. Łuczak, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. "New Results in Free-Viewpoint Television Systems for Horizontal Virtual Navigation". W: *2016 IEEE International Conference on Multimedia and Expo (ICME)*. 2016, s. 1–6. DOI: 10.1109/ICME.2016.7552993

- A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner i M. Domański. "Multiview Synthesis — Improved View Synthesis for Virtual Navigation". W: *2016 Picture Coding Symposium (PCS)*. 2016, s. 1–5. DOI: 10.1109/PCS.2016.7906380

- A. Dziembowski, M. Domański, A. Grzelka, D. Mieloch, J. Stankowski i K. Wegner. "The Influence of a Lossy Compression on the Quality of Estimated Depth Maps". W: *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2016, s. 1–4. DOI: 10.1109/IWSSIP.2016.7502730

- A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz i M. Domański. "Depth Map Upsampling and Refinement for FTV Systems". W: *2016 International Conference on Signals and Electronic Systems (ICSES)*. 2016, s. 89–92

- M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, Ł. Kowalski, M. Kurc, A. Luczak, D. Mieloch, R. Ratajczak, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. "Methods of High Efficiency Compression for Transmission of Spatial Representation of Motion Scenes". W: *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 2015, s. 1–4

Pozostałe publikacje

- D. Mieloch, A. Dziembowski, B. Szydełko, D. Klóska, A. Grzelka, J. Stankowski, M. Domański, G. Lee i J. Jeong. *[MIV] New natural content – MartialArts*. ISO/IEC JTC1/SC29/WG4 MPEG dok. m61949, OnLine. Sty. 2023

- A. Grzelka D. Mieloch A. Dziembowski. *[MIV] Exploration Experiments on Future MIV: PUT results*. ISO/IEC JTC1/SC29/WG4 MPEG dok. m58835, OnLine. Sty. 2022

- A. Grzelka, A. Dziembowski, D. Mieloch, M. Domański, G. Lee i J. Jeong. *MIV CE2.11: Encoder-side rendering*. ISO/IEC JTC1/SC29/WG4 MPEG dok. m56369, OnLine. Kw. 2021

- A. Dziembowski, D. Mieloch, A. Grzelka, J. Stankowski, M. Domański, G. Lee i J. Jeong. *CE3-related: Color-based patch splitting and sorting*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m53408, Alpbah, Austria. Czer. 2020

- A. Dziembowski, D. Mieloch, A. Grzelka, J. Stankowski, M. Domański i G. Lee. *Immersive Video CE3-related: Spatio-temporal patch redundancy removal*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m51604, Bruksela, Belgia. Sty. 2020

- A. Dziembowski, D. Mieloch, A. Grzelka, J. Stankowski, M. Domański i G. Lee. *Immersive Video CE3.2: Temporal Patch Redundancy Removal*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m51603, Bruksela, Belgia. Sty. 2020

- A. Dziembowski, D. Mieloch, A. Grzelka, J. Stankowski, M. Domański i G. Lee. *Immersive Video CE3.1: Patch Splitting*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m51602, Bruksela, Belgia. Sty. 2020

- A. Grzelka, M. Lorkiewicz, A. Dziembowski i D. Mieloch. *PUT Proposal of PoznanFencing Posetraces*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m50646, Genewa, Szwajcaria. Paź. 2019

- M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, J. Y. Jeong i J. Seo. *Call for Proposals on 3DoF+: Aggregation of the Results of Submission Assessments Obtained by the Subjective Tests With Naïve Subjects*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m47985, Genewa, Szwajcaria. Mar. 2019

- M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, J. Y. Jeong i J. Seo. *Technical Description of Proposal for Call for Proposals on 3DoF+ Visual Prepared by Poznań University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI)*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m47407, Genewa, Szwajcaria. Mar. 2019

- K. Wegner, T. Grajek, A. Grzelka, M. Lorkiewicz, R. Ratajczak, O. Stankiewicz, J. Stankowski, H. Żabiński i M. Domański. *Depth Estimation Reference Software extension for lightfield images*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m46126, Marrakech, Maroko. Sty. 2019

- D. Łosiewicz, T. Grajek, K. Wegner, A. Grzelka, O. Stankiewicz i M. Domański. *360 Degree Test Image With Depth*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m41991, Gwangju, Korea Południowa. Sty. 2018

- D. Łosiewicz, T. Grajek, K. Wegner, A. Grzelka, O. Stankiewicz i M. Domański. *360 Degree Test Image With Depth*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m41991, Gwangju, Korea Południowa. Sty. 2018

- M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, J. Stankowski, O. Stankiewicz i K. Wegner. *Experimental Video Coding Software for Free Navigation Applications,* ISO/IEC JTC1/SC29/WG11 MPEG dok. m39527, Chengdu, Chiny. Paź. 2016

- M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, J. Stankowski, O. Stankiewicz i K. Wegner. *Coding Results for Poznan Fencing 2 and Poznan Blocks 2 Test Sequences in Free Navigation Scenario.* ISO/IEC JTC1/SC29/WG11 MPEG dok. m39215, Chengdu, Chiny. Paź. 2016

- M. Domański, A. Dziembowski, A. Grzelka i D. Mieloch. *Study on Nouniform Distributions of Cameras Located on an Arc.* ISO/IEC JTC1/SC29/WG11 MPEG dok. m38248, Genewa, Switzerland. Maj 2016

- A. Grzelka, D. Mieloch, O. Stankiewicz i K. Wegner. *Multiview Test Video Sequences for Free Navigation Exploration Obtained Using Pairs of Cameras.* ISO/IEC JTC1/SC29/WG11 MPEG dok. m38247, Genewa, Switzerland. Maj 2016

- M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. *[FTV AHG] Extended Results of Poznan University of Technology Proposal for Call for Evidence on Free-Viewpoint Television.* ISO/IEC JTC1/SC29/WG11 MPEG dok. m38246, Genewa, Switzerland. Maj 2016

- M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. *Technical Description of Poznan University of Technology Proposal for Call for Evidence on Free-Viewpoint Television.* ISO/IEC JTC1/SC29/WG11 MPEG dok. m37893, San Diego, USA. Lut. 2016

- M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz i K. Wegner. *[FTV AHG] Video and Depth Multiview Test Sequences Acquired With Circular Camera Arrangement – "Poznan Service" and "Poznan People".* ISO/IEC JTC1/SC29/WG11 MPEG dok. m36569, Warszawa, Polska. Czer. 2015

Bibliografia

[00420] J. Boyce, Ł. Kondrad i B. Chupeau. *Potential Improvements of MIV*. ISO/IEC JTC1/SC29/WG4 MPEG dok. n00004, OnLine. List. 2020.

[00620] J. Jung i B. Kroon. *Common Test Conditions for MPEG Immersive Video*. ISO/IEC JTC1/SC29/WG4 MPEG dok. n0006, OnLine. List. 2020.

[48220] *Committee Draft of MPEG Immersive Video*. ISO/IEC JTC1/SC29/WG11 MPEG dok. n19482, OnLine. Lip. 2020.

[49520] *Software Manual of IV-PSNR for Immersive Video*. ISO/IEC JTC1/SC29/WG11 MPEG dok. n19495, OnLine. Lip. 2020.

[AD03] M. Agrawal i L. Davis. "Complete Camera Calibration Using Spheres: Dual Space Approach". W: *IEEE ICCV*. Citeseer. 2003, s. 782–789.

[Aki+15] A. Akin, R. Capoccia, J. Narinx, J. Masur, A. Schmid i Y. Leblebici. "Real-Time Free Viewpoint Synthesis Using Three-Camera Disparity Estimation Hardware". W: *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2015, s. 2525–2528. DOI: 10.1109/ISCAS.2015.7169199.

[Akr14] S. Akramullah. "Video Quality Metrics". W: *Digital Video Concepts, Methods, and Metrics: Quality, Compression, Performance, and Power Trade-off Analysis*. Berkeley, CA: Apress, 2014, s. 101–160. ISBN: 978-1-4302-6713-3. DOI: 10.1007/978-1-4302-6713-3_4.

[Ala+17] A. D. Aladagli, E. Ekmekcioglu, D. Jarnikov i A. Kondoz. "Predicting Head Trajectories in 360 Virtual Reality Videos". W: *2017 International Conference on 3D Immersion (IC3D)*. 2017, s. 1–6. DOI: 10.1109/IC3D.2017.8251913.

[AO23] Y. Al-Obaidi. "Modeling of Codecs for 3-Dimensional Video". Prac. dokt. Politechnika Poznańska, maj 2023.

[AOG20] Y. Al-Obaidi i T. Grajek. "Estimation of the Optimum Depth Quantization Parameter for a Given Bitrate of Multiview Video Plus Depth in 3D-HEVC Coding". W: *Journal of WSCG* 28 (sty. 2020), s. 145–152. DOI: 10.24132/CSRN.2020.3001.17.

[Bao+16] Y. Bao, H. Wu, T. Zhang, A. A. Ramli i X. Liu. "Shooting a Moving Target: Motion-Prediction-Based Transmission for 360-Degree Videos". W: *2016 IEEE International Conference on Big Data (Big Data)*. 2016, s. 1161–1170. DOI: 10.1109/BigData.2016.7840720.

[Ber+20] D. Berjón, P. Carballeira, J. Cabrera, C. Carmona, D. Corregidor, C. Díaz, F. Morán i N. García. "FVV Live: Real-Time, Low-Cost, Free Viewpoint Video". W: *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 2020, s. 1–2. DOI: 10.1109/ICMEW46912.2020.9105977.

[Biba] *Methodology for the subjective assessment of the quality of television pictures*. ITU-R BT.500-12. Wrz. 2009.

