Dawid Mieloch
Adrian Dziembowski
Adam Grzelka
Katedra Telekomunikacji Multimedialnej i Mikroelektroniki
Politechnika Poznańska
Polanka 3, 60-965 Poznań
{dmieloch, adziembowski, agrzelka}@multimedia.edu.pl

KKRRiT 2017
Poznań, 21-23 czerwca 2017

ESTYMACJA GŁĘBI DLA SYSTEMÓW WIELOWIDOKOWYCH

DEPTH ESTIMATION FOR MULTIVIEW SYSTEMS

Streszczenie: **W artykule przedstawiono nowatorską metodę estymacji spójnej międzywidokowo głębi dla systemów wielowidokowych. Proponowana metoda oparta jest na optymalizacji funkcji energii opartej na segmentach, a nie na punktach zarejestrowanych obrazów. Pozwala to na kontrolowanie końcowej jakości map głębi oraz czasu ich estymacji poprzez zmianę liczby użytych segmentów. Przedstawione wyniki eksperymentalne potwierdzają znaczącą poprawę jakości estymowanej głębi względem metody odniesienia udostępnianej przez grupę ekspercką MPEG.**

Abstract: **In this article a new method of an inter-view consistent depth estimation for multiview systems was shown. The proposed method is based on the optimization of the energy function formulated over segments instead of over pixels of input views. The quality of depth maps and the time of their estimation can be controlled through the number of used segments. Results of experiments show significant increase of estimated depth maps quality in comparison to reference method of MPEG group.**

Słowa kluczowe: **mapa głębi, segmentacja obrazu, system wielowidokowy**

Keywords: **depth map, image segmentation, multiview system**

1. WSTĘP

Nowe zastosowania trójwymiarowej reprezentacji zarejestrowanych scen, takie jak telewizja swobodnego punktu widzenia [1][12], czy systemy rzeczywistości wirtualnej, pokazują znaczenie i konieczność dalszego rozwoju metod tworzenia treści trójwymiarowej. Jedną z najpowszechniejszych reprezentacji takich treści jest użycie widoków zarejestrowanych przez kamery wraz z odpowiadającymi im mapami głębi. Mapy głębi, najczęściej reprezentowane jako obrazy w skali szarości, przedstawiają odległość punktów zarejestrowanej sceny od płaszczyzny przetwornika kamery. W tym artykule rozważana będzie estymacja głębi na podstawie widoków zarejestrowanych przez kamery systemu wielokamerowego.

Mapy głębi mogą być estymowane dla każdej kamery niezależnie, korzystając jedynie z sąsiednich widoków (lewego i prawego). Taka metoda nie zapewnia jednak odpowiedniej spójności przestrzennej, ponieważ głębia pewnych obiektów może być wtedy różna dla sąsiednich widoków. Różnice w głębi w różnych widokach są jedną z przyczyn błędnie tworzonych widoków wirtualnych, stanowiących podstawę swobodnej nawigacji dookoła sceny.

Tworzenie map głębi podczas estymacji wielowidokowej, wykorzystującej wszystkie zarejestrowane widoki, zapewnia spójność głębi obiektów w sąsiednich widokach. Jednakże, taka estymacja jest bardzo czasochłonna, szczególnie jeżeli kamery systemu są swobodnie rozstawione dookoła sceny i nie przyjęto żadnych założeń dotyczących ich położenia i liczby. Dla przykładu, metoda [14] pozwala na tworzenie map głębi w czasie rzeczywistym. Niestety, konieczne jest jednak zastosowanie czterech kamer o równoległych osiach optycznych i ściśle określonym rozstawieniu, co zmniejsza liczbę ewentualnych zastosowań takiego systemu.

Czas estymacji jest również ściśle powiązany z rozdzielczością zarejestrowanych widoków. Możliwe jest tworzenie map głębi dla zdecymowanych obrazów wejściowych, co znacząco skraca czas estymacji, a następnie zwiększenie mapy głębi do pełnej rozdzielczości [4]. Tracona jest jednak ponownie spójność głębi w sąsiednich widokach, ponieważ proces zwiększania rozdzielczości jest niezależny dla każdego widoku.

Proponowana metoda ma za zadanie zapewnić możliwość tworzenia map głębi wysokiej jakości dla systemu wielowidokowego o dowolnym rozstawieniu kamer, przy zachowaniu ich pełnej rozdzielczości. Podstawą nowej metody jest oparcie procesu estymacji głębi na segmentach widoków wejściowych, a nie na punktach zarejestrowanych obrazów. Istniejące metody estymacji głębi wykorzystujące segmentację [9][16] mogą być wykorzystywane jedynie dla par kamer i nie zapewniają zmniejszonego czasu obliczeń.

Podstawy estymacji map głębi opartej na optymalizacji funkcji energii opisano w rozdziale 2. W proponowanej metodzie, której szczegółowy opis zawarty jest w rozdziale 3, wykorzystanie segmentacji obrazu zapewnia skalowalność między jakością estymowanych map głębi oraz czasem wykonywania całego procesu. Wyniki przeprowadzonego eksperymentu mającego na celu porównanie proponowanej metody z metodą odniesienia DERS [5], udostępnianą przez międzynarodową grupę ekspercką MPEG, przedstawiono w rozdziale 4.

2. ESTYMACJA MAP GŁĘBI

Estymacja map głębi może być przeprowadzana poprzez optymalizację funkcji energii [3][8][18]. W podstawowej formie funkcja ta składa się z dwóch składników. Pierwszy z nich, błąd dopasowania, obliczany jest dla każdego punktu obrazów wejściowych i wyraża miarę dopasowania danego punktu do punktu w innej kamerze dla aktualnie rozpatrywanej głębi. Aby zapewnić odpowiednią spójność głębi między wszystkimi widokami, błąd dopasowania może być zależny również od głębi punktu w innej kamerze.

Drugi składnik, zwany gładkością, określany jest pomiędzy parami sąsiadujących ze sobą punktów. Określa on wewnątrzobrazowy koszt nieciągłości głębi. Dla rzeczywistych scen większość obszaru mapy głębi powinna być gładka. Jeżeli sąsiadujące punkty mają podobną głębię, to koszt związany z gładkością jest niski.

Optymalizacja funkcji energii jest bardzo czasochłonna i wymaga bardzo dużych zasobów pamięci operacyjnej. Dla przykładu, używając metody DERS [5], w której minimalizowana funkcja energii jest oparta na puntach obrazu, dla obrazów o rozdzielczości Full HD i ograniczeniu rozważanych poziomów głębi do 250, wykorzystana pamięć operacyjna to prawie 9 GB, a czas estymacji to średnio 30 minut.

3. PROPONOWANA METODA

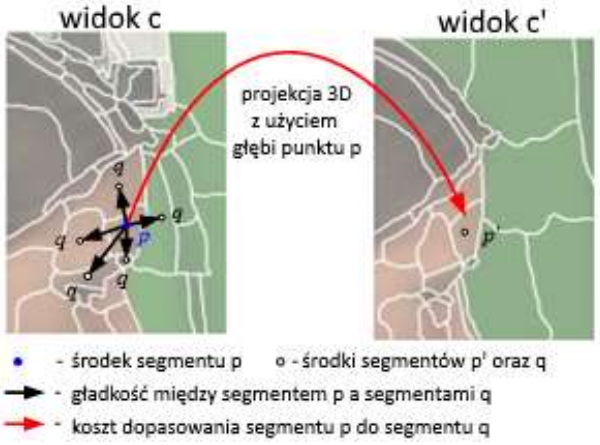
W proponowanej nowej metodzie wyznaczania map głębi dla systemów wielowidokowych optymalizacja funkcji energii przeprowadzana jest dla segmentów obrazu, a nie dla pojedynczych punktów. Tym sposobem, jednakowa głębia jest wyznaczana dla całego segmentu. Liczba użytych segmentów stała się parametrem optymalizacji, który daje możliwość sterowania jakością i czasem estymacji głębi. Dla przykładu, kiedy zostaną użyte małe segmenty, o rozmiarze do 20 punktów, jakość estymowanej głębi jest nadal wysoka, lecz czas estymacji jest znacząco zmniejszony.

Wykorzystanie segmentacji obrazu pozwala również na lepsze odwzorowanie krawędzi obiektów w estymowanych mapach głębi. Błędne odwzorowanie krawędzi w mapach głębi powoduje znaczące zmniejszenie jakości syntezowanych widoków wirtualnych [7].

Błąd dopasowania jest określony dla segmentów, jednak jego wartość jest zależna od punktów obrazu. Pozwala to na zachowanie dokładności wyznaczania głębi porównywalnej do typowej estymacji wykorzystującej funkcję energii opartą na punktach. Błąd dopasowania jest wyrażany jako suma różnic luminancji i chrominancji geometrycznego środka segmentu i punktu w sąsiednim obrazie, który odpowiada opisywanemu środkowi po projekcji 3D z użyciem rozpatrywanej głębi. Optymalizowana funkcja jest sformułowana w następujący sposób:

$$E(d_p) = \sum_{c,c'} \sum_{p,p'} M_{p,p'}(d_p, d_{p'}) + \sum_{p,q} V_{p,q}(d_p, d_q), \quad (1)$$

gdzie p to punkt (będący środkiem pewnego segmentu) w widoku c, który odpowiada punktowi p' w widoku c' (dla głębi punktu p wyrażonej jako d_p). $M_{p,p'}$ to błąd dopasowania punktów p i p', a $V_{p,q}$ to gładkość między segmentami. Szczegółowy opis obliczenia gładkości w adaptacyjny sposób, dopasowany do charakterystyki obrazu wejściowego, opisano w [15]. Optymalizacja funkcji celu jest przeprowadzana z użyciem metody graph cut [18]. Sposób formułowania energii dla przykładowego segmentu obrazu zaprezentowano na rysunku 1.



Rys. 1. Czynniki funkcji energii dla segmentu p

Głębia dla wszystkich kamer jest wyznaczana jednocześnie, umożliwiając spójność głębi w sąsiednich widokach. Co więcej, nie jest narzucany żaden sposób rozmieszczenia kamer systemu. Metoda może być użyta dla dowolnej liczby kamer, zarówno ustawionych na łuku, jak i gdy ich osie optyczne są równoległe.

Wymagana segmentacja jest przeprowadzana dla każdego widoku wejściowego niezależnie, za pomocą metody SLIC [17]. Segmentacja może być przez to dokonana równolegle dla wszystkich widoków wejściowych. W ten sposób czas potrzebny na dokonanie segmentacji obrazu jest krótki i w niewielki sposób wpływa na złożoność całego algorytmu.

4. WYNIKI EKSPERYMENTALNE

Proponowaną metodę porównano z metodą odniesienia DERS. Z uwagi na brak dostępnych map głębi odniesienia dla sekwencji testowych, zdecydowano się na zmierzenie jakości map głębi pośrednio, poprzez syntezę widoków wirtualnych. Używając dwóch widoków oraz odpowiadających im map głębi tworzono widok wirtualny pomiędzy tymi widokami. Syntezy dokonano przy użyciu metody odniesienia VSRS udostępnianej przez grupę MPEG [6]. Widok wirtualny porównywano z rzeczywistym widokiem znajdującym się w tym samym miejscu i określono obiektywną jakość jako PSNR pomiędzy nimi.

W eksperymentach użyto 8 wielowidokowych sekwencji testowych, przedstawionych w tabeli 1. Różniły się one złożonością przedstawianej sceny oraz rozdzielczością widoków. Mapy głębi estymowane były dla 5 widoków i rozpatrywano 250 poziomów głębi. Wyniki eksperymentu przedstawiono w tabeli 2.

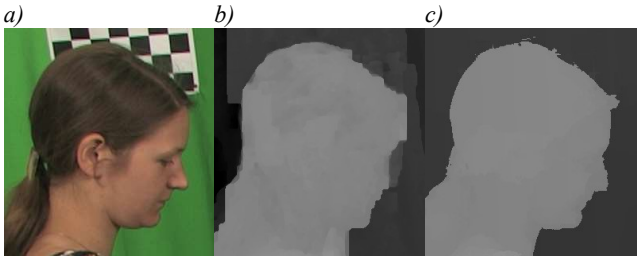
Tab. 1. Wykorzystane sekwencje testowe

Sekwencja testowa	*Rozdzielczość*	*Źródło sekwencji*
Ballet	1024×768	Microsoft Research [10]
Breakdancers		
BBB Butterfly	1280×1080	Holografika [2]
BBB Rabbit		
Poznan Blocks	1920×1080	Politechnika Poznańska [11][13]
Poznan Blocks2		
Poznan Fencing2		
Poznan Service2		

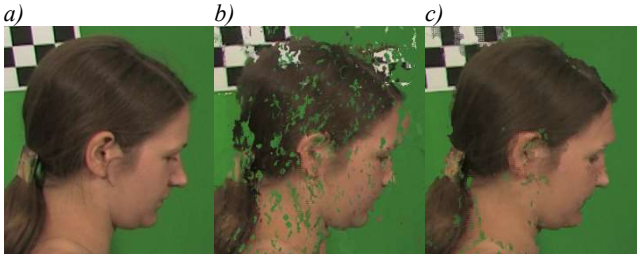
Tab. 2. Porównanie jakości widoków wirtualnych syntezowanych z użyciem map głębi estymowanych przez metodę DERS oraz metodę proponowaną

Sekwencja testowa	*PSNR widoku wirtualnego [dB]*	
	Metoda DERS	*Metoda proponowana*
Ballet	27.81	28.48
Breakdancers	31.81	32.82
BBB Butterfly	29.67	30.36
BBB Rabbit	20.90	24.50
Poznan Blocks	22.58	23.63
Poznan Blocks2	30.59	31.18
Poznan Fencing2	27.53	31.01
Poznan Service2	23.87	25.26
Średnie zwiększenie PSNR: 1.56 dB		

Średni wzrost obiektywnej jakości dla proponowanej metody wyniósł aż 1.56 dB. Zobrazowanie tej różnicy poprzez porównanie fragmentów wyestymowanych map głębi oraz fragmentów wirtualnych widoków stworzonych z ich użyciem dla sekwencji Poznan Blocks przedstawiono na rys. 2 i 3.

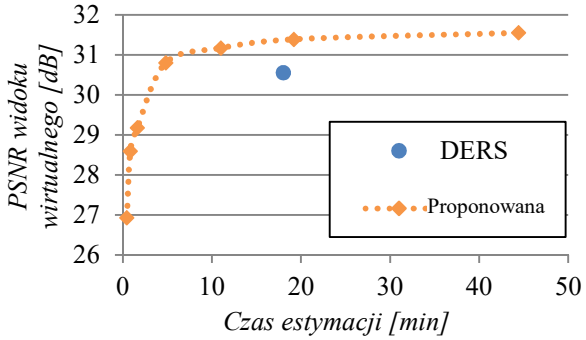


Rys. 2. Porównanie a) fragmentu widoku, b) głębi estymowanej przez DERS, c) głębi estymowanej przez proponowaną metodę

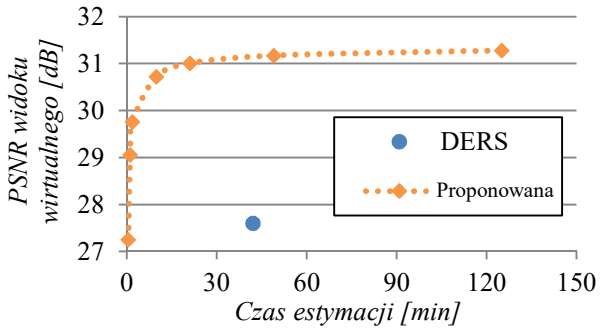


Rys. 3. Porównanie a) fragmentu widoku, b) syntezy z użyciem głębi estymowanej przez DERS, c) syntezy z użyciem głębi estymowanej przez proponowaną metodę

Eksperymenty powtórzono dla różnej liczby segmentów użytych w estymacji dla proponowanej metody, aby określić zależność między czasem estymacji map głębi a jakością widoku wirtualnego. Wyniki tego eksperymentu przeprowadzonego dla sekwencji Poznan Blocks2 oraz Poznan Fencing2 przedstawiono na rysunkach 4 i 5.



Rys. 4. Porównanie jakości widoku wirtualnego i czasu estymacji map głębi dla metody DERS oraz metody proponowanej dla sekwencji Poznan Blocks2



Rys. 5. Porównanie jakości widoku wirtualnego i czasu estymacji map głębi dla metody DERS oraz metody proponowanej dla sekwencji Poznan Fencing2

Przedstawione wyