



## PRZYGOTOWANIE WIELOWIDOKOWYCH SEKWENCJI WIZYJNYCH DLA BADAŃ NAD TELEWIZJĄ TRÓJWYMIAROWĄ

**Streszczenie:** W pracy przedstawiono proces przygotowania wielowidokowych sekwencji wizyjnych na potrzeby badań nad telewizją trójwymiarową. Przedstawiono założenia dotyczące cech użytecznych sekwencji, sposób ich akwizycji oraz przetwarzania wstępnego. Szczegółowo przedstawiono autorskie algorytmy rektyfikacji i korekcji kolorów dla systemów wielokamerowych. Zaprezentowano zarejestrowane sekwencje wielowidokowe, które zostały udostępnione środowisku naukowemu dla celów badawczych oraz zgłoszone jako sekwencje testowe w ramach prac grupy roboczej ISO/IEC MPEG-3DV.

### 1. WSTĘP

Następuje szybki rozwój technologii związanych z telewizją trójwymiarową. Technologie te, takie jak: wyznaczanie map głębi, synteza widoków czy kompresja sekwencji wielowidokowych, ulegają stopniowej ewolucji i udoskonalaniu. Ocena jakości tych technologii ma charakter czysto empiryczny i odbywa się z wykorzystaniem sekwencji testowych. Istotnym jest aby zbiór sekwencji wykorzystywanych przez badaczy był ujednolicony, aby możliwe było porównywanie otrzymywanych wyników.

Zbiór takich sekwencji testowych jest utrzymywany przez grupę roboczą MPEG-3DV, prowadzącą z ramienia ISO/IEC prace nad nowym standardem w dziedzinie telewizji trójwymiarowej [1]. Sekwencje te okazały się jednak być zbyt proste (zarówno pod względem treści jak i głębi sceny) by sprostać wymaganiom badań dotyczących m.in. algorytmów wyznaczania map głębi.

Przykładowo, sekwencje [2] cechują się nieskomplikowanymi teksturami i niewielkim ruchem. Dodatkowo, sekwencje te [2], przedstawiają sceny wewnętrzne o niskiej dynamice głębi i nieskomplikowanej jej strukturze lub (jak „Alt Moabit”) scenę zewnętrzną, o dużej dynamice głębi, lecz laminarnej (planarnej) strukturze głębi. Z kolei, sekwencja „Newspaper” [3], mimo iż zawiera więcej ruchu i ma bardziej skomplikowaną głębię, to jej użyteczność jest ograniczona ze względu na przeświecenie. Sekwencje „Champagne Tower” i „Pantomime” [4] posiadają ciekawy ruch, są także wyzwaniem pod względem obiektów przezroczystych, jednakże są to sekwencje wewnętrzne, kręcone na czarnym tle, które zajmuje większość sceny. Sekwencje „Lovebird” (1 i 2) [5] są sekwencjami zewnętrznymi, jednakże brak w nich złożonego ruchu oraz skomplikowanej struktury głębi sceny. Ponadto, wszystkie wymienione sekwencje kręcone były z wykorzystaniem nieruchomych zestawów kamer przemysłowych.

Brak dostępności zadowalających, wystarczająco złożonych sekwencji testowych dobrej jakości, spowodował ogłoszenie przez grupę MPEG-FTV wezwania [1] do dostarczenia nowych, bardziej zaawansowanych treści trójwymiarowych. W szczególności pożądane są sekwencje spełniające poniższe postulaty:

- zróżnicowana treść,
- duża dynamika głębi i skomplikowana jej struktura,
- dobre oświetlenie (przeświecenie/niedoświecenie),
- ruch kamery,
- złożony ruch w scenie,
- powierzchnie odbijające i przezroczyste,
- przedstawienie zarówno zewnętrznych i wewnętrznych scen naturalnych.

Niniejsza praca przedstawia proces przygotowania wielowidokowych sekwencji wizyjnych spełniających powyższe wymagania.

### 2. REJESTRACJA SEKWENCJI

#### 2.1. System kamerowy

Prezentowane sekwencje wielowidokowe zarejestrowane zostały z wykorzystaniem systemu kamerowego zbudowanego w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki (KTMiM) [6] Politechniki Poznańskiej. System ten składa się z 9 identycznych kamer Canon XH-G1 (Rys.1a). Jest on jest oryginalny i unikatowy w skali światowej, gdyż wykorzystywane w nim kamery są profesjonalnymi kamerami telewizyjnymi stosowanymi w kinematografii, nagrywającymi obraz w wysokiej rozdzielczości (1920×1080), z prędkością 30 klatek na sekundę. Systemy wykorzystywane w innych ośrodkach na Świecie [2-5] wykorzystują kamery prototypowe lub kamery przemysłowe o niższych rozdzielczościach np. XGA (1024×768).

Wybór kamer jest jednym z kluczowych elementów przy projektowaniu systemów wielokamerowych. Przy projektowaniu omawianego systemu, przydatność kamer badana była z wykorzystaniem autorskiego wyposażenia będącego przedmiotem zgłoszenia patentowego autorów. Wspomniane kamery Canon XH-G1 zostały wybrane ze względu na ich specjalne wejście synchronizacji umożliwiające zgranie momentu wyzwolenia przetworników z dokładnością do pojedynczych punktów obrazu.

Jednoczesne sterowanie wszystkimi kamerami umożliwia specjalny sterownik opracowany w Katedrze, sterujący pracą kamer za pośrednictwem protokołu LANC i zawierający generator synchronizacji.

Kolejną unikalną cechą omawianego systemu jest mobilny stelaż, zaprojektowany z myślą o stawieniu czoła wyzwaniom związanym z kręceniem sekwencji z ruchomymi kamerami. Stelaż ten, o wymiarach 163cm×189cm×350cm (Rys. 1b), wyposażony jest w zestaw kół o obrotowych osiach, umożliwiających jego łatwe poruszanie. Kamery zostały rozmieszczone na statywie równomiernie wzdłuż jednej linii co 13,75cm (co odpowiada podwójnej odległości rozstawu ludzkich oczu).

Sygnał wizyjny wyprowadzany jest 9 kablami koncentrycznymi SDI, umożliwiającymi przesyłanie go na duże odległości, sięgające kilkuset metrów. Sygnał doprowadzany jest do zestawu 5 stacjonarnych komputerów. Komputery te wyposażone są w karty DeckLink HD przechwytyjące sygnał wizyjny. Strumień danych zapisywany jest bezpośrednio na szybkich dyskach SSD.



Rys. 1. System kamerowy zbudowany w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki [5]

## 2.2. Opis scen

Zrealizowano wszystkich postulatów cechujących dobry materiał testowy, przedstawionych we wstępie, w jednej sekwencji byłoby trudne. Zrealizowano więc trzy niezależne sekwencje opisane poniżej [7].



Rys. 2. Wybrane ramki z przygotowanych sekwencji a) Poznań Street, b) Poznań CarPark, c) Poznań Hall

**Poznań Street** - Sekwencja obrazująca scenę przedstawiającą perspektywę na ulicę. Kamery są nieruchome, zaś po scenie porusza się dwójka ludzi oraz trzy samochody w różnych kierunkach. W oknach budynków i na samochodach widoczne są refleksy świetlne. Wiatr porusza liśćmi drzew. Oświetlenie dzienne. Dynamika głębi jest duża - od bliskiej (samochód wyjeżdżający z za kamery) do nieskończoności (niebo ponad budynkami).

**Poznań CarPark** - Sekwencja na dworze, przedstawiająca scenę na parkingu przed budynkiem. Dużo lustrzanych powierzchni w oknach. Po scenie porusza się sześcioro ludzi w różnych kierunkach oraz jeden samochód zbliżającym się do kamery. Dynamika głębi jest duża: od małej głębi (samochód jadący wyjeżdżający ze sceny koło kamery) do dużej (oddalony budynek w tle). Kamery są nieruchome. Oświetlenie jest dzienne.

**Poznań Hall** - Sekwencja przedstawiająca hol budynku Politechniki Poznańskiej. Scena zawiera korytarz idący w głąb sceny, windę oraz schody. Po scenie porusza się trójka osób. Występuje również obrotowy ruch parasolki i pionowy ruch windy. Występują odbicia świetlne na podłodze i na windzie. Dynamika głębi jest umiarkowana (od dalekiej w głąb korytarza) do bardzo bliskiej (filary znajdujące się przy kamerach) jednakże bardzo urozmaicona (schody, ściany, windy, osoby). Scena oświetlana była sztucznie lampami studyjnymi.

Cechą wyróżniającą tą sekwencję jest ruch kamer (jest to jedna z trzech publicznie dostępnych takich sekwencji na Świecie). Kamery poruszają się wzdłuż linii prostej w poprzek holu pokazując jego panoramę. Podczas kręcenia tej sekwencji napotkano na problem drgań podczas ruchu. W celu ich eliminacji, wykorzystano gumową matę po której toczono statyw.

## 3. PRZETWARZANIE WSTĘPNE

Sekwencje wielowidokowe pozyskane bezpośrednio z zestawu kamer posiadają szereg negatywnych cech. Cechy te wynikają głównie z nieidealnego ustawiania kamer w systemie kamerowym, nierównoległości osi optycznych kamer, zniekształceń geometrycznych wprowadzanych przez obiektywy kamer, różnic w oświetleniu oraz różnej charakterystyki przetworników zastosowanych w kamerach.

Przetwarzanie wstępne sekwencji ma na celu usunięcie tych niekorzystnych cech i składa się z dwóch procesów: rektyfikacji i korekcji kolorów.

### 3.1. Rektyfikacja

Celem procesu rektyfikacji jest transformacja obrazu w taki sposób aby powstała sekwencja przedstawiała obraz zarejestrowany pewnym wirtualnym zestawem idealnych kamer ustawionych idealnie wzdłuż linii prostej. Ma to na celu skorygowanie występujących w oryginalnych obrazach zniekształceń geometrycznych oraz zminimalizowanie wpływu niedokładności geometrii systemu kamerowego na przetworzone obrazy.

Do przeprowadzenia takiej transformacji konieczna jest znajomość dokładnych parametrów systemu kamerowego. Parametry te dzielimy na dwie kategorie: parametry wewnętrzne (opisujące samą kamerę) oraz parametry zewnętrzne (opisujące położenie kamery w przestrzeni).

Choć w literaturze znane są metody rektyfikacji pary obrazów [8], brak jest algorytmów jednoczesnej rektyfikacji wielu obrazów pozyskiwanych w systemach wielowidokowych. Opracowany przez autorów algorytm jednoczesnej rektyfikacji wielu widoków [9] składa się z trzech głównych etapów:

- wyznaczenia parametrów opisujących zestaw kamerowy,

- wyznaczania parametrów wirtualnego zestawu idealnych kamer,
- transformacji samego obrazu.

### Wyznaczenie parametrów kamer

Celem tego etapu jest wyznaczenie parametrów opisujących zestaw fizyczny kamerowy użyty do rejestracji sekwencji wielowidokowej.

Jako pierwsze wyznaczane są macierze parametrów wewnętrznych kamer zawierające takie informacje jak długość ogniskowej, współrzędne rzutu środka optycznego kamery na płaszczyznę przetwornika, oraz parametry zniekształceń geometrycznych (ograniczone jedynie do zniekształceń radialnych).

W tym celu, każdą z kamer niezależnie zarejestrowano obraz nieruchomej tablicy kalibrującej pokazanej na rys. 3. Uzyskane sekwencje uśredniono w czasie w celu zredukowania szumów.