[Bibb] *Subjective assessment methods for image quality in high-definition television*. ITU-R BT.710-4. 1998.

[Bibc] *Subjective video quality assessment methods for multimedia applications*. ITU-T P.910-04. Kw. 2008.

[Bibd] *Subjective video quality assessment methods for recognition tasks*. ITU-T P.912-08. 2008.

[BK04] Y. Boykov i V. Kolmogorov. "An Experimental Comparison Of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision". W: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.9 (2004), s. 1124–1137.

[BLB19] C. G. Bampis, Z. Li i A. C. Bovik. "Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment". W: *IEEE Transactions on Circuits and Systems for Video Technology* 29.8 (2019), s. 2256–2270. DOI: 10.1109/TCSVT.2018.2868262.

[Bn15] L. Bolecek i V. Říčný. "Influence of Stereoscopic Camera System Alignment Error on the Accuracy of 3D Reconstruction". W: *Radioengineering* 24 (czer. 2015), s. 610–620. DOI: 10.13164/re.2015.0610.

[Boy+21] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital i L. Yu. "MPEG Immersive Video Coding Standard". W: *Proceedings of the IEEE* 109.9 (2021), s. 1521–1536. DOI: 10.1109/JPROC.2021.3062590.

[Bro+21] B. Bross, Y. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan i J. Ohm. "Overview of the Versatile Video Coding (VVC) Standard and its Applications". W: *IEEE Transactions on Circuits and Systems for Video Technology* 31.10 (2021), s. 3736–3764. DOI: 10.1109/TCSVT.2021.3101953.

[Bui+21] M. Bui, L. Chang, H. Liu, Q. Zhao i G. Chen. "Comparative Study of 3D Point Cloud Compression Methods". W: *2021 IEEE International Conference on Big Data (Big Data)*. 2021, s. 5859–5861. DOI: 10.1109/BigData52589.2021.9671822.

[bullet360] *360Heros Debuts New Bullet360 Control Board for Multi-Camera Virtual Reality 360 Video Rigs at CES 2016*. https://www.prweb.com/releases/360heros_debuts_new_bullet360_control_board_for_multi_camera_virtual_reality_360_video_rigs_at_ces_2016/prweb13150794.htm. Dostęp: 2016-07-07.

[Car+03] J. Carranza, Ch. Theobalt, M. A. Magnor i H. Seidel. "Free-Viewpoint Video of Human Actors". W: *ACM Trans. Graph.* 22.3 (lip. 2003), s. 569–577. ISSN: 0730-0301. DOI: 10.1145/882262.882309. URL: https://doi.org/10.1145/882262.882309.

[CD08] X. Chen i J. Davis. "An Occlusion Metric for Selecting Robust Camera Configurations". W: *Machine Vision and Applications* 19.4 (2008), s. 217–222. ISSN: 1432-1769. DOI: 10.1007/s00138-007-0094-y. URL: https://doi.org/10.1007/s00138-007-0094-y.

[Ceu+18] B. Ceulemans, S. Lu, G. Lafruit i A. Munteanu. "Robust Multiview Synthesis for Wide-Baseline Camera Arrays". W: *IEEE Transactions on Multimedia* 20.9 (2018), s. 2235–2248.

[Che+18] Y. Chen, D. Murherjee, J. Han, A. Grange, Y. Xu, Z. Liu, S. Parker, C. Chen, H. Su, U. Joshi, C. Chiang, Y. Wang, P. Wilkins, J. Bankoski, L. Trudeau, N. Egge, J. Valin, T. Davies, S. Midtskogen, A. Norkin i P. de Rivaz. "An Overview of Core Coding Tools in

the AV1 Video Codec". W: *2018 Picture Coding Symposium (PCS)*. 2018, s. 41–45. DOI: 10.1109/PCS.2018.8456249.

[Che+21a] D. Chen, J. Wu, X. Zhu i T. Jia. "Depth Image Restoration Based on Bimodal Joint Sequential Filling". W: *Infrared Physics & Technology* 116 (2021), s. 103663. ISSN: 1350-4495. DOI: https://doi.org/10.1016/j.infrared.2021.103663. URL: https://www.sciencedirect.com/science/article/pii/S1350449521000359.

[Che+21b] R. Chen, F. Chen, G. Xu, X. Li, H. Shen i J. Yuan. "Precision Analysis Model and Experimentation of Vision Reconstruction with Two Cameras and 3D Orientation Reference". W: *Scientific Reports* 11 (lut. 2021). DOI: 10.1038/s41598-021-83390-y.

[Cor+17] X. Corbillon, G. Simon, A. Devlic i J. Chakareski. "Viewport-Adaptive Navigable 360-Degree Video Delivery". W: *2017 IEEE International Conference on Communications (ICC)*. 2017, s. 1–7. DOI: 10.1109/ICC.2017.7996611.

[CS03] Stephen L. Casner i Henning Schulzrinne. *RTP Profile for Audio and Video Conferences with Minimal Control*. RFC 3551. Lip. 2003. DOI: 10.17487/RFC3551. URL: https://www.rfc-editor.org/info/rfc3551.

[CS09] B. Cyganek i J. P. Siebert. *An Introduction to 3D Computer Vision Techniques and Algorithms*. 1 wyd. John Wiley & Sons Ltd, sty. 2009.

[CSL12] L. Chen, N. Shashidhar i Q. Liu. "Scalable Secure MJPEG Video Streaming". W: (mar. 2012). DOI: 10.1109/WAINA.2012.163.

[CVO11] G. Cheung, V. Velisavljevic i A. Ortega. "On Dependent Bit Allocation for Multiview Image Coding With Depth-Image-Based Rendering". W: *IEEE Transactions on Image Processing* 20.11 (2011), s. 3179–3194. ISSN: 1057-7149. DOI: 10.1109/TIP.2011.2158230.

[Cyg02] B. Cyganek. *Komputerowe Przetwarzanie Obrazów Trójwymiarowych*. Problemy Współczesnej Nauki: Informatyka. Akademicka Ofi-

cyna Wydawnicza EXIT, 2002. ISBN: 9788387674342. URL: https://books.google.pl/books?id=d6eftgAACAAJ.

[Dí+18] C. Díaz, J. Cabrera, M. Orduna, L. Muñoz, P. Pérez, J. Ruiz i N. García. "Viability Analysis of Content Preparation Configurations to Deliver 360VR Video via MPEG-DASH Technology". W: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. 2018, s. 1–2. DOI: 10.1109/ICCE.2018.8326235.

[DAOG21] M. Domański, Y. Al-Obaidi i T. Grajek. "Universal Modeling of Monoscopic and Multiview Video Codecs with Applications to Encoder Control". W: *2021 IEEE International Conference on Image Processing (ICIP)*. 2021, s. 2144–2148. DOI: 10.1109/ICIP42928.2021.9506735.

[DBT18] R. Doré, G. Briand i T. Tapie. *Technicolor 3DoF+ Test Materials*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m42349, San Diego, USA. Kw. 2018.

[DD19] M. Domański i A. Dziembowski. *Objective Quality Metric for Immersive Video*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m48093, Gothenburg , Szwecja. Czer. 2019.

[DGM16] A. Dziembowski, A. Grzelka i D. Mieloch. "Wielowidokowa Synteza w Systemach Telewizji Swobodnego Punktu Widzenia". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* (czer. 2016), s. 233–236. DOI: 10.15199/59.2016.6.14.

[DGM17] A. Dziembowski, A. Grzelka i D. Mieloch. "Zwiększanie Rozdzielczości Obrazu i Mapy Głębi w Celu Poprawy Jakości Syntezy Widoków Wirtualnych". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* 2017 (czer. 2017). DOI: 10.15199/59.2017.6.57.

[Dim18] E. Dima. "Multi-Camera Light Field Capture : Synchronization, Calibration, Depth Uncertainty, and System Design". Prac. dokt. Mid Sweden University, czer. 2018. DOI: 10.13140/RG.2.2.28982.04166.

[Dim+19] E. Dima, K. Brunnström, M. Sjöström, M. Andersson, J. Edlund, M. Johanson i T. Qureshi. "View Position Impact on QoE in an Immersive Telepresence System for Remote Operation". W: *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. 2019, s. 1–3.

[Dom10] M. Domański. *Obraz cyfrowy*. Warszawa: Wydawnictwa Komunikacji i Łączności, 2010.

[Dom+14a] M. Domański, A. Dziembowski, A. Kuehn, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz i K. Wegner. "Experiments on Acquisition and Processing of Video for Free-Viewpoint Television". W: *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. 2014, s. 1–4. DOI: 10.1109/3DTV.2014.6874727.

[Dom+14b] M. Domański, A. Dziembowski, A. Kuehn i D. Mieloch. "Telewizja Swobodnego Punktu Widzenia - Nowa Usługa czy Futurystyczna Wizja?" W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* nr 8-9 (2014), s. 734–737. ISSN: 1230-3496.

[Dom+15a] M. Domanski, A. Dziembowski, K. Klimaszewski, A. Łuczak, D. Mieloch, O. Stankiewicz i K. Wegner. *Comments on Further Standardization for Free-Viewpoint Television*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m35842, Genewa, Szwajcaria. Lut. 2015.

[Dom+15b] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, Ł. Kowalski, M. Kurc, A. Luczak, D. Mieloch, R. Ratajczak, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. "Methods of High Efficiency Compression for Transmission of Spatial Representation of Motion Scenes". W: *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 2015, s. 1–4.

[Dom+15c] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz i K. Wegner. *[FTV AHG] Video and Depth Multiview Test Sequences Acquired With Circular Camera Arrangement – "Poznan Service" and "Poznan People"*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m36569, Warszawa, Polska. Czer. 2015.

[Dom+15d] M. Domański, A. Dziembowski, D. Mieloch, A. Łuczak, O. Stankiewicz i K. Wegner. "A Practical Approach to Acquisition and Processing of Free Viewpoint Video". W: *2015 Picture Coding Symposium (PCS)*. 2015, s. 10–14. DOI: 10.1109/PCS.2015.7170037.

[Dom+16a] M. Domański, M. Bartkowiak, A. Dziembowski, T. Grajek, A. Grzelka, A. Łuczak, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. "New Results in Free-Viewpoint Television Systems for Horizontal Virtual Navigation". W: *2016 IEEE International Conference on Multimedia and Expo (ICME)*. 2016, s. 1–6. DOI: 10.1109/ICME.2016.7552993.

[Dom+16b] M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. *[FTV AHG] Extended Results of Poznan University of Technology Proposal for Call for Evidence on Free-Viewpoint Television*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m38246, Genewa, Switzerland. Maj 2016.

[Dom+16c] M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski i K. Wegner. *Technical Description of Poznan University of Technology Proposal for Call for Evidence on Free-Viewpoint Television*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m37893, San Diego, USA. Lut. 2016.

[Dom+16d] M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, J. Stankowski, O. Stankiewicz i K. Wegner. *Coding Results for Poznan Fencing 2 and Poznan Blocks 2 Test Sequences in Free Navigation Scenario*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m39215, Chengdu, Chiny. Paź. 2016.

[Dom+16e] M. Domański, A. Dziembowski, A. Grzelka, Ł. Kowalski, D. Mieloch, J. Samelak, J. Stankowski, O. Stankiewicz i K. Wegner. *Experimental Video Coding Software for Free Navigation Applications,* ISO/IEC JTC1/SC29/WG11 MPEG dok. m39527, Chengdu, Chiny. Paź. 2016.

[Dom+16f] M. Domański, A. Dziembowski, A. Grzelka i D. Mieloch. "Optimization of Camera Positions for Free-Navigation Applications". W: *2016 International Conference on Signals and Electronic Systems (ICSES)*. 2016, s. 118–123. DOI: 10.1109/ICSES.2016.7593833.

[Dom+16g] M. Domański, A. Dziembowski, A. Grzelka i D. Mieloch. *Study on Nouniform Distributions of Cameras Located on an Arc*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m38248, Genewa, Switzerland. Maj 2016.

[Dom+17] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, K. Klimaszewski, D. Mieloch, R. Ratajczak, O. Stankiewicz, J. Siast, J. Stankowski i K. Wegner. "Demonstration of a Simple Free Viewpoint Television System". W: *2017 IEEE International Conference on Image Processing (ICIP)*. 2017, s. 4589–4591.

[Dom+18a] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, K. Klimaszewski, D. Mieloch, R. Ratajczak, O. Stankiewicz, J. Siast, J. Stankowski i K. Wegner. *Free-Viewpoint Television Demonstration for Sports Events*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m41992, Gwangju, Korea Południowa. Sty. 2018.

[Dom+18b] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, D. Mieloch, R. Rataiczak, O. Stankiewicz, J. Stankowski i K. Wegner. "Real-Time Virtual Navigation Provision by Simple Means". W: *2018 International Conference on Signals and Electronic Systems (ICSES)*. 2018, s. 69–73.

[Dom+19a] M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, J. Y. Jeong i J. Seo. *Call for Proposals on 3DoF+: Aggregation of the Results of Submission Assessments Obtained by the Subjective Tests With Naïve Subjects*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m47985, Genewa, Szwajcaria. Mar. 2019.

[Dom+19b] M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, J. Y. Jeong i J. Seo. *Technical Description of Proposal for Call for Proposals on 3DoF+ Visual Prepared by Poznań University of Technology (PUT) and Electro-*

nics and Telecommunications Research Institute (ETRI). ISO/IEC JTC1/SC29/WG11 MPEG dok. m47407, Genewa, Szwajcaria. Mar. 2019.

[DS18] A. Dziembowski i J. Stankowski. "Szybka Synteza Widoków Wirtualnych Dla Systemów Telewizji Swobodnego Punktu Widzenia". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* nr 6 (2018), s. 330–333.

[DSD18] A. Dziembowski, J. Samelak i M. Domański. "View Selection for Virtual View Synthesis in Free Navigation Systems". W: *2018 International Conference on Signals and Electronic Systems (ICSES)*. 2018, s. 83–87.

[DSO17] E. Dima, M. Sjöström i R. Olsson. "Modeling Depth Uncertainty of Desynchronized Multi-Camera Systems". W: *2017 International Conference on 3D Immersion (IC3D)*. 2017, s. 1–6.

[DTL19] B. Demir, J. Thiran i Y. Leblebici. "Real-Time Textureless-Region Tolerant High-Resolution Depth Estimation System". W: *2019 22nd Euromicro Conference on Digital System Design (DSD)*. 2019, s. 69–73. DOI: 10.1109/DSD.2019.00020.

[Dub+18] R. Dube, M. Gollub, H. Sommer, . Gilitschenski, R. Siegwart, C. Cadena i J. Nieto. "Incremental Segment-Based Localization in 3D Point Clouds". W: *IEEE Robotics and Automation Letters* PP (lut. 2018), s. 1–1. DOI: 10.1109/LRA.2018.2803213.

[Duf17] Frederic Dufaux. *Academic Press Library in Signal Processing, Volume 6 : Image and Video Processing and Analysis and Computer Vision (Section Editor)*. List. 2017.

[Dzi+16a] A. Dziembowski, M. Domański, A. Grzelka, D. Mieloch, J. Stankowski i K. Wegner. "The Influence of a Lossy Compression on the Quality of Estimated Depth Maps". W: *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2016, s. 1–4. DOI: 10.1109/IWSSIP.2016.7502730.

[Dzi+16b] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz i M. Domański. "Depth Map Upsampling and Refinement for FTV Systems". W: *2016 International Conference on Signals and Electronic Systems (ICSES)*. 2016, s. 89–92.

[Dzi+16c] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner i M. Domański. "Multiview Synthesis — Improved View Synthesis for Virtual Navigation". W: *2016 Picture Coding Symposium (PCS)*. 2016, s. 1–5. DOI: 10.1109/PCS.2016.7906380.

[Dzi+17] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz i M. Domański. "Enhancing View Synthesis With Image and Depth Map Upsampling". W: *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2017, s. 1–4.

[Dzi18] A. Dziembowski. "Synteza Widoków Wirtualnych w Rzadkich Systemach Wielokamerowych dla Zastosowań w Swobodnej Nawigacji". Prac. dokt. Politechnika Poznańska, wrz. 2018.

[Dzi+19] A. Dziembowski, D. Mieloch, O. Stankiewicz, M. Domański, G. Lee i J. Seo. "Virtual View Synthesis for 3DoF+ Video". W: *2019 Picture Coding Symposium (PCS)*. 2019, s. 1–5. DOI: 10.1109/PCS48520.2019.8954502.

[Dzi20] A. Dziembowski. *Software manual of IV-PSNR for Immersive Video*. ISO/IEC JTC1/SC29/WG4 MPEG dok. n0014, OnLine. Paź. 2020.

[Dzi+22] A. Dziembowski, D. Mieloch, J. Stankowski i A. Grzelka. "IV-PSNR—The Objective Quality Metric for Immersive Video Applications". W: *IEEE Transactions on Circuits and Systems for Video Technology* 32.11 (2022), s. 7575–7591. DOI: 10.1109/TCSVT.2022.3179575.

[FBP06] C. Fehn, R. de la Barre i S. Pastoor. "Interactive 3-DTV-Concepts and Key Technologies". W: *Proceedings of the IEEE* 94.3 (2006), s. 524–538.

[Feh03] C. Fehn. "A 3D-TV Approach Using Depth-Image-Based Rendering DIBR". W: *in Proceedings of 3rd IASTED Conference on Visuali-*

zation, Imaging, and Image Processing, Benalmadena. 2003, s. 482–487.

[Feh04] C. Fehn. "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV". W: *Stereoscopic Displays and Virtual Reality Systems XI*. Red. Mark T. Bolas, Andrew J. Woods, John O. Merritt i Stephen A. Benton. T. 5291. International Society for Optics i Photonics. SPIE, 2004, s. 93 –104. DOI: 10.1117/12.524762. URL: https://doi.org/10.1117/12.524762.

[Fis+20] R. Fischer, P. Dittmann, Ch. Schröder i G. Zachmann. "Improved Lossless Depth Image Compression". W: *Journal of WSCG 28*. 2020, s. 168–176. DOI: 10.24132/JWSCG.2020.28.21.

[Foo+13] F. Fooladgar, S. Samavi, S. M. R. Soroushmehr i S. Shirani. "Geometrical Analysis of Localization Error in Stereo Vision Systems". W: *IEEE Sensors Journal* 13.11 (2013), s. 4236–4246. ISSN: 1558-1748. DOI: 10.1109/JSEN.2013.2264480.

[FSL17] S. Fu, F. Safaei i W. Li. "Optimization of Camera Arrangement Using Correspondence Field to Improve Depth Estimation". W: *IEEE Transactions on Image Processing* 26.6 (2017), s. 3038–3050.

[FT04] T. Fujii i M. Tanimoto. "A real-time ray-space acquisition system". W: *Stereoscopic Displays and Virtual Reality Systems XI*. Red. Mark T. Bolas, Andrew J. Woods, John O. Merritt i Stephen A. Benton. T. 5291. International Society for Optics i Photonics. SPIE, 2004, s. 179 –187. DOI: 10.1117/12.530575. URL: https://doi.org/10.1117/12.530575.

