

Adrian Dziembowski
Agnieszka Kuehn
Adam Łuczak
Dawid Mieloch
Krzysztof Wegner
Katedra Telekomunikacji Multimedialnej i Mikroelektroniki
Politechnika Poznańska
ul. Polanka 3
60-965 Poznań
{adziembowski, akuehn, aluczak, dmieloch, kwegner}@multimedia.edu.pl

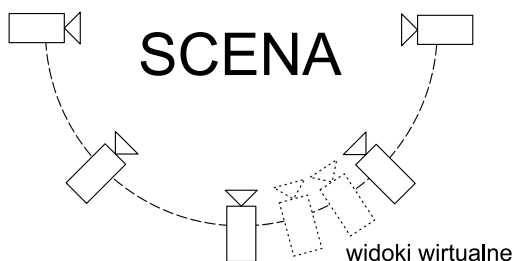
Realizacja eksperymentalnego systemu telewizji swobodnego punktu widzenia z łukowym ustawieniem kamer

Streszczenie: W artykule opisano eksperymentalny system telewizji swobodnego punktu widzenia zbudowany w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki Politechniki Poznańskiej. Przedstawione zostały szczegóły dotyczące budowy systemu wraz z metodami jego kalibracji, a także rezultaty w postaci przykładowej sekwencji wielowidokowej.

1. WSTĘP

Telewizja swobodnego punktu widzenia (ang. Free Viewpoint Television – FTV) jest w ostatnim czasie tematem intensywnych prac prowadzonych na całym świecie [5]. Możliwości przez nią oferowane sprawiają, że interesują się nią największe firmy związane z usługami telewizyjnymi.

Pomimo, że nie ma jeszcze w pełni gotowego i komercyjnie dostępnego systemu FTV, wiele zespołów badawczych tworzy własne systemy eksperymentalne. Również zespół z Katedry Telekomunikacji Multimedialnej i Mikroelektroniki Politechniki Poznańskiej (KTMiM) od lat pracuje nad systemami FTV. Systemy takie składają się zazwyczaj z kilku lub kilkunastu kamer rozmieszczonych wokół pewnej sceny (rys. 1.) (boiska, placu, ringu itp.). Widz może swobodnie zmieniać swój punkt obserwacji sceny i nie jest ograniczony tylko do widoków, które zarejestrowały rzeczywiste kamery. Aby osiągnąć efekt swobodnego przemieszczania się po scenie, w systemach takich stosuje się proces syntezy wielu widoków pośrednich. Widoki te tworzone są z pewnego modelu przestrzennego obserwowanej sceny który estymowany jest na podstawie zarejestrowanych obrazów.



Rys. 1. Schemat działania systemu FTV

2. CHARAKTERYSTYKA SYSTEMU

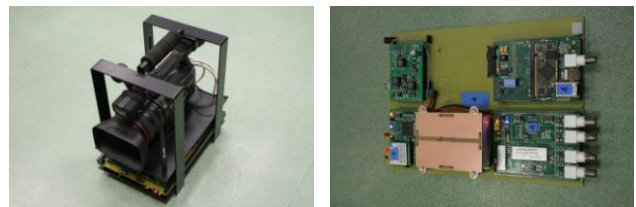
W skład zaprojektowanego przez zespół z KTMiM systemu telewizji swobodnego punktu widzenia wchodzi dziesięć wolnostojących stanowisk kamerowych. Każde z nich zawiera wysokiej klasy kamerę i oprzyrządowanie

pomocnicze, umieszczone we wspólnej, specjalnie zaprojektowanej dla potrzeb systemu, klatce kamerowej (rys. 2).

2.1. Stanowisko kamerowe

Do stworzenia systemu użyto dziesięciu kamer tego samego modelu – Canon XH G1. Ponadto, wszystkie kamery pochodzą z tej samej serii produkcyjnej, aby zapewnić powtarzalność parametrów kamer. Pomimo tego, charakterystyka optyczna poszczególnych urządzeń jest nieco różna. Różnice te są kompensowane w wyniku zastosowania przetwarzania wstępnego opisanego w rozdziałach 3.1, 3.2 i 3.4.

W celu zapewnienia mobilności systemu, wszystkie połączenia kablowe w systemie zostały wyeliminowane i zastąpione komunikacją bezprzewodową. W tym celu opracowano specjalną klatkę kamerową z układami bezprzewodowej komunikacji i opracowanym unikalnym bezprzewodowym systemem synchronizacji kamer. Każde stanowisko kamerowe wyposażone jest także w baterię litowo-jonową, pozwalającą na przynajmniej 6 godzin ciągłej pracy urządzenia. Wbudowany dysk HDD o pojemności 500GB pozwala na zarejestrowanie niemal dwóch godzin nieskompresowanego materiału wizyjnego o rozdzielczości Full-HD (nagrywanej ze standardową szybkością 25 ramek na sekundę, przy schemacie podpróbkowania chrominancji 4:2:0). Co więcej, modułowa budowa (rys. 2) stanowiska kamerowego umożliwia elastyczne dopasowanie systemu do indywidualnych wymagań – zarówno baterię, jak i dysk można zastąpić urządzeniami o innej charakterystyce.



Rys. 2. Klatka kamerowa / moduły dołączone do kamery

Praca poszczególnych bloków wchodzących w skład stanowiska kamerowego sterowana jest za pomocą urządzenia opartego na programowalnym układzie cyfrowym (FPGA).

2.2. Rozmieszczenie kamer w systemie

Opracowany system jest w pełni mobilny, a każda kamera pracuje niezależnie. W odróżnieniu od innych istniejących rozwiązań, w których kamery umieszczone

są na jednej sztywnej ramie lub ustawione na stałe w ściśle określonych (z bardzo dużą precyzją) pozycjach, w realizowanym systemie kamery znajdują się na oddzielnych statywach i mogą zostać szybko i dowolnie ustawione dookoła rejestrowanej sceny (rys. 3).

W opisywanym systemie założono łukowe ustawienie kamer. Takie rozmieszczenie urządzeń rejestrujących scenę jest bardzo korzystne z punktu widzenia użytkownika telewizji swobodnego punktu widzenia który może w takim przypadku wirtualnie okrążyć zarejestrowaną scenę. Zwiększa to znacząco atrakcyjność zarejestrowanego materiału, umożliwiając obserwację tych części sceny, które nie byłyby widoczne przy zastosowaniu tradycyjnego podejścia.



Rys. 3. System FTV z łukowym ustawieniem kamer

Jako, że system z założenia charakteryzuje się dużą mobilnością, przy jego rozstawianiu przedłożono szybkość nad dokładność. Do rozmieszczania kamer wystarczy dalmierz laserowy, z prostym kątomierzem umieszczonym na obrotowej głowicy statywu w centrum rejestrowanej sceny. Kamery ustawiane są w równej odległości od centrum sceny co z góry zadany kąt (do rejestracji sekwencji testowej przyjęto odległość 2 metrów i kąt 11 stopni). Po wstępnym ustawieniu statywów (przyjęto, że wystarczająca dokładność wynosi 5%) wyrównywane są ich wysokości. Operacja ta niweluje w końcowym materiale wrażenie unoszenia się i opadania punktu widzenia wraz z okrążaniem sceny. Ostatnim krokiem jest skierowanie osi optycznych poszczególnych kamer w określony punkt sceny (np. środek sceny z ustawianym dalmierzem). Im dokładniej wykonany zostanie ten krok, tym większy będzie obszar sceny zarejestrowany przez wszystkie kamery.

3. KALIBRACJA SYSTEMU

Zastosowanie niestandardowego podejścia bazującego na niezależnym rozmieszczeniu kamer wiąże się z tym, iż ich wzajemne pozycje nie są znane. Po ustawieniu kamer koniecznym staje się dokładne skalibrowanie systemu. W tradycyjnym systemie kalibracji dokonuje się poprzez fizyczne precyzyjne przemieszczanie kamer. W opracowanym systemie o swobodnym rozmieszczeniu kamer zastosowano inne podejście. Dokładne parametry systemu estymowane są programowo, na podstawie krótkiej sekwencji kalibrującej. W szczególności estymowane są parametry torów optycznych poszczególnych kamer, wzajemne położenie wszystkich kamer w systemie, oraz charakterystyki barwne wszystkich kamer.

3.1. Korekcja zniekształceń soczewkowych

Przeprowadzenie korekcji zniekształceń powstałych w układzie optycznym, wynikających z fizycznej budowy stosowanych soczewek, jest niezbędne nawet dla studyjnych kamer wysokiej klasy. W opisywanym systemie wykorzystano algorytm korekcji [11] estymujący i usuwający jedynie zniekształcenia radialne. Modelują one zniekształcenia wprowadzane przez soczewkę, które powstają szczególnie na obrzeżach rejestrowanego obrazu. W celu estymacji dokładnych wartości współczynników zniekształceń wykorzystywana jest tablica kalibracyjna o wzorze szachownicy. Dla każdej z kamer w systemie niezależnie należy dokonać rejestracji obrazu tablicy w taki sposób, by zajmowała ona jak największy obszar obrazu. Pozwala to na zredukowanie zniekształceń geometrycznych niemalże do zera.

3.2. Estymacja parametrów wewnętrznych

Zarejestrowane wcześniej widoki tablicy kalibrującej wykorzystywane są również podczas estymacji parametrów wewnętrznych. Opisują one sposób odwzorowania rejestrowanej sceny przez poszczególne kamery systemu. Zestaw parametrów dla jednej kamery zawiera informacje o jej ogniskowej (w kierunku poziomym i pionowym), położeniu środka optycznego (punktu centralnego), a także kąta pomiędzy osiami przetwornika (wynikającego z nieregularności w rozmieszczeniu elementów światłoczułych).

W opisywanym systemie wykorzystano algorytm Zhanga [11], który to na podstawie wyznaczonych punktów odniesienia (narożników pól szachownicy) wyznacza parametry wewnętrzne kamer.

Wraz z opisywaną wcześniej korekcją zniekształceń soczewkowych, proces estymacji parametrów wewnętrznych wystarczy wykonać tylko raz dla danych ustawień parametrów optycznych kamer.

3.3. Estymacja parametrów zewnętrznych

Względne położenie kamer systemu opisywane jest za pomocą parametrów zewnętrznych. Określają one zarówno położenie poszczególnych kamer we wspólnym trójwymiarowym układzie współrzędnych, jak i ich kierunki obserwacji w postaci macierzy rotacji.

Tradycyjne metody wyznaczania parametrów zewnętrznych kamer wymagają rejestracji złożonych sekwencji kalibrujących, jednocześnie oferując niewystarczającą dokładność, by mogły być one wykorzystane w opisywanym systemie. Zastosowano zatem podejście wykorzystujące optymalizację funkcji zależności między kamerami systemu. W stworzonym algorytmie, w przeciwieństwie do metod tradycyjnych, w których przeprowadzana była kalibracja co najwyżej 3 kamer, możliwa jest jednoczesna estymacja parametrów dla wszystkich kamer systemu. Dzięki temu dla wszystkich kamer zachowana jest ta sama skala odwzorowania przestrzeni, a proces kalibracji staje się krótszy. Podczas kalibracji wykorzystywane są ręcznie zaznaczone punkty odniesienia, które oznaczane są na zarejestrowanej wszystkich kamerami systemu sekwencji. Używając początkowo zakładanych pozycji kamer (np. odległości od środka sceny i kąta między kamerami), dla każdego punktu tworzone są tzw. linie epipolarne. Te same punkty sceny

na obrazie z innej kamery znajdują się właśnie na wspomnianych liniach.

Wykorzystując algorytm Lipschitza [13] dokonywana jest minimalizacja błędu, określanego jako odległość każdego z punktów odniesienia wobec odpowiadającej mu linii epipolarnej. Algorytm zmienia wartości parametrów zewnętrznych tak, aby błąd odwzorowania był jak najbliższy zeru – czyli sytuacji, gdy w poprawny sposób wyestymowano pozycje kamer. Błąd jest wyznaczany dla każdej pary kamer.

3.4. Korekcja barwna

Para obrazów zarejestrowana przez dwie kamery tego samego typu (nawet o identycznych ustawieniach) ma zazwyczaj różną charakterystykę barwną. Skutkuje to tym iż dany przedmiot na każdym z obrazów ma inny kolor. W systemach FTV kluczowe jest aby obrazy ze wszystkich kamer miały identyczną charakterystykę barwną. Wyrównanie charakterystyk barwnych ma na celu: zwiększenie efektywności predykcji międzywideo-owej w kompresji sekwencji wielowidokowych, poprawne dopasowanie elementów sceny w procesie wyznaczenia map głębi oraz uniknięcie artefaktów kolorystycznych podczas tworzenia dodatkowych widoków pośrednich (wirtualnych).

Znane z literatury metody pozwalają na korekcję jedynie pary obrazów. Korekcja ta polega na znalezieniu pewnej zależności pomiędzy przestrzeniami kolorystycznymi dwóch obrazów. Często używa się w tym celu specjalnego wzorca kolorystycznego, jednak jego zastosowanie ogranicza się jedynie do niewielkich scen i wydłuża czas nagrywania.

Dlatego też w opisywanym systemie wyznacza się słownik przekształcenia na podstawie uśrednionych po sekwencji histogramów dla każdej kamery w sposób, jaki opisano w artykule [4]. Korekcja ta została jednak przeprowadzona dla 10 kamer jednocześnie.

Wyzwaniem w systemie wielokamerowym jest więc odpowiednia kolejność obrazów poddanych korekcji barwnej. Dla każdego z wymienionych wcześniej etapów przetwarzania najważniejsza jest podobieństwo pomiędzy sąsiednimi obrazami. Z tego powodu korekcja jest dokonywana od lewej do prawej strony, ze zmieniającym obrazem odniesienia.

4. ESTYMACJA MAP GŁĘBI

Do wytworzenia dodatkowych pośrednich widoków sceny niezbędne jest uprzednie wyznaczenie przestrzennego modelu zarejestrowanej sceny w postaci map głębi.

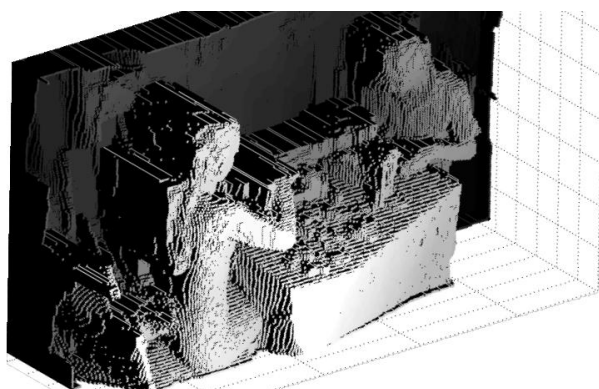
Mapy głębi określają odległość zarejestrowanych w scenie punktów od kamery systemu. Najczęściej przyjmują one formę obrazu w skali szarości, o rozdzielczości rejestrowanego obrazu. Im ciemniejszy jest punkt mapy głębi, tym dalej od kamery się znajduje (rys. 4).

Mapy głębi estymuje się algorytmicznie na podstawie zarejestrowanych uprzednio obrazów sceny. Wyznaczenie map głębi jest najbardziej wymagającym obliczeniowo elementem prezentowanego w artykule systemu telewizyjnego swobodnego punktu widzenia.



Rys. 4. Przykładowa mapa głębi dla sekwencji 'Poznan Blocks' [8]

Dla prezentowanego systemu autorzy stworzyli wiele dedykowanych narzędzi uogólniających znane z literatury narzędzia estymacji map głębi na przypadek wielokamerowy z łukowym ustawieniem kamer [10, 12]. W szczególności opracowano technikę pasowania fragmentów obrazów miękkimi segmentami [10] oraz techniki estymacji map głębi z jednoczesną estymacją przesłoneń związanych ze zmianą perspektywy widzenia kamery [12]. Fragmenty opracowanych technik zostały włączone do oprogramowania modelowego (Depth Estimation Reference Software - DERS) [6,7] grupy MPEG Międzynarodowej Organizacji Normalizacyjnej. Oprogramowanie to jest powszechnie wykorzystywane jako technika odniesienia w badaniach na algorytmami wyznaczania głębi. Przykładowy model przestrzenny zarejestrowanej sceny przedstawiono na rys. 5.



Rys. 5. Przestrzenne odwzorowanie sceny na podstawie mapy głębi [8]

5. SYNTEZA WIRTUALNYCH WIDOKÓW

Synteza dodatkowych widoków sceny widzianej z wirtualnej kamery umieszczonej w dowolnym punkcie pomiędzy dwiema rzeczywistymi kamerami bazuje na oprogramowaniu opracowanym w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki [2]. Fragmenty opracowanego algorytmu stały się częścią oprogramowania modelowego VSRS [3] grupy MPEG.

Technika syntezy obrazów wymaga dwóch obrazów sceny zarejestrowanych przez rzeczywiste kamery i odpowiadających im map głębi. Pozycje kamer określa się w postaci macierzy projekcji łączącej wyznaczone wcześniej parametry wewnętrzne i zewnętrzne poszczególnych kamer. Do algorytmu podaje się także pozycję kamery wirtualnej w postaci odpowiedniej macierzy projekcji.

W pierwszym kroku algorytmu widok wirtualny jest syntezowany niezależnie na podstawie każdego z obrazów wejściowych. W tym celu tworzona jest mapa głębi widoku wirtualnego - punkty mapy głębi widoków wejściowych rzutuje się na płaszczyznę obrazu wirtualnego za pomocą wyznaczonej macierzy homografii pomiędzy obrazem rzeczywistym a obrazem wirtualnym (na podstawie podanych wcześniej macierzy projekcji). Znajdąc mapę głębi widoku wirtualnego, rzutuje się kolor pobierając go z odpowiednich punktów obrazów rzeczywistych.



Rys. 6. Zsyntezowany widok wirtualny pomiędzy kamerami 4 i 5 dla sekwencji 'Poznan Blocks' [1]

Otrzymane w ten sposób dwa zsyntezowane widoki (na podstawie lewego i prawego obrazu) posiadają także obszary, w których nie udało się ustalić ich głębi, a więc też informacja o ich kolorze jest niedostępna. W zależności od tego, czy obraz wirtualny znajduje się bliżej obrazu lewego czy prawego, do wyjściowego obrazu kopiowany jest widok zsyntezowany odpowiadający bliższemu obrazowi. Do nieokreślonych obszarów jest kopiowana informacja z drugiego obrazu. Obszary, których nie udało się zsyntezować ani z lewego, ani z prawego obrazu, są interpolowane na podstawie sąsiadujących punktów obrazu.

W ramach prac stworzono także specjalną implementację algorytmu na procesory kart graficznych umożliwiającą syntezę dwóch wysokiej jakości widoków wirtualnych (stereopary) w czasie rzeczywistym.

6. PODSUMOWANIE

W pracy przedstawiono opracowany w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki unikatowy w skali światowej eksperymentalny system telewizji swobodnego punktu widzenia. W ramach realizacji systemu przygotowano dedykowane oprogramowanie i przeprowadzono eksperymenty dla systemu złożonego z 10 kamer rozstawionych co 11 stopni, na łuku o promieniu 3,5 metra. Wynikiem prac jest powstanie sekwencji testowej "Poznan Blocks" [1], która została zgłoszona grupie MPEG jako propozycja sekwencji testowej w pracach nad technikami przetwarzania w przyszłych systemach telewizji swobodnego punktu widzenia. Dotychczasowe badania potwierdzają uzyskanie porównywalnej jakości zsyntezowanych widoków wirtualnych jak w systemach z liniowym ustawieniem kamer. Opracowany system stanowi bazę do dalszych prac nad rozwojem telewizji swobodnego punktu widzenia w Polsce i na świecie.

Sfinansowano ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2012/05/B/ST7/01279.

SPIS LITERATURY

- [1] M. Domański, A. Dziembowski, A. Kuehn, M. Kurc, A. Łuczak, D. Mieloch, J. Siast, O. Stankiewicz, K. Wegner, „Poznan Blocks – a multiview video test sequence and camera parameters for Free Viewpoint Television”, ISO/IEC JTC1/SC29/WG11 MPEG 2014 / M32248, San Jose, USA, styczeń 2014.
- [2] M. Domański, M. Gotfryd, K. Wegner, „View synthesis for multiview video transmission”, The 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, USA, 13-16 lipiec 2009.
- [3] K. Wegner, O. Stankiewicz, M. Tanimoto, M. Domański, „Enhanced View Synthesis Reference Software (VSRS) for Free-viewpoint Television”, ISO/IEC JTC1/SC29/WG11 MPEG 2013 / M31520 Genewa, Szwajcaria, październik 2013.
- [4] U. Fecker, M. Barkowsky, A. Kaup, „Histogram-Based Prefiltering for Luminance and Chrominance Compensation of Multiview Video”, *IEEE Transactions on Circuits and Systems for Video Technology*, wol. 18, str. 1258-1267, 2008.
- [5] M. Domański, *Obraz cyfrowy*, WKiŁ, 2010.
- [6] K. Wegner, O. Stankiewicz, M. Tanimoto, M. Domański, „Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television”, ISO/IEC JTC1/SC29/WG11 MPEG 2013 / M31520, Genewa, Szwajcaria, październik 2013.
- [7] O. Stankiewicz, K. Wegner, „Depth Map Estimation Software version 3”, ISO/IEC JTC1/SC29/WG11 MPEG 2008 / M15540, Hanower, Niemcy, lipiec 2008.
- [8] K. Wegner, O. Stankiewicz, A. Dziembowski, D. Mieloch, M. Domański, „First version of depth maps for Poznan Blocks multiview video test sequence”, ISO/IEC JTC1/SC29/WG11 MPEG 2014 / M32248, San Jose, USA, styczeń 2014.
- [9] M. Domański, A. Łuczak, K. Klimaszewski, „Telewizja swobodnego punktu widzenia”, *Przegląd Telekomunikacyjny* 2-3, str. 84-86, 2013.
- [10] K. Wegner, O. Stankiewicz M. Domański, „Stereoscopic depth estimation using fuzzy segment matching”, 28th Picture Coding Symposium PCS2010, Nagoya, Japonia, 8-10 grudzień 2010.
- [11] Z. Zhang, "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations", *Proceedings International Conf. on Computer Vision*, str. 666-673, 1999.
- [12] K. Wegner, O. Stankiewicz „Improved depth estimation with advanced occlusion handling” ISO/IEC JTC1/SC29/WG11 MPEG 2011 / M21365, Torino, Włochy, lipiec 2011.
- [13] D. R. Jones, C. D. Perttunen, B. E. Stuckman, “Lipschitzian optimization without the Lipschitz constant”, *Journal of Optimization Theory and Application*, wol. 79, nr. 1, 1993.