Na podstawie zarejestrowanego położenia wierzchołków pól szachownicy i ich porównania z wzorcem, wyznaczone zostały macierze parametrów wewnętrznych i parametry zniekształceń geometrycznych każdej z kamer (za pomocą metody Zhanga [8]).

Macierze parametrów uzyskane dla różnych ustawień tablicy kalibrującej względem kamery poddane zostały uśrednieniu w celu zredukowania wpływu konkretnego ustawiania tablicy kalibrującej na wyznaczone parametry.

Następnie wyznaczane zostały parametry zewnętrzne kamer określające położenie kamer w przestrzeni (translacja, rotacja). W tym celu obraz tablicy kalibrującej nagrany został jednocześnie wszystkimi kamerami. Wymagało to ustawiania tablicy kalibrującej w taki sposób aby znalazła się w polu widzenia wszystkich kamer w zestawie. Przysparza to jednak dużo trudności w systemie złożonym z wielu kamer. W szczególności w stosowanym na PP systemie odległość między sąsiednimi kamerami wynosi 13,75 cm, co daje 108 cm pomiędzy skrajnymi kamerami w zestawie. Zbyt bliskie ustawienie tablicy powoduje, że nie wszystkie kamery obejmują szachownicę w całości, zaś zbyt dalekie powoduje zmniejszenie obrazu tablicy a związku z tym trudności w wyznaczeniu dokładnych położenia wierzchołków pól szachownicy.

Na podstawie położenia wierzchołków szachownicy w obrazach z kolejnych kamer wyznaczane są [8] wektory translacji (**T**) i macierze rotacji (**R**) poszczególnych kamer względem tablicy kalibrującej.

Następnie parametry **R** i **T** podlegają przeliczeniu na wartości względne względem kamery o numerze 0, z użyciem poniższych równań:

$$R_{rel}(i) = R(0) \cdot R(i)^{-1}$$

$$T_{rel}(i) = R_{rel}(i) \cdot T(i) - T(0)$$

gdzie:

- $R_{rel}(i)$  - względna macierz rotacji kamery względem kamery 0,
- $T_{rel}(i)$  - wektor translacji kamery  $i$  względem kamery 0,
- $R(i)$  - macierz rotacji kamery  $i$  względem tablicy kalibrującej
- $T(i)$  - wektor translacji kamery  $i$  względem tablicy kalibrującej.

### Wyznaczenie parametrów idealnego zestawu kamer

Ten etap ma na celu wyznaczenie parametrów pewnego wirtualnego zestawu idealnych kamer w taki sposób aby przekształcenie obrazu rzeczywistego wprowadzało jak najmniejsze zniekształcenia.

Położenie idealnych kamer w wirtualnym zestawie kamer, wyznaczana się za pomocą regresji liniowej trójwymiarowych współrzędnych kolejnych kamer w zastosowanym zestawie kamer. Zastosowanie takiej procedury zapewnia iż przesunięcia  $T_c$  pomiędzy odpowiadającymi sobie kamerami w wirtualnym i fizycznym zestawie kamer będą najmniejsze z możliwych.

Przejdzie między układem współrzędnych wirtualnego i rzeczywistego systemu kamer umożliwi macierz rotacji  $R_c$ . Położenie wirtualnych kamer oblicza się za pomocą następującego równania:

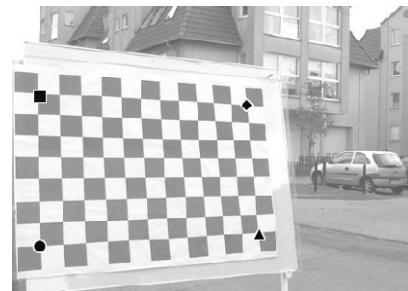
$$T_t(i) = R_c \cdot T_{rel}(i) + T_c \quad (1)$$

Aby osie optyczne kamer wirtualnych były równoległe, wszystkie kamery muszą posiadać taką samą macierz rotacji  $R_t(i)$ .

Ostatnim krokiem jest wyliczenie macierzy parametrów wewnętrznych idealnej kamery, przez uśrednienie macierzy parametrów wewnętrznych kamer rzeczywistych.

### Transformacja obrazu

Znając docelowe parametry idealnej kamery  $R_t(i)$ ,  $T_t(i)$  oraz jej parametry wewnętrzne, można wyznaczyć współrzędne rogów tablicy kalibracyjnej na obrazie z kamery wirtualnej. Przekształcenie perspektywiczne płaszczyzny obrazu z rzeczywistej kamery zapewniające pokrywanie się wyznaczonych właśnie współrzędnych rogów tablicy w kamerze wirtualnej z punktami na obrazie z rzeczywistej kamery jest przekształceniem rektyfikującym dla tej kamery.



Rys. 3. Tablica kalibracyjna z zaznaczonymi pokrywającymi się punktami na narożnikach

### 3.2. Korekcja kolorów

Pomimo iż sekwencje zostały nagrane kamerami bardzo wysokiej jakości, które zostały skalibrowane na etapie produkcji i posiadają prawie nie rozróżnialne profile kolorów obrazu z poszczególnych kamer są nieznacznie inne. Wynika to nie tylko z różnych profili kolorów w poszczególnych kamerach ale przede wszystkim z innych warunków ekspozycyjnych każdej z kamer.

Zaproponowana przez autorów metoda opiera się o transformacje histogramów. Ze względu na użycie trójprzetwornikowych kamer zastosowano korekcje kolorów niezależnie dla każdej składowej RGB.

Dla każdego widoku  $n$  wyznaczono histogram obrazu  $H_n$  dla każdej składowej. Następnie, na podstawie porównania histogramów między dwoma widokami, obliczono parametr korekcyjny, transformujący profil kolorów widoku  $m$  do profilu kolorów widoku  $n$ .

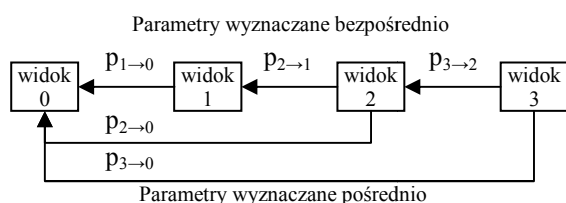
$$p_{m \rightarrow n} = \frac{\sum_i H_m(i) \cdot i}{\sum_i H_n(i) \cdot i} \quad (2)$$

Aby ujednoczyć wszystkie widoki, jeden z widoków wybrano jako widok referencyjny (widok referencyjny oznaczono jako widok numer 0). Ponadto aby zminimalizować wpływ różnic zarejestrowanego materiału przez każdą kamerę, tylko histogramy sąsiadujących widoków są ze sobą porównywane bezpośrednio. W związku z tym aby wyznaczyć współczynnik korygujący dla kamer bardziej oddalonych od kamery referencyjnej (rys. 4) wykorzystano następujący wzór

$$p_{n \rightarrow 0} = \prod_i p_{i+1 \rightarrow i} \quad (3)$$

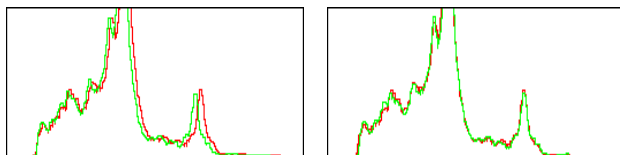
Następnie każdy z widoków skorygowano (rys. 5) zgodnie w wyznaczonym współczynniku korekcji

$$I'_n(x, y) = I_n(x, y) \cdot p_{n \rightarrow 0}$$



Rys. 4. Wyznaczanie parametrów korekcyjnych.