[Fue+19] Y. S. de la Fuente, G. S. Bhullar, R. Skupin, C. Hellge i T. Schierl. "Delay Impact on MPEG OMAF's Tile-Based Viewport-Dependent 360 Video Streaming". W: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9.1 (2019), s. 18–28. ISSN: 2156-3357. DOI: 10.1109/JETCAS.2019.2899516.

[Fuj+06] T. Fujii, K. Mori, K. Takeda, K. Mase, M. Tanimoto i Y. Suenaga. "Multipoint Measuring System for Video and Sound - 100-camera and microphone system". W: *2006 IEEE International Conference*

on Multimedia and Expo. 2006, s. 437–440. DOI: 10.1109/ICME.2006.262566.

[Gar+22] P. Garus, F. Henry, J. Jung, T. Maugey i Ch. Guillemot. "Immersive Video Coding: Should Geometry Information Be Transmitted as Depth Maps?" W: *IEEE Transactions on Circuits and Systems for Video Technology* 32.5 (2022), s. 3250–3264. DOI: 10.1109/TCSVT.2021.3100006.

[Gen13] J. Geng. "Three-Dimensional Display Technologies". W: *Advances in optics and photonics* 5 (grud. 2013), s. 456–535. DOI: 10.1364/AOP.5.000456.

[Goo+12] P. Goorts, M. Dumont, S. Rogmans i P. Bekaert. "An End-To-End System for Free Viewpoint Video for Smooth Camera Transitions". W: *2012 International Conference on 3D Imaging (IC3D)*. 2012, s. 1–7.

[Goo+14] P. Goorts, M. Javadi, S. Rogmans i G. Lafruit. *San Miguel Test Images With Depth Ground Truth*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m33163, Valencia, Hiszpania. Mar. 2014.

[GPS07] P. Gargallo, E. Prados i P. Sturm. "Minimizing the Reprojection Error in Surface Reconstruction from Images". W: *2007 IEEE 11th International Conference on Computer Vision*. 2007, s. 1–8. DOI: 10.1109/ICCV.2007.4409003.

[Gra+13] T. Grajek, R. Ratajczak, K. Wegner i M. Domański. "Limitations of Vehicle Length Estimation Using Stereoscopic Video Analysis". W: *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2013, s. 27–30. DOI: 10.1109/IWSSIP.2013.6623441.

[Gra+20] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki i A. Tabatabai. "An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC)". W: *APSIPA Transactions on Signal and Information Processing* 9 (2020), e13. DOI: 10.1017/ATSIP.2020.12.

[Grz+16] A. Grzelka, D. Mieloch, O. Stankiewicz i K. Wegner. *Multiview Test Video Sequences for Free Navigation Exploration Obtained Using Pairs of Cameras*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m38247, Genewa, Switzerland. Maj 2016.

[Grz+19a] A. Grzelka, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski i M. Domański. "Impact of Video Streaming Delay on User Experience with Head-Mounted Displays". W: *2019 Picture Coding Symposium (PCS)*. 2019, s. 1–5.

[Grz+19b] A. Grzelka, M. Lorkiewicz, A. Dziembowski i D. Mieloch. *PUT Proposal of PoznanFencing Posetraces*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m50646, Genewa, Szwajcaria. Paź. 2019.

[Grz+22] A. Grzelka, A. Dziembowski, D. Mieloch i M. Domański. "The Study of the Video Encoder Efficiency in Decoder-Side Depth Estimation Applications". W: *30th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision : WSCG*. 2022, s. 248–255. DOI: 10.24132/CSRN.3201.31.

[GSY17] B. Guan, Y. Shang i Q. Yu. "Planar self-calibration for stereo cameras with radial distortion". W: *Applied Optics* 56.33 (2017), s. 9257–9267.

[Guo+19] X. Guo, K. Yang, W. Yang, X. Wang i H. Li. *Group-wise Correlation Stereo Network*. 2019. arXiv: 1903.04025 [cs.CV].

[Guo+24] S. Guo, J. Hu, K. Zhou, J. Wang, L. Song, R. Xie i W. Zhang. "Real-Time Free Viewpoint Video Synthesis System Based on DIBR and a Depth Estimation Network". W: *IEEE Transactions on Multimedia* 26 (2024), s. 6701–6716. DOI: 10.1109/TMM.2024.3355639.

[Han+15] M. M. Hannuksela, Y. Yan, X. Huang i H. Li. "Overview of The Multiview High Efficiency Video Coding (Mv-HEVC) Standard". W: *2015 IEEE International Conference on Image Processing (ICIP)*. 2015, s. 2154–2158.

[HL15] J. Huang i J. Leou. "Virtual view synthesis for multi-view video plus depth sequences using spatial-temporal information". W: *2015 3DTV-Conference: The True Vision - Capture, Transmission and*

Display of 3D Video (3DTV-CON). 2015, s. 1–4. DOI: 10.1109/3DTV.2015.7169359.

[Hor+16] R. Horaud, M. Hansard, G. Evangelidis i C. Ménier. "An overview of depth cameras and range scanners based on time-of-flight technologies". W: *Machine Vision and Applications* 27.7 (czer. 2016), s. 1005–1020. ISSN: 1432-1769. DOI: 10.1007/s00138-016-0784-4. URL: http://dx.doi.org/10.1007/s00138-016-0784-4.

[Hor+17] Koki Horita, Shota Shiobara, Takao Okamawari, Fumio Teraoka i Kunitake Kaneko. "PTP accuracy measurement comparison between boundary clock and VLAN priority". W: *2017 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*. 2017, s. 1–6. DOI: 10.1109/ISPCS.2017.8056748.

[How11] P. Howarth. "Potential Hazards of Viewing 3-D Stereoscopic Television, Cinema And Computer Games: A Review". W: *Ophthalmic and physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)* 31 (mar. 2011), s. 111–22. DOI: 10.1111/j.1475-1313.2011.00822.x.

[HS16] M. Hosseini i V. Swaminathan. "Adaptive 360 VR Video Streaming Based on MPEG-DASH SRD". W: *2016 IEEE International Symposium on Multimedia (ISM)*. 2016, s. 407–408. DOI: 10.1109/ISM.2016.0093.

[Hu+14] Q. Hu, X. Zhang, Z. Gao i J. Sun. "Analysis and optimization of x265 encoder". W: *2014 IEEE Visual Communications and Image Processing Conference*. 2014, s. 502–505. DOI: 10.1109/VCIP.2014.7051616.

[Hua+19] B. Huang, D. Shen, G. Lin i S. D. Chai. "Multi-Camera Video Synchronization Based on Feature Point Matching and Refinement". W: *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*. 2019, s. 136–139.

[HZ04] R. Hartley i A. Zisserman. *Multiple View Geometry in Computer Vision*. 2 wyd. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511811685.

[Ike14] K. Ikeuchi. *Computer Vision: A Reference Guide*. Springer, 2014. ISBN: 9781785395932. URL: https://books.google.pl/books?id=EpHwjwEACAAJ.

[Irs+23] M. Z. Irshad, S. Zakharov, K. Liu, V. Guizilini, T. Kollar, A. Gaidon, Z. Kira i R. Ambrus. "NeO 360: Neural Fields for Sparse View Synthesis of Outdoor Scenes". W: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023, s. 9153–9164. DOI: 10.1109/ICCV51070.2023.00843.

[Jia+18] Y. Jiang, D. Russell, T. Godisart, N. Kholgade Banerjee i S. Banerjee. "Hardware Synchronization of Multiple Kinects and Microphones for 3D Audiovisual Spatiotemporal Data Capture". W: *2018 IEEE International Conference on Multimedia and Expo (ICME)*. 2018, s. 1–6.

[Jin+16] J. Jin, A. Wang, Y. Zhao i B. Zeng. "Region-Aware 3-D Warping for DIBR". W: *IEEE Transactions on Multimedia* 18 (czer. 2016), s. 1–1. DOI: 10.1109/TMM.2016.2539825.

[Jun+18] J. Jung, B. Kroon, R. Doré, G. Lafruit i J. Boyce. *Common Test Conditions on 3DoF+ and Windowed 6DoF*. ISO/IEC JTC1/SC29/WG11 MPEG dok. n18089, Macao, Chiny. Paź. 2018.

[KEV18] A. Khatiullin, M. Erofeev i D. Vatolin. "Fast Occlusion Filling Method for Multiview Video Generation". W: *2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. 2018, s. 1–4. DOI: 10.1109/3DTV.2018.8478562.

[Kim+13] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung i M. Gross. "Scene Reconstruction from High Spatio-Angular Resolution Light Fields". W: *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)* 32.4 (2013), 73:1–73:12.

[Kim+18] J. Kim, W. Kim, S. Ahn, J. Kim i S. Lee. "Virtual Reality Sickness Predictor: Analysis of visual-vestibular conflict and VR contents". W: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. 2018, s. 1–6. DOI: 10.1109/QoMEX.2018.8463413.

[Kli12] K. Klimaszewski. "Algorytmy kompresji sekwencji wielowidokowych". Prac. dokt. Politechnika Poznańska, 2012.

[KM16] R. Kijima i K. Miyajima. "Measurement of Head Mounted Display's latency in rotation and side effect caused by lag compensation by simultaneous observation — An example result using Oculus Rift DK2". W: *2016 IEEE Virtual Reality (VR)*. 2016, s. 203–204.

[Kov+15] P. Kovacs, A. Fekete, K. Lackner, V. Kiran, A. Zare i T. Balogh. *Big Buck Bunny Light-Field Test Sequences*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m35721, Genewa, Szwajcaria. Lut. 2015.

[KPP13] J. Kowalczuk, E. T. Psota i L. C. Pérez. "Real-time Temporal Stereo Matching Using Iterative Adaptive Support Weights". W: *IEEE International Conference on Electro-Information Technology , EIT 2013*. 2013, s. 1–6.

[KRN97] T. Kanade, P. Rander i P.J. Narayanan. "Virtualized Reality: Constructing Virtual Worlds from Real Scenes". W: *IEEE MultiMedia* 4.1 (1997), s. 34–47. DOI: 10.1109/93.580394.

[Kro18] B. Kroon. *3DoF+ Test Sequence ClassroomVideo*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m42415, San Diego, USA. Kw. 2018.

[KSG19] A. Kaushik, I. Sreedevi i D. Gupta. "A Grey Wolf Optimization Based Algorithm for Optimum Camera Placement". W: *Wireless Personal Communications* 105 (lut. 2019). DOI: 10.1007/s11277-019-06140-4.

[KSN13] K. Kawamura, H. Sankoh i S. Naito. *Requirements and Use Cases of Free-Navigation for Fly-Through Experience*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m31521, Genewa, Switzerland. Paź. 2013.

[Kuk+19] D. Kukolj, L. Bolecek, L. Polak, T. Kratochvil, O. Zach, J. Kufa, M. Slanina, T. Grajek, J. Samelak, M. Domański i D. A. Milovanovic. "3D Content Acquisition and Coding". W: *3D Visual Content Creation, Coding and Delivery*. Red. Pedro Amado Assunção i Atanas Gotchev. Cham: Springer International Publishing, 2019, s. 41–95. ISBN: 978-3-319-77842-6. DOI: 10.1007/978-3-319-77842-6_3.

[Kur19] M. Kurc. “Hybrid Techniques of Depth Map Estimation and Their Application in Three-dimensional Video Systems”. Prac. dokt. Politechnika Poznańska, 2019.

[KWD14] K. Klimaszewski, K. Wegner i M. Domański. “Video and Depth Bitrate Allocation in Multiview Compression”. W: *2014 21th International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2014, s. 207–210.

[KWZ99] S. B. Kang, J. Webb i C. Zitnick. “An Active Multibaseline Stereo System With Real-Time Image Acquisition”. W: (list. 1999).

[KZ02] V. Kolmogorov i R. Zabih. “Multi-camera Scene Reconstruction via Graph Cuts”. W: *Computer Vision — ECCV 2002*. Red. Anders Heyden, Gunnar Sparr, Mads Nielsen i Peter Johansen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, s. 82–96.

[KZ04] V. Kolmogorov i R. Zabin. “What Energy Functions Can Be Minimized via Graph Cuts?” W: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.2 (2004), s. 147–159. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2004.1262177.

[LB19] A. Lustica i J. Bozek. “5G Technology for Broadcast TV Applications”. W: *2019 International Symposium ELMAR*. 2019, s. 97–100. DOI: 10.1109/ELMAR.2019.8918666.

[LG18] K. Li i J. Guo. “A TOF Laser Radar Receiver with 100dB-Level Dynamic Range and mm-Level Accuracy”. W: *2018 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA)*. 2018, s. 132–133. DOI: 10.1109/CICTA.2018.8706052.

[LHL13] D. Li, H. Hang i Y. Liu. “Virtual view synthesis using backward depth warping algorithm”. W: *2013 Picture Coding Symposium (PCS)*. 2013, s. 205–208. DOI: 10.1109/PCS.2013.6737719.

[Li+13] B. Li, L. Heng, K. Koser i M. Pollefeys. “A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern”. W: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013, s. 1301–1307. DOI: 10.1109/IROS.2013.6696517.

[Li+18] L. Li, S. Zhang, X. Yu i L. Zhang. "PMSC: PatchMatch-Based Superpixel Cut for Accurate Stereo Matching". W: *IEEE Transactions on Circuits and Systems for Video Technology* 28.3 (2018), s. 679–692.

[Lip+18] M. Lipiński, E. van der Bij, J. Serrano, T. Włostowski, G. Daniluk, A. Wujek, M. Rizzi i D. Lampridis. "White Rabbit Applications and Enhancements". W: *2018 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*. 2018, s. 1–7.

[Liu+09] R. Liu, H. Zhang, M. Liu, X. Xia i T. Hu. "Stereo Cameras Self-Calibration Based on SIFT". W: *2009 International Conference on Measuring Technology and Mechatronics Automation*. T. 1. 2009, s. 352–355.

[Liu+14] J. Liu, S. Sridharan, C. Fookes i T. Wark. "Optimal Camera Planning Under Versatile User Constraints in Multi-Camera Image Processing Systems". W: *IEEE Transactions on Image Processing* 23.1 (2014), s. 171–184.

[LTT15] C. Lee, A. Tabatabai i K. Tashiro. "Free Viewpoint Video (FVV) Survey and Future Research Direction". W: *APSIPA Transactions on Signal and Information Processing* 4 (2015), e15. DOI: 10.1017/ATSIP.2015.18.

[Lv+17] D. Lv, X. Ying, Y. Cui, J. Song, K. Qian i M. Li. "Research on the Technology of LIDAR Data Processing". W: *2017 First International Conference on Electronics Instrumentation Information Systems (EIIS)*. 2017, s. 1–5.

[LZ17] G. Luo i Y. Zhu. "Foreground Removal Approach for Hole Filling in 3D Video and FVV Synthesis". W: *IEEE Transactions on Circuits and Systems for Video Technology* 27.10 (2017), s. 2118–2131. ISSN: 1051-8215. DOI: 10.1109/TCSVT.2016.2583978.

[LZS18] S. Li, C. Zhu i M. Sun. "Hole Filling With Multiple Reference Views in DIBR View Synthesis". W: *IEEE Transactions on Multimedia* 20.8 (2018), s. 1948–1959.

[Mar+10] J. Martin, J. Burbank, W. Kasch i D. L. Mills. *Network Time Protocol Version 4: Protocol and Algorithms Specification*. RFC 5905. Czer. 2010. DOI: 10.17487/RFC5905. URL: https://rfc-editor.org/rfc/rfc5905.txt.

[Mat+00] W. Matusik, Ch. Buehler, R. Raskar, S. J. Gortler i L. McMillan. "Image-Based Visual Hulls". W: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. USA: ACM Press/Addison-Wesley Publishing Co., 2000, s. 369–374. ISBN: 1581132085. DOI: 10.1145/344779.344951. URL: https://doi.org/10.1145/344779.344951.

[MDG16] D. Mieloch, A. Dziembowski i A. Grzelka. "Segmentacja Obrazu w Estymacji Map Głębi". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* (czer. 2016), s. 241–244. DOI: 10.15199/59.2016.6.16.

[MDG17] D. Mieloch, A. Dziembowski i A. Grzelka. "Estymacja Głębi Dla Systemów Wielowidokowych". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* 2017 (czer. 2017). DOI: 10.15199/59.2017.6.75.

[MFW07] Y. Morvan, D. Farin i P. With. "Joint Depth/texture Bit-Allocation for Multi-View Video Compression". W: sty. 2007.

[MG18] D. Mieloch i A. Grzelka. "Segmentation-based Method of Increasing The Depth Maps Temporal Consistency". W: *International Journal of Electronics and Telecommunications* vol. 64.No 3 (2018). DOI: 10.24425/123521.

[Mie+17a] D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz i M. Domański. "Graph-Based Multiview Depth Estimation Using Segmentation". W: *2017 IEEE International Conference on Multimedia and Expo (ICME)*. 2017, s. 217–222. DOI: 10.1109/ICME.2017.8019532.

[Mie+17b] D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz i M. Domański. "Temporal Enhancement of Graph-Based Depth Estimation Method". W: *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2017, s. 1–4.

[Mie18] D. Mieloch. “Depth Estimation in Free-Viewpoint Television”. Prac. dokt. Politechnika Poznańska, wrz. 2018.

[Mie+22] D. Mieloch, P. Garus, M. Milovanović, J. Jung, J. Jeong, S. Ravi i B. Salahieh. “Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video”. W: *IEEE Transactions on Circuits and Systems for Video Technology* 32.9 (2022), s. 6360–6374. DOI: 10.1109/TCSVT.2022.3162916.

[Mie+23a] D. Mieloch, A. Dziembowski, J. Jeong i G. Lee. “On the future of decoder-side depth estimation in MPEG immersive video coding”. W: *2023 Data Compression Conference (DCC)*. 2023, s. 354–354. DOI: 10.1109/DCC55655.2023.00042.

[Mie+23b] D. Mieloch, A. Dziembowski, B. Szydełko, D. Klóska, A. Grzelka, J. Stankowski, M. Domański, G. Lee i J. Jeong. *[MIV] New natural content – MartialArts*. ISO/IEC JTC1/SC29/WG4 MPEG dok. m61949, OnLine. Sty. 2023.

[Mil06a] D.L. Mills. *Computer Network Time Synchronization: The Network Time Protocol*. CRC Press, 2006. ISBN: 9781420006155. URL: https://books.google.pl/books?id=pdTcJBfnbq8C.

[Mil06b] Professor David L. Mills. *Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI*. RFC 4330. Sty. 2006. DOI: 10.17487/RFC4330. URL: https://rfc-editor.org/rfc/rfc4330.txt.

[Mil+20] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi i R. Ng. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”. W: *ECCV*. 2020.

[MMW11] K. Muller, P. Merkle i T. Wiegand. “3-D Video Representation Using Depth Maps”. W: *Proceedings of the IEEE* 99.4 (2011), s. 643–656. ISSN: 0018-9219. DOI: 10.1109/JPROC.2010.2091090.

[Mo+23] Z. Mo, W. Wu, W. Yu, T. Zhang, Z. Ke i J. Huang. “Fast Generalizable Novel View Synthesis with Uncertainty-Aware Sampling”. W: *Artificial Neural Networks and Machine Learning – ICANN 2023*. Red. Lazaros Iliadis, Antonios Papaleonidas, Plamen Angelov

i Chrisina Jayne. Cham: Springer Nature Switzerland, 2023, s. 389–401. ISBN: 978-3-031-44213-1.

[MSH06] G. Miller, J. Starck i A. Hilton. "Proj Ective Surface Refinement for Free-Viewpoint Video". W: *The 3rd European Conference on Visual Media Production (CVMP 2006) - Part of the 2nd Multimedia Conference 2006*. 2006, s. 153–162. DOI: 10.1049/cp:20061937.

[Muk+13] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu i R. Bultje. "The latest open-source video codec VP9 - An overview and preliminary results". W: *2013 Picture Coding Symposium (PCS)*. 2013, s. 390–393. DOI: 10.1109/PCS.2013.6737765.

[NDC17] G. Nieto, F. Devernay i J. Crowley. "Linearizing the Plenoptic Space". W: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, s. 1714–1725. DOI: 10.1109/CVPRW.2017.218.

[NM19] T. Nunome i R. Miyazaki. "The Effect of Contents and Available Viewpoints on QoE of Multi-view Video and Audio over WebRTC". W: *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*. 2019, s. 80–85.

[Ło+18] D. Łosiewicz, T. Grajek, K. Wegner, A. Grzelka, O. Stankiewicz i M. Domański. *360 Degree Test Image With Depth*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m41991, Gwangju, Korea Południowa. Sty. 2018.

[OFI12] J. Osada, N. Fukushima i Y. Ishibashi. "Influence of network delay on viewpoint change in free-viewpoint video transmission". W: *2012 18th Asia-Pacific Conference on Communications (APCC)*. 2012, s. 110–115.

[OM02] G. Olague i R. Mohr. "Optimal Camera Placement for Accurate Reconstruction". W: *Pattern Recognition* 35 (kw. 2002), s. 927–944. DOI: 10.1016/S0031-3203(01)00076-0.

[Pé+22] P. Pérez, D. Corregidor, E. Garrido, I. Benito, E. González-Sosa, J. Cabrera, D. Berjón, C. Díaz, F. Morán, N. García, J. Igual i J. Ruiz. "Live Free-Viewpoint Video in Immersive Media Production Over 5G Networks". W: *IEEE Transactions on Broadcasting* 68.2 (2022), s. 439–450. DOI: 10.1109/TBC.2022.3154612.

[Pal+17] K. Palacio-Baus, M. Mejía-Pesántez, L. Muñoz-Guillen i M. Espinoza-Mejía. "Current Challenges of Interactive Digital Television". W: *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*. 2017, s. 1–6.

[Pau+13] R. Pauliks, K. Tretjaks, K. Belahs i Romass Pauliks. "A survey on some measurement methods for subjective video quality assessment". W: *2013 World Congress on Computer and Information Technology (WCCIT)*. 2013, s. 1–6. DOI: 10.1109/WCCIT.2013.6618758.

[PCB14] M. H. Pinson, L. K. Choi i A. C. Bovik. "Temporal Video Quality Model Accounting for Variable Frame Delay Distortions". W: *IEEE Transactions on Broadcasting* 60.4 (2014), s. 637–649.

[Pen+15] Y. Peng, L. Chen, F. Ou-Yang, W. Chen i J. Yong. "JF-Cut: A Parallel Graph Cut Approach for Large-Scale Image and Video". W: *IEEE Transactions on Image Processing* 24.2 (2015), s. 655–666.

[Pla+16] H. Plank, G. Holweg, T. Herndl i N. Druml. "High Performance Time-of-Flight and Color Sensor Fusion With Image-Guided Depth Super Resolution". W: *2016 Design, Automation Test in Europe Conference Exhibition (DATE)*. 2016, s. 1213–1218.

[Pon+07] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola i V. Lukin. "On Between-Coefficient Contrast Masking of DCT Basis Functions". W: *Proc of the 3rd Int Workshop on Video Processing and Quality Metrics for Consumer Electronics* (sty. 2007).

[Poy12] C. Poynton. *Digital Video and HD (Second Edition)*. Boston: Morgan Kaufmann, 2012.

[PW04] M. H. Pinson i S. Wolf. "A New Standardized Method for Objectively Measuring Video Quality". W: *IEEE Transactions on Broadcasting* 50.3 (2004), s. 312–322.

[QL15] N. Qian i C. Lo. "Optimizing Camera Positions for Multi-View 3D Reconstruction". W: *2015 International Conference on 3D Imaging (IC3D)*. 2015, s. 1–8. DOI: 10.1109/IC3D.2015.7391816.

[Rav+22] Smitha Lingadahalli Ravi, Marta Milovanović, Luce Morin i Félix Henry. "A Study of Conventional and Learning-Based Depth Estimators for Immersive Video Transmission". W: *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*. 2022, s. 1–5. DOI: 10.1109/MMSP55362.2022.9948931.

[RD20] J. Fleureau R. Doré G. Briand. *Fan content proposal for MIV*. ISO/IEC JTC1/SC29/WG4 MPEG dok. m54732, OnLine. Lip. 2020.

[RK15] P. Rahimian i J. K. Kearney. "Optimal Camera Placement for Motion Capture Systems in the Presence of Dynamic Occlusion". W: list. 2015, s. 129–138. DOI: 10.1145/2821592.2821596.

[RK17] P. Rahimian i J. K. Kearney. "Optimal Camera Placement for Motion Capture Systems". W: *IEEE Transactions on Visualization and Computer Graphics* 23.3 (2017), s. 1209–1221. ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2637334.

[Rog+19] S. Rogge, D. Bonatto, J. Sancho, R. Salvador, E. Juarez, A. Munteanu i G. Lafruit. "MPEG-I Depth Estimation Reference Software". W: grud. 2019, s. 1–6. DOI: 10.1109/IC3D48390.2019.8975995.

[Run00] D. Runde. "How to Realize a Natural Image Reproduction Using Stereoscopic Displays With Motion Parallax". W: *IEEE Transactions on Circuits and Systems for Video Technology* 10.3 (2000), s. 376–386. ISSN: 1051-8215. DOI: 10.1109/76.836282.

[Saf+13] F. Safaei, P. Mokhtarian, H. Shidanshidi, W. Li, M. Namazi-Rad i A. Mousavinia. "Scene-adaptive Configuration of Two Cameras Using the Correspondence Field Function". W: *2013 IEEE International Conference on Multimedia and Expo (ICME)*. 2013, s. 1–6.

[San+17] W. Sankowski, M. Wodarczyk, D. Kacperski i K. Grabowski. "Estimation of Measurement Uncertainty in Stereo Vision System". W: *Image Vision Comput.* 61.C (maj 2017), s. 70–81. ISSN: 0262-8856.

DOI: 10.1016/j.imavis.2017.02.005. URL: https://doi.org/10.1016/j.imavis.2017.02.005.

[SB06] H. R. Sheikh i A. C. Bovik. "Image Information and Visual Quality". W: *IEEE Transactions on Image Processing* 15.2 (2006), s. 430–444.

[Sch+17] T. Schöps, J. L. Schönberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys i A. Geiger. "A Multi-view Stereo Benchmark with High-Resolution Images and Multi-camera Videos". W: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, s. 2538–2547. DOI: 10.1109/CVPR.2017.272.

[SD20] J. Stankowski i A. Dziembowski. "Fast View Synthesis for Immersive Video Systems". W: *Journal of WSCG* 28 (czer. 2020), s. 137–144. DOI: 10.24132/CSRN.2020.3001.16.

[SDW15] O. Stankiewicz, M. Domański i K. Wegner. "Estimation Of Temporally-consistent Depth Maps from Video With Reduced Noise". W: *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. 2015, s. 1–4.

[Sen+14] T. Senoh, A. Ishikawa, M. Okui, K. Yamamoto i N. Inoue. *EE1 and EE2 Results With Bee by NICT*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m32995, San Jose, USA. Sty. 2014.

[SFO11] D. Sandberg, P. Forssen i J. Ogniewski. "Model-Based Video Coding Using Colour and Depth Cameras". W: *2011 International Conference on Digital Image Computing: Techniques and Applications*. 2011, s. 158–163. DOI: 10.1109/DICTA.2011.33.

[Sha+14] F. Shao, W. Lin, G. Jiang, M. Yu i Q. Dai. "Depth Map Coding for View Synthesis Based on Distortion Analyses". W: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 4.1 (2014), s. 106–117. ISSN: 2156-3357. DOI: 10.1109/JETCAS.2014.2298314.

[She+13] L. Shen, Z. Liu, X. Zhang, W. Zhao i Z. Zhang. "An Effective CU Size Decision Method for HEVC Encoders". W: *IEEE Transactions on Multimedia* 15.2 (2013), s. 465–470. DOI: 10.1109/TMM.2012.2231060.

[Shi02] T. Shibata". "Head Mounted Display". W: *Displays* 23.1 (2002), s. 57 –64. ISSN: 0141-9382. DOI: https://doi.org/10.1016/S0141-9382(02)00010-0. URL: http://www.sciencedirect.com/science/article/pii/S0141938202000100.

[Shi+16] H. Shi, H. Zhu, J. Wang, S. Yu i Z. Fu. "Segment-based Adaptive Window And Multi-feature Fusion for Stereo Matching". W: *Journal of Algorithms & Computational Technology* 10 (lut. 2016). DOI: 10.1177/1748301815618299.

