

TECHNIKI KODOWANIA I WYZNACZANIA MAP GŁĘBI STEREOSKOPOWEJ W SYSTEMACH TELEWIZJI WIELOWIDOKOWEJ

Olgierd Stankiewicz, ostank@multimedia.edu.pl

Politechnika Poznańska, Katedra Telekomunikacji Multimedialnej i Mikroelektroniki
Ul. Polanka 3, 61-131 Poznań

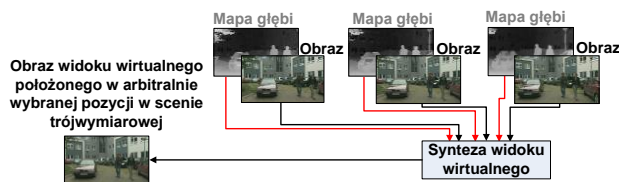
Streszczenie: Artykuł stanowi podsumowanie i streszczenie rozprawy doktorskiej [1]. Przedstawiono osiągnięcia autora związane z opracowaniem nowych technik estymacji głębi oraz nowej techniki kodowania głębi wykorzystującej reprezentację nieliniową, która została włączona do specyfikacji międzynarodowych norm opisujących nowe techniki kodowania obrazu trójwymiarowego. Opisano również udział autora w opracowaniu poznańskiego kodeka obrazu trójwymiarowego oraz eksperymentalnego systemu wielokamerowego i sekwencji testowych.

1. WSTĘP

Rozprawa [1] dotyczy nowej generacji systemów obrazu trójwymiarowego. Swoim zakresem obejmuje zwłaszcza systemy swobodnego punktu widzenia (free viewpoint navigation) i autostereoskopię. W stosunku do dotychczas znanych i stosowanych rozwiązań wykorzystujących stereoskopię, nowe systemy mają zapewnić lepsze wrażenia obrazu przestrzennego, lepszą reprodukcję struktury sceny oraz umożliwić widzowi oddziaływanie na widzianą scenę i jej oglądanie bez użycia specjalnych okularów.

Zarówno nawigacja swobodnego punktu widzenia, jak i autostereoskopia, wymagają znacznej liczby widoków sceny, których akwizycja, przetwarzanie i transmisja byłaby niepraktyczna. Sposobem, który umożliwia efektywny opis sceny trójwymiarowej jest zastosowanie reprezentacji MVD (Multiview Video plus Depth) [2]. Reprezentacja sceny trójwymiarowej w formacie MVD składa się z ograniczonej liczby widoków oraz odpowiadających im map głębi, które przenoszą dodatkową informację na temat trójwymiarowej struktury sceny.

Mapy głębi są obrazami, których wartości odzwierciedlają odległości pomiędzy daną kamerą a punktami w scenie. Wykorzystanie map głębi (rys. 1) pozwala na generowanie tzw. **widoków wirtualnych**. Widok wirtualny, czyli widok pochodzący z kamery umiejscowionej wirtualnie w scenie, nie jest rejestrowany a powstaje syntetycznie na podstawie treści zarejestrowanej przez kamery rzeczywiste. Taka wirtualna kamera może być umieszczona w dowolnym położeniu – także w położeniu nieosiągalnym w rzeczywistym świecie lub w położeniu rzeczywistej kamery. Ta druga możliwość może być wykorzystana przy predykcji określonego widoku na podstawie innych widoków. Jest to kluczowa zaleta reprezentacji MVD, gdyż umożliwia znaczną redukcję liczby bezpośrednio kodowanych widoków np. do 3 widoków, które są przesyłane wraz z odpowiadającymi im mapami głębi. Na tej podstawie pozostałe widoki mogą zostać zrekonstruowane po stronie dekodera/odbiornika.

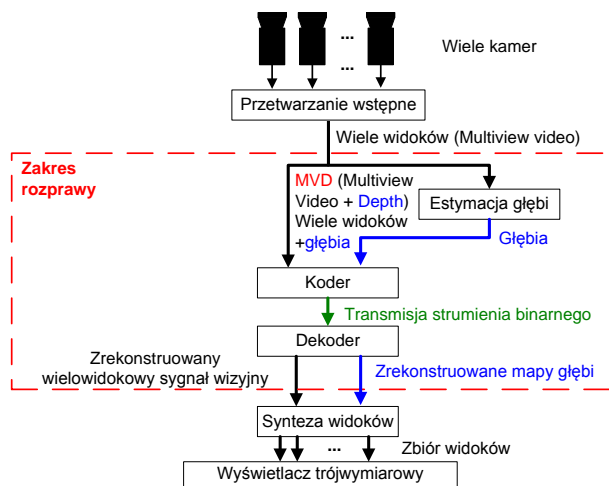


Rys. 1 Wykorzystanie reprezentacji MVD (np. 3 widoki + 3 mapy głębi) do syntezy widoku wirtualnego

2. PROBLEMY BADAWCZE

Wymagane przez reprezentację MVD **mapy głębi trzeba wyznaczyć oraz efektywnie przesłać**. Problemy te są rozważane w rozprawie w odniesieniu do czterech tez: T1÷T4. W szczególności:

1. Mapy głębi wysokiej jakości są potrzebne przy tworzeniu treści zawierających obraz trójwymiarowy. W rozprawie rozpatrywana jest algorytmiczna estymacja głębi, zaś opracowanie nowych algorytmów i technik estymacji map głębi jest jednym z celów pracy (**tezy T1 i T2** rozprawy). Jest to zagadnienie skomplikowane, stanowiące spore **wyzwanie badawcze**, gdyż w przeciwieństwie do np. grafiki komputerowej, rozpatrywane są tutaj złożone sceny naturalne w których występują liczne problemy związane m.in. z odbiciami i przesłonięciami.



Rys. 2 Zakres rozprawy (zaznaczony czerwoną linią) w kontekście systemu trójwymiarowego - estymacja i kompresja map głębi

2. Oczekuje się, iż mapy głębi powinny być **spójne w dziedzinie czasu**. Temat ten jest związany zarówno z estymacją głębi, jak i jej kompresją. Brak spójności czasowej wyznaczanych map głębi objawia się migotaniem w widokach wirtualnych syntezowanych z ich wykorzystaniem. Ponadto, ograniczona spójność map głębi w dziedzinie czasu wpływa negatywnie na spraw-

ność ich kompresji. **Poprawa spójności czasowej** map głębi jest kolejnym z celów rozprawy (teza T3).

3. Stosownie formatu MVD stwarza potrzebę efektywnego przesyłania map głębi. Ponieważ mapy głębi różnią się pod względem charakteru swojej treści od obrazów naturalnych poszukuje się **nowych, wydajnych metod ich kompresji**. Zagadnienia tego dotyczy **teza T4** rozprawy.

3. METODOLOGIA BADAŃ

Celem rozprawy są badania nad poprawą **efektywności kodowania map głębi** i poprawą **jakości map głębi estymowanych algorytmicznie**. W przypadku obu tych problemów **jedyną metodą rzetelnej oceny propozycji jest przeprowadzenie serii eksperymentów wykorzystujących sekwencje testowe**. Z tego względu, na poparcie postawionych tez, w rozprawie przedstawiono wyniki licznych eksperymentów. Aby je przeprowadzić, propozycje przedstawione w rozprawie zaimplementowano i zintegrowano z istniejącymi pakietami oprogramowania, takimi jak Depth Estimation Reference Software (DERS) [3] i 3D-ATM (AVC based 3D Test Model) [4], stosowanymi w literaturze jako odniesienie do badań naukowych w tej dziedzinie. Sumarycznie, podczas tworzenia rozprawy, **autor pracował z oprogramowaniem liczącym około 385 000 linii kodu źródłowego, wśród których około 41 000 zostało stworzonych przez autora**.

Do eksperymentów wykorzystano szeroko stosowany w literaturze zbiór sekwencji testowych. Zbiór ten, wraz z szeregiem wytycznych dotyczących przeprowadzania eksperymentów został opracowany przy **aktywnym udziale autora** w ramach prac międzynarodowej grupy ekspertów MPEG (Motion Picture Experts Group) działającej z ramienia organizacji ISO/IEC.