Aby jeszcze bardziej zmniejszyć zniekształcenie jako kamera referencyjna została wykorzystana kamera środkowa, a proces kalibracji został przeprowadzony niezależnie dla kamer po lewej i prawej strony kamery referencyjnej.



Rys. 5. Przykładowe histogramy obrazów z pary kamer (czarny/szary) przed korekcją (po lewej) i po korekcji (po prawej).

#### 4. PODSUMOWANIE

W pracy przedstawiono proces przygotowania wielowidokowych sekwencji wizyjnych na potrzeby badań nad telewizją trójwymiarową. Przedstawiono założenia dotyczące cech użytecznych sekwencji, sposób ich akwizycji oraz przetwarzania wstępnego. Szczegółowo przedstawiono autorskie algorytmy rektyfikacji i korekcji kolorów. Algorytmy te, stanowią rozwinięcie technik stosowanych dla pary kamer na przypadek systemu wielokamerowego.

Zaprezentowane zarejestrowane sekwencje wielowidokowe:

- Poznan\_Street,
- Poznan\_CarPark,
- Poznan\_Hall,

zostały udostępnione środowisku naukowemu [10] dla celów badawczych. Sekwencje te, zostały zgłoszone również jako materiał testowy dla badań nad telewizją trójwymiarową w ramach prac grupy roboczej ISO/IEC MPEG-3DV. Ich zgłoszenie spotkało się z pozytywnym odzewem ze strony środowiska: jedna z nich została już wykorzystana w eksperymencie dotyczącym kodowania sekwencji trójwymiarowych, zaś dwie pozostałe zostały przeznaczone do badań nad nową generacją algorytmów wyznaczających mapy głębi.

Praca finansowana ze środków na naukę jako projekt „badawczy rozwojowy”.

#### SPIS LITERATURY

- [1] “Call Call for 3D Test Material: Depth Maps & Supplementary Information”, ISO/IEC JTC1/SC29/WG11 MPEG/N10359, Lozanna, Szwajcaria, luty 2009.
- [2] I. Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, T. Wiegand „HHI Test Material for 3D Video”, ISO/IEC JTC1/SC29/WG11 MPEG/M15413, Archamps, Francja, kwiecień 2008.
- [3] Y.-S. Ho, E.-K. Lee, and C. Lee “Multiview Video Test Sequence and Camera Parameters”, ISO/IEC JTC1/SC29/WG11, MPEG/M15419, Archamps, Francja, kwiecień 2008.
- [4] M. Tanimoto, T. Fujii, N. Fukushima, “1D Parallel Test Sequences for MPEG-FTV”, ISO/IEC JTC1/SC29/WG11 MPEG/M15378, Archamps, Francja, kwiecień 2008.
- [5] Gi-M. Um, G. Bang, N. Hur, J. Kim and Yo-S. Ho “Video Test Material of Outdoor Scene”, ISO/IEC JTC1/SC29/WG11, MPEG/M15371, Archamps, Francja, kwiecień 2008.
- [6] M. Domański, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, O. Stankiewicz, K. Wegner “An experimental Free-view Television System”, 1st International Conference on Image Processing & Communications, Bydgoszcz, Polska, wrzesień 2009.
- [7] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner “Poznań Multiview Video Test Sequences and Camera Parameters” ISO/IEC JTC1/SC29/WG11 MPEG/M17050, Xian, Chiny, październik 2009.
- [8] Z. Zhang. A Flexible New Technique for Camera Calibration. Technical Report MSRTR-98-71, Microsoft Research, December 1998. Available together with the software at <http://research.microsoft.com/~zhang/Calib/>.
- [9] J. Stankowski, K. Klimaszewski, O. Stankiewicz, K. Wegner, M. Domański, “Preprocessing methods used for Poznań 3D/FTV test sequences” MPEG 2010 / M17174, Kyoto, Japonia, styczeń 2010.
- [10] <ftp://ftp.multimedia.edu.pl/3DV>.