[Sik97] T. Sikora. "MPEG digital video-coding standards". W: *IEEE Signal Processing Magazine* 14.5 (1997), s. 82–100. DOI: 10.1109/79.618010.

[SJ15] Z. Sun i C. Jung. "Real-Time Depth-Image-Based Rendering on GPU". W: *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*. 2015, s. 324–328. DOI: 10.1109/CyberC.2015.97.

[SLY17] Y. Sun, A. Lu i L. Yu. "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video". W: *IEEE Signal Processing Letters* 24.9 (2017), s. 1408–1412.

[Smo+05] A. Smolic, K. Müller, P. Merkle, M. Kautzner i T. Wiegand. "3D Video Objects for Interactive Applications". W: *2005 13th European Signal Processing Conference*. 2005, s. 1–4.

[Son+14] Y. Song, S. Noh, J. Yu, C. Park i B. Lee. "Background Subtraction Based on Gaussian Mixture Models Using Color and Depth Information". W: *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*. 2014, s. 132–135.

[Sre+16] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela i M. Gabbouj. "Viewport-Adaptive Encoding and Streaming of 360-Degree Video for Virtual Reality Applications". W: *2016 IEEE International Symposium on Multimedia (ISM)*. 2016, s. 583–586. DOI: 10.1109/ISM.2016.0126.

[SS13] S. Sugimoto i S. Shimizu. *NICT-3D Test Sequences*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m30121, Vienna, Austria. Lip. 2013.

[SS15] N. Seth-Smith. "Network-Based Timing, Genlock and Time Code Using Precision Time Protocol". W: *SMPTE Motion Imaging Journal* 124.2 (2015), s. 16–20. DOI: 10.5594/j18512.

[Sta+13] O. Stankiewicz, K. Wegner, M. Tanimoto i M. Domański. *Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m31518, Genewa, Switzerland. Paź. 2013.

[Sta14] O. Stankiewicz. "Stereoscopic Depth Map Estimation and Coding Techniques for Multiview Video Systems". Prac. dokt. Politechnika Poznańska, 2014.

[Sta+18a] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch i J. Samelak. "A Free-Viewpoint Television System for Horizontal Virtual Navigation". W: *IEEE Transactions on Multimedia* 20.8 (2018), s. 2182–2195. ISSN: 1520-9210. DOI: 10.1109/TMM.2018.2790162.

[Sta+18b] O. Stankiewicz, K. Wegner, A. Dziembowski, M. Lorkiewicz, G. Lee, J. Seo i M. Domański. *Proposed test materials for 3DoF+ or Omnidirectional 6DoF*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m44461, Macao, Chiny. Paź. 2018.

[Sta+18c] O. Stankiewicz, K. Wegner, T. Grajek i M. Domański. "Nowe media immersyjne". W: *Przegląd Telekomunikacyjny + Wiadomości Telekomunikacyjne* nr 6 (2018), s. 397–400, CD. ISSN: 1230-3496.

[Sta+18d] J. Stankowski, A. Grzelka, D. Mieloch i K. Wegner. "Processing Pipeline for Real-Time Remote Delivery of Virtual View in FTV Systems". W: *2018 International Conference on Signals and Electronic Systems (ICSES)*. 2018, s. 118–123.

[Str+05] P. Strumiłło, D. Rzeszotarski, P. Pełczyński, B. Więcek i A. Lorenc. "System Obrazowania Stereoskopowego Sekwencji Scen Trójwymiarowych". W: *Elektronika : prace naukowe* nr 10 (2005), s. 165–184.

[Sul+12] G. J. Sullivan, J. Ohm, W. Han i T. Wiegand. "Overview of the High Efficiency Video Coding (HEVC) Standard". W: *IEEE Transactions on Circuits and Systems for Video Technology* 22.12 (2012), s. 1649–1668. DOI: 10.1109/TCSVT.2012.2221191.

[Tan05] M. Tanimoto. "FTV (free viewpoint television) creating ray-based image engineering". W: *IEEE International Conference on Image Processing 2005*. T. 2. 2005, s. II–25. DOI: 10.1109/ICIP.2005.1529982.

[Tan10] M. Tanimoto. "FTV (Free-viewpoint TV)". W: *2010 IEEE International Conference on Image Processing*. 2010, s. 2393–2396. DOI: 10.1109/ICIP.2010.5652084.

[Tan+12] M. Tanimoto, M. Panahpour Tehrani, T. Fujii i T. Yendo. "FTV for 3-D Spatial Communication". W: *Proceedings of the IEEE* 100.4 (2012), s. 905–917. ISSN: 0018-9219. DOI: 10.1109/JPROC.2011.2182101.

[Tec+16] G. Tech, Y. Chen, K. Müller, J. Ohm, A. Vetro i Y. Wang. "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding". W: *IEEE Transactions on Circuits and Systems for Video Technology* 26.1 (2016), s. 35–49. DOI: 10.1109/TCSVT.2015.2477935.

[Teh+18] M. Tehrani, B. Kroon, P. Nikitin, T. Senoh, K. Wegner, G. Lafruit, M. Tanimoto i Y. Sun. *MPEG-I Visual Test Material Summary*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m17461, Gwangju, Korea. Sty. 2018.

[Ter+20] M. Teratani, T. Senoh, B. Kroon, K. Wegner, B. Salahieh, D. Didier, J. Jung, G. Bang, R. Dore, Y. Sun, X. Jin, M. Tanimoto, G. Lafruit i Y. Lu. *Overview of MPEG-I Visual Test Materials*. ISO/IEC JTC1/SC29/WG11 MPEG dok. n19489, OnLine. Lip. 2020.

[TFF08] M. Tanimoto, T. Fujii i N. Fukushima. *1D Parallel Test Sequences for MPEG-FTV*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m15378, Archamps, France. Kw. 2008.

[Ure+11] H. Urey, K. V. Chellappan, E. Erden i P. Surman. "State of the Art in Stereoscopic and Autostereoscopic Displays". W: *Proceedings of the IEEE* 99.4 (2011), s. 540–555.

[Val+13] J. Valin, G. Maxwell, T. Terriberry i K. Vos. "High-Quality, Low-Delay Music Coding in the Opus Codec". W: *135th Audio Engineering Society Convention 2013* (sty. 2013), s. 73–82.

[VB16] Jean-Marc V. i Cary B. *WebRTC Audio Codec and Processing Requirements*. RFC 7874. Maj 2016. DOI: 10.17487/RFC7874. URL: https://www.rfc-editor.org/info/rfc7874.

[Ved+00] S. Vedula, S. Baker, S. Seitz i T. Kanade. "Shape And Motion Carving In 6D". W: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. T. 2. 2000, 592–598 vol.2. DOI: 10.1109/CVPR.2000.854926.

[VWS11] A. Vetro, T. Wiegand i G. J. Sullivan. "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard". W: *Proceedings of the IEEE* 99.4 (2011), s. 626–642. DOI: 10.1109/JPROC.2010.2098830.

[Wan+04a] Z. Wang, A. C. Bovik, H. R. Sheikh i E. P. Simoncelli. "Image Quality Assessment: From Error Visibility To Structural Similarity". W: *IEEE Transactions on Image Processing* 13.4 (2004), s. 600–612.

[Wan+04b] Z. Wang, A. C. Bovik, H. R. Sheikh i E. P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity". W: *IEEE Transactions on Image Processing* 13.4 (2004), s. 600–612. ISSN: 1057-7149. DOI: 10.1109/TIP.2003.819861.

[Wan+17] R. Wang, J. Luo, X. Jiang, Z. Wang, W. Wang, G. Li i W. Gao. "Accelerating Image-Domain-Warping Virtual View Synthesis on GPGPU". W: *IEEE Transactions on Multimedia* 19.6 (2017), s. 1392–1400. ISSN: 1520-9210. DOI: 10.1109/TMM.2017.2654120.

[Wan+21] Q. Wang, Z. Wang, K. Genova, P. Srinivasan, H. Zhou, J. T. Barron, R. Martin-Brualla, N. Snavely i T. Funkhouser. "IBRNet: Learning Multi-View Image-Based Rendering". W: *2021 IEEE/CVF Confe-*

rence on Computer Vision and Pattern Recognition (CVPR). 2021, s. 4688–4697. DOI: 10.1109/CVPR46437.2021.00466.

[wArstechnica] K. Orland. *How fast does "virtual reality" have to be to look like "actual reality"?* https://arstechnica.com/gaming/2013/01/how-fast-does-virtual-reality-have-to-be-to-look-like-actual-reality/?comments=1&comments-page=1. Dostęp: 2020-03-01.

[wBA] *Dokumentacja kamery Basler Ace*. https://www.baslerweb.com/fp-1559566827/media/downloads/documents/brochure/BAS1906_ace_Brochure_SAP0025_No37_EN_Web.pdf. Dostęp: 2019-07-24.

[wBL] *Basler Lens Selector*. https://www.baslerweb.com/en/tools/lens-selector/. Dostęp: 2024-04-22.

[wDC-BGH1] *LUMIX DC-BS1H*. https://www.panasonic.com/uk/consumer/cameras-camcorders/lumix-mirrorless-cameras/lumix-box-style-cameras/dc-bgh1.specs.html. Dostęp: 2024-05-28.

[Weg+19] K. Wegner, T. Grajek, A. Grzelka, M. Lorkiewicz, R. Ratajczak, O. Stankiewicz, J. Stankowski, H. Żabiński i M. Domański. *Depth Estimation Reference Software extension for lightfield images*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m46126, Marrakech, Maroko. Sty. 2019.

[Wei12] N.A. Weiss. *Introductory Statistics*. Pearson Education, 2012. ISBN: 9780321691224. URL: https://books.google.pl/books?id=_r5ucgAACAAJ.

[Wei+21] Z. Wei, Q. Zhu, C. Min, Y. Chen i G. Wang. *AA-RMVSNet: Adaptive Aggregation Recurrent Multi-view Stereo Network*. 2021. arXiv: 2108.03824 [cs.CV].

[wFFmpeg] *FFmpeg encoder*. www.ffmpeg.org. Dostęp: 2016-01-18.

[wGoPro] *GoPro*. https://gopro.com/content/dam/help/hero4-black/music-bundle-manuals/UM_H4Black-Music_ENG_REVB_WEB.pdf. Dostęp: 2020-05-06.

[wGoProDH] *Dual HERO System*. https://gopro.com/content/dam/help/dual-hero-system/manuals/DualHeroSystem_UG_SPA_REVD_WEB.pdf. Dostęp: 2020-05-06.

[WHC13] K. Wei, Y. Huang i S. Chien. "Point-Based Model Construction for Free-Viewpoint TV". W: *2013 IEEE Third International Conference on Consumer Electronics ¿ Berlin (ICCE-Berlin)*. 2013, s. 220–221. DOI: 10.1109/ICCE-Berlin.2013.6698017.

[wHEVC] *3D HEVC reference codec*. https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-13.0. Dostęp: 2017-01-16.

[Wie+03] T. Wiegand, G.J. Sullivan, G. Bjontegaard i A. Luthra. "Overview of the H.264/AVC video coding standard". W: *IEEE Transactions on Circuits and Systems for Video Technology* 13.7 (2003), s. 560–576. DOI: 10.1109/TCSVT.2003.815165.

[Wie+21] A. Wieckowski, J. Brandenburg, T. Hinz, Ch. Bartnik, V. George, G. Hege, Ch. Helmrich, A. Henkel, Ch. Lehmann, Ch. Stoffers, I. Zupancic, B. Bross i D. Marpe. "Vvenc: An Open And Optimized Vvc Encoder Implementation". W: *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 2021, s. 1–2. DOI: 10.1109/ICMEW53276.2021.9455944.

[wKITTI] *The KITTI Vision Benchmark Suite*. http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo. Dostęp: 2020-01-06.

[WLH04] J. Wang, M. Lewis i S. Hughes. "Gravity-Referenced Attitude Display for Teleoperation of Mobile Robots". W: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48 (wrz. 2004). DOI: 10.1177/154193120404802312.

[wMewpro] *MewPro4 Github*. https://github.com/orangkucing/MewPro4. Dostęp: 2016-01-25.

[wMid] *Middlebury Stereo Vision Page*. https://vision.middlebury.edu/stereo/. Dostęp: 2020-01-06.

[wMidm] *Middlebury Multi-View Stereo Vision Page*. https://vision.middlebury.edu/mview/. Dostęp: 2022-11-06.

[WML14] J. Wang, Z. Miao i Q. Liang. "Synchronization of Cameras from Human Motion Using Feature Points". W: *2014 12th International Conference on Signal Processing (ICSP)*. 2014, s. 1281–1284.

[wNetflix] *Toward A Practical Perceptual Video Quality Metric*. https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652. Dostęp: 2020-08-20.

[wOcu] *Oculus*. https://www.oculus.com. Dostęp: 2019-01-14.

[wOLED] *Our TV Motion Tests Response Time*. https://www.rtings.com/tv/tests/motion/motion-blur-and-response-time. Dostęp: 2023-06-02.

[wOmni] *Omni is here*. https://gopro.com/en/us/news/omni-is-here. Dostęp: 2022-01-28.

[wPTP20] "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems". W: *IEEE Std 1588-2019 (Revision ofIEEE Std 1588-2008)* (2020), s. 1–499.

[WSB03] Z. Wang, E. P. Simoncelli i A. C. Bovik. "Multiscale Structural Similarity for Image Quality Assessment". W: *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*. T. 2. 2003, 1398–1402 Vol.2.

[WSD15] K. Wegner, O. Stankiewicz i M. Domański. *Depth Based View Blending in VSRS*. ISO/IEC JTC1/SC29/WG11 MPEG dok. m37232, Genewa, Switzerland. Paź. 2015.

[wTM-1400] *TM-1400 Series Progressive Scan Shutter Cameras*. https://www.jai.com/uploads/documents/Discontinued-Products/English-Manuals-Datasheets/TM-Series/Manual_TM-1400series_Discontinued.pdf. Dostęp: 2024-05-28.

[Wu+17] P. Wu, Y. Liu, M. Ye, J. Li i S. Du. "Fast and Adaptive 3D Reconstruction With Extensively High Completeness". W: *IEEE Transactions on Multimedia* 19.2 (2017), s. 266–278. ISSN: 1520-9210. DOI: 10.1109/TMM.2016.2612761.

[Wu+23] G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian i Wang X. "4D Gaussian Splatting for Real-Time Dynamic Scene Rendering". W: *arXiv preprint arXiv:2310.08528* (2023).

[wX264] *Encoder x264*. www.videolan.org/developers/x264.html. Dostęp: 2017-01-16.

[wX265] *Encoder x265*. www.x265.org. Dostęp: 2017-01-16.

[wXHG1] *Canon XH G1*. https://www.canon.pl/support/consumer/products/camcorders/professional/xh-series/xh-g1.html?type=specifications. Dostęp: 2024-05-28.

[Xia+13] S. Xiang, L. Yu, Q. Liu i Z. Xiong. "A Gradient-Based Approach for Interference Cancelation in Systems With Multiple Kinect Cameras". W: maj 2013, s. 13–16. ISBN: 978-1-4673-5760-9. DOI: 10.1109/ISCAS.2013.6571770.

[Yam01] R. Yamashita. *AHG on 3D Video Coding in MPEG*. ISO/IEC JTC1/SC29/WG11 MPEG dok. n4524, Pattaya , Tajlandia. Grud. 2001.

[Yan+15] L. Yang, L. Zhang, H. Dong, A. Alelaiwi i A. E. Saddik. "Evaluating and Improving the Depth Accuracy of Kinect for Windows v2". W: *IEEE Sensors Journal* 15.8 (2015), s. 4275–4285. DOI: 10.1109/JSEN.2015.2416651.

[Yao+18] Y. Yao, Z. Luo, S. Li, T. Fang i L. Quan. *MVSNet: Depth Inference for Unstructured Multi-view Stereo*. 2018. arXiv: 1804.02505 [cs.CV].

[YBM20] A. Yaqoob, T. Bi i G. Muntean. "A Survey on Adaptive 360° Video Streaming: Solutions, Challenges and Opportunities". W: *IEEE Communications Surveys Tutorials* 22.4 (2020), s. 2801–2838. DOI: 10.1109/COMST.2020.3006999.

[YNT10] M. Yamamoto, T. Nunome i S. Tasaka. "The Effects of Camera Arrangements and Contents on QoE In Multi-view Video And Audio Ip Transmission". W: *TENCON 2010 - 2010 IEEE Region 10 Conference*. 2010, s. 1450–1455.

[YWB02] R Yang, G. Welch i G. Bishop. "Real-time Consensus-Based Scene Reconstruction Using Commodity Graphics Hardware". W: *10th Pacific Conference on Computer Graphics and Applications, 2002. Proceedings.* 2002, s. 225–234. DOI: 10.1109/PCCGA.2002.1167864.

[ZB11] T. Zhang i T. Boult. "Realistic Stereo Error Models and Finite Optimal Stereo Baselines". W: *2011 IEEE Workshop on Applications of Computer Vision (WACV).* 2011, s. 426–433. DOI: 10.1109/WACV.2011.5711535.

[Zha+19] F. Zhang, V. Prisacariu, R. Yang i P. H. S. Torr. *GA-Net: Guided Aggregation Net for End-to-end Stereo Matching.* 2019. arXiv: 1904.06587 [cs.CV].

[Zha99] Z. Zhang. "Flexible Camera Calibration by Viewing a Plane From Unknown Orientations". W: *Proceedings of the Seventh IEEE International Conference on Computer Vision.* T. 1. 1999, 666–673 vol.1. DOI: 10.1109/ICCV.1999.791289.

[Zhe20] L. Zheng. "Discussion of Design and Application of Live Broadcasting System Based on 5G + VR Technology". W: *2020 International Conference on Computer Engineering and Application (ICCEA).* 2020, s. 757–760. DOI: 10.1109/ICCEA50009.2020.00165.

[Zhu12] Q. Zhu. "A technique of camera calibration using single planar calibration image". W: *2012 5th International Congress on Image and Signal Processing.* 2012, s. 824–827. DOI: 10.1109/CISP.2012.6469808.

[Zit+04] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder i R. Szeliski. "High-Quality Video View Interpolation Using a Layered Representation". W: *ACM Trans. Graph.* 23.3 (2004), s. 600–608. ISSN: 0730-0301. DOI: 10.1145/1015706.1015766. URL: https://doi.org/10.1145/1015706.1015766.

[ZL16] C. Zhu i S. Li. "Depth Image Based View Synthesis: New Insights and Perspectives on Hole Generation and Filling". W: *IEEE Transactions on Broadcasting* 62.1 (2016), s. 82–93. ISSN: 0018-9316. DOI: 10.1109/TBC.2015.2475697.

[ZYZ13] P. Zhou, L. Yu i G. Zhong. "The non-Lambertian reflection in plenoptic sampling". W: *2013 IEEE International Conference on Image Processing*. 2013, s. 2154–2157. DOI: 10.1109/ICIP.2013.6738444.