Rys. 3. Eksperymentalny system wielokamerowy (po lewej) powstały w Katedrze Telekomunikacji Multimedialnej i Mikroelektroniki z aktywnym udziałem autora rozprawy, wykorzystany min. do przygotowania sekwencji testowej „Poznan Hall” (po prawej)

Jednym z elementów wspomnianych prac było wytworzenie przez Katedrę Telekomunikacji Multimedialnej i Mikroelektroniki **eksperymentalnego systemu wielokamerowego** (rys. 3). System ten posłużył do nagrania sekwencji testowych o unikalnych w skali światowej cechach, takich jak wysoka rozdzielczość obrazu Full-HD, naturalność nagranych scen, wykorzystanie powszechnie używanych kamer telewizyjnych. Cechy te zadecydowały, że dwie z nagranych sekwencji zostały włączone do zbioru sekwencji testowych rekomendowanego przez MPEG.

Eksperymenty przeprowadzone przez autora wymagały **sporych nakładów mocy obliczeniowej**. Sumarycznie, przeprowadzenie badań niezbędnych dla ukończenia rozprawy zajęłoby około **6 610 dni nieustannej pracy jednowątkowego komputera dużej mocy**. Ich wykonanie w krótszym czasie, możliwe było jedynie dzięki silnemu sprawnemu zarządzaniu obliczeniami i ich zrównolegleniu na klastrze obliczeniowym.

4. NOWE TECHNIKI ESTYMACJI GŁĘBI

W rozprawie zaproponowano **trzy nowe autorskie techniki** dla estymacji map głębi. Posłużyły one dla wsparcia tez T1-T3 rozprawy, omówionych poniżej.

4.1. Teza T1

Estymacja głębi może być usprawniona dzięki wykorzystaniu modelowania funkcji kosztu bazującego na regule maksymalizacji prawdopodobieństwa a posteriori.

W rozprawie zaproponowano nowe sformułowanie funkcji celu dla algorytmów optymalizacyjnych wyznaczających mapy głębi, np. belief propagation. Dwa składniki tego sformułowania - *DataCost* i *Transition-Cost*, przedstawiono odpowiednio w rozdziałach 3.4 i 3.5 rozprawy.

Posługując się wyprowadzeniem teoretycznym, wykorzystującym regułę maksymalizacji prawdopodobieństwa a posteriori MAP (Maximum A posteriori Probability) oraz badaniami empirycznymi, w rozprawie wykazano, że **wykorzystanie zaproponowanego nowego sformułowania prowadzi do poprawy jakości estymowanych map głębi, wyrażonej wzrostem współczynnika PSNR do 2,8 dB**, liczonego dla widoków wirtualnych syntezowanych z użyciem map głębi wyznaczonych za pomocą zaproponowanej techniki. Odniesieniem dla badań był algorytm DERS reprezentujący aktualny stan wiedzy w zakresie estymacji głębi.

Wspomnieć należy, że jedną z głównych zalet zaproponowanej metody jest to iż, **nie wymaga ona manualnego doboru parametrów sterujących**, tak jak ma to miejsce np. w algorytmie odniesienia DERS.

4.2. Teza T2

Precyzja i dokładność estymowanych map głębi może być zwiększona w kroku przetwarzania końcowego, poprzez iteracyjne wstawianie wartości pośrednich głębi, kontrolowane za pomocą syntezy widoków.

Zaproponowano nową metodę estymacji głębi wykorzystującą autorski algorytm MLH (Mid-level Hypothesis). W rozprawie wykazano, iż jego użycie poprawia precyzję i dokładność estymowanych map głębi.

Zaproponowana metoda jako dane wejściowe wykorzystuje wyznaczoną wcześniej mapę głębi o niskiej precyzji np. wyznaczoną z dokładnością do pełnego okresu próbkowania. Następnie, taka mapa głębi jest przetwarzana za pomocą zaproponowanego algorytmu MLH. Wyszukiwane są sąsiadujące ze sobą punkty, różniące się wartościami głębi o jeden poziom kwantyzacji. Stawiana jest hipoteza, że punkty te mają w rze-

czywistości pośrednią wartość głębi, a więc niereprezentowalną przy początkowej dokładności mapy głębi. Hipoteza ta jest następnie weryfikowana poprzez porównanie jakości widoków wirtualnych syntezowanych z uwzględnieniem początkowej i hipotetycznej wartości głębi. Następnie, jeśli hipoteza zostaje zweryfikowana pozytywnie, hipotetyczna wartość głębi staje się wartością faktyczną, zaś rozpatrywana hipoteza rozszerza się na kolejne sąsiadujące punkty, aż do iteracyjnego sprawdzenia wszystkich hipotez.

Uzyskane wyniki wskazują, że zastosowanie algorytmu MLH prowadzi do dwojakiego rodzaju korzyści, zależnie od sposobu interpretacji wyników. Z jednej strony, pozwala on uzyskać wyższą precyzję i dokładność estymowanych map głębi, co wyraża się **poprawą jakości widoku wirtualnego, zmierzoną wzrostem współczynnika PSNR o około 2 dB**. Z drugiej strony, algorytm MLH pozwala uzyskać porównywalną jakość jak algorytm jednokrokwie estymujący głębię z wysoką precyzją, przy **znacznie mniejszych - około 3,7 razy - nakładach obliczeniowych**.

4.3. Teza T3

Spójność w dziedzinie czasu estymowanych map głębi może być poprawiona poprzez redukcję szumów w wejściowych sekwencjach wielowidokowych.

W rozdziale 3.8 rozprawy przedstawiono badania dotyczące zwiększania spójności czasowej estymowanych map głębi. Zaprezentowano autorską metodę estymacji głębi, polegającą na wykorzystaniu redukcji szumu w sekwencjach wejściowych a priori samej estymacji. Zaproponowaną metodę przebadano eksperymentalnie z wykorzystaniem dwóch technik redukcji szumu, opracowanych przez autora: SBNR (Still Background Noise Reduction) i MCNRR (Motion-Compensated Noise Reduction with Refinement).

Posługując się zbiorem sekwencji testowych, wykazano iż zaproponowana metoda umożliwia uzyskanie **poprawy spójności czasowej estymowanych map głębi**, która zaowocowała **30% redukcją prędkości bitowej**, potrzebnej na przesłanie wyestymowanych map głębi, opisujących z tą samą jakością tę samą scenę trójwymiarową.

5. NOWA METODA NIELINIOWEJ REPRZENTACJI GŁĘBI DLA KODOWANIA

W rozprawie zaprezentowano oryginalną ideę zastosowania nieliniowego przekształcenia dla reprezentacji i kodowania map głębi. Wykazano, że jej wykorzystanie przynosi znaczną redukcję skompresowanego strumienia, polepszając tym samym sprawność kompresji. Wyniki te wspierają tezę T4 rozprawy. Poniżej zawarto streszczenie osiągnięć związanych z tą tezą.

5.1. Teza T4

Nieliniowa reprezentacja głębi umożliwia zwiększenie sprawności kompresji map głębi w systemach obrazu trójwymiarowego.

W pierwszej kolejności, pokazano proste przekształcenie, które znalazło zastosowanie w propozycji

kodeka trójwymiarowego Katedry Telekomunikacji Multimedialnej i Mikroelektroniki (**KTMiM**) zgłoszonej na konkurs „Call for Proposals on 3D Video Coding Technology” zorganizowany przez grupę MPEG w 2011r.

Sukces kodeka odniesiony w konkursie jak i późniejsze badania badające udział poszczególnych technik kompresji, stały się motywacją dla systematycznego opracowania propozycji nieliniowego przekształcenia. Propozycja ta, korzystająca z wyprowadzenia teoretycznego, przedstawiona jest w rozprawie wraz szczegółami implementacji i wynikami weryfikacji eksperymentalnej.

Jak pokazano, **wykorzystanie zaproponowanej techniki kodowania wykorzystującej nieliniową reprezentację głębi NDR (Non-linear Depth Representation) pozwala na uzyskanie około 25% redukcji strumienia bitowego głębi. Przekłada się to na poprawę jakości subiektywnej obrazu wirtualnego o około 1 punkt w 11-punktowej skali MOS.**

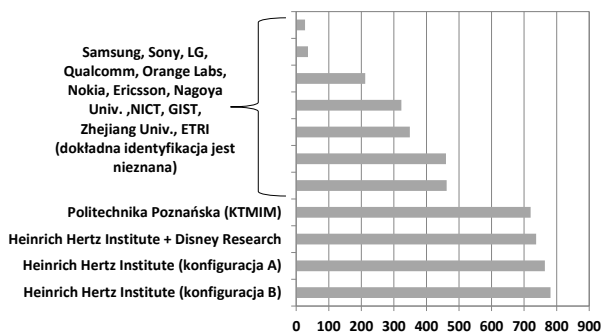
Zaproponowana przez autora technika kodowania wykorzystująca nieliniową reprezentacji głębi NDR (Non-linear Depth Representation) **została przyjęta do nowych międzynarodowych norm** opisujących trójwymiarowe rozszerzenia technik kodowania AVC (ISO/IEC IS 14496 10 i ITU Rec. H.264), znanych pod nazwami MVC+D [5] i AVC-3D [6]. Warto nadmienić, że idea nieliniowej reprezentacji głębi jest tematem zgłoszenia patentowego Politechniki Poznańskiej w Polsce i w Stanach Zjednoczonych.

6. NOWA TECHNIKA KODOWANIA SEKWENCJI TRÓJWYMIAROWYCH

W roku 2011 grupa MPEG, działająca z ramienia ISO/IEC, ogłosiła konkurs „Call for Proposals on 3D Video Coding Technology” (CfP). Jego celem było wyłonienie najlepszych z dostępnych technik kompresji scen trójwymiarowych w celu ich standaryzacji i **wprowadzenia na rynek**.

W konkursie tym wzięły liczne instytuty badawcze, firmy i uniwersytety z całego świata (rys. 4) oraz Katedra Telekomunikacji Multimedialnej i Mikroelektroniki (KTMiM) Politechniki Poznańskiej. Wszystkie zgłoszone propozycje były poddane ocenie obiektywnej i subiektywnej przez niezależne laboratoria badawcze.

Wyniki, podsumowane przez grupę MPEG pod koniec roku 2011, wyłoniły **dwa najlepsze rozwiązania** kodeków opartych na HEVC: przygotowane przez Heinrich Hertz Institute (HHI) we współpracy z Disney Research Labs (w kilku wariantach) oraz przez **KTMiM** Politechniki Poznańskiej. **Wysoki poziom zaproponowanego kodeka z KTMiM** (rys. 4) został w rozprawie przedstawiony na tle innych zgłoszeń na konkurs względem tzw. relacji przewyższania (outranking).



Rys. 4. Relacja przewyższania (outranking) propozycji różnych uczestników konkursu Call for Proposals grupy MPEG

Wykorzystana podczas rozstrzygnięcia konkursu relacja przewyższania (rys. 4) przedstawia ilość przypadków w testach subiektywnych, w których dana propozycja była najlepsza wśród badanych.

Autor rozprawy był **jednym z głównych twórców** wspomnianej propozycji kodeka z KTMiM. Jako że jej opis zawarto już w publikacjach (np. [7], w rozprawie zamieszczono jedynie krótki opis, uwypuklający wkład samego autora:

- Autor rozprawy współtworzył **ogólną strukturę** kodeka.
- Autor koordynował **dobór algorytmów**, które zostały zawarte w kodeku.
- Autor współtworzył schemat kompresji wykorzystujący podział kodowanej treści na **warstwę główną i warstwę szumu (narzędzie Layer separation)**, m.in. przez propozycję oryginalnej koncepcji i dostarczenie autorskiego algorytmu redukcji szumów.
- Autor współuczestniczył w powstawaniu metody kodowania wykorzystującej **zunifikowaną reprezentację głębi (Unified Depth Representation)**, dostarczając m.in. autorski algorytm **MLH (Mid-Level Hypothesis)** opisany w rozdziale 3.7 rozprawy.
- Autor zaproponował i zaimplementował algorytm kodowania wykorzystujący **nieliniowej reprezentacji głębi NDR**, opisany w rozdziale 5 rozprawy.

Wyniki przeprowadzonych badań **wykazaly, że zastosowanie techniki kodowania nieliniową reprezentacją głębi NDR (Non-linear Depth Representation) przyniosło redukcję strumienia bitowego o około 20% względem HEVC (simulcast)**, przy zachowaniu tej samej jakości subiektywnej. Wszystkie techniki kompresji (włączając te nie objęte omawianą rozprawą) **umożliwiają ok. 50%÷60% redukcję strumienia binarnego**. Można więc ocenić, że znaczna część zysków w kodeku trójwymiarowym pochodzi od opisanej w rozprawie autorskiej technice nieliniowej reprezentacji głębi (NDR).

7. PODSUMOWANIE

Omówiona rozprawa [1] przedstawia szereg osiągnięć autora związanych z opracowaniem nowych tech-

nik i algorytmów. Posłużyły one udowodnieniu czterech tez postawionych w rozprawie. Za największy sukces rozprawy uznać należy opracowanie nieliniowa reprezentacji głębi NDR (Non-linear Depth Representation), która może być wykorzystana do zwiększenia efektywności kompresji map głębi w systemach trójwymiarowych. Propozycja tej techniki została włączona do nowych norm opisujących trójwymiarowe rozszerzenia techniki kodowania AVC (ISO/IEC IS 14496-10 i ITU Rec. H.264), znanych pod nazwami MVC+D [5] i AVC-3D [6]. W rozprawie pokazano, że NDR przynosi znaczącą poprawę jakości subiektywnej syntezowanych widoków wirtualnych, wynoszącą nawet do 1 punktu w 11-punktowej skali MOS oraz znaczącą poprawę sprawności kompresji, zapewniając redukcję wielkości strumienia bitowego głębi o około 25%.

Realizacja rozdziałów pracy doktorskiej, dotyczących estymacji głębi, była częściowo sfinansowana ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2012/07/N/ST6/02267.

Realizacja rozdziałów pracy doktorskiej, dotyczących kodowania głębi, była częściowo sfinansowana ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2012/05/B/ST7/01279.

Praca doktorska współfinansowana ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

SPIS LITERATURY

- [1] O. Stankiewicz, „*Stereoscopic depth map estimation and coding techniques for multiview video systems*”, rozprawa doktorska, promotor: prof. dr hab. inż. M. Domański, recenzenci: prof. M. Bober, prof. dr hab. inż. P. Strumiłło, Politechnika Poznańska, Wydział Elektroniki i Telekomunikacji, 2014.
- [2] P. Merkle, A. Smolic, K. Müller, T. Wiegand, „*Multi-view video plus depth representation and coding*”, Proceedings of IEEE International Conference on Image Processing (ICIP'07), Tom I, str. 201-204, USA, 2007.
- [3] M. Tanimoto, T. Fujii, K. Suzuki, „*Video Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching*”, ISO/IEC MPEG M16092, Lozanna, Szwajcaria, luty 2009.
- [4] D. Rusanovskyy, F. C. Chen, L. Zhang, T. Suzuki, „*3D-AVC Test Model 5*”, ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-C1003, Genewa, Szwajcaria, styczeń 2013.
- [5] „*MVC extensions for inclusion of depth maps*”, Amendment 2 of ISO/IEC 14496-10:2012/Amd 2:2012, Wydanie 1, 2012.
- [6] „*AVC compatible video with depth information*”, Amendment 3 of ISO/IEC 14496-10:2013/Amd 3:2013, Wydanie 1, 2013.
- [7] M. Domański, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, J. Stankowski, R. Ratajczak, T. Grajek, „*High Efficiency 3D Video Coding Using New Tools Based on View Synthesis*”, IEEE Transactions on Image Processing, tom 22, nr 9, str. 3517-3527, 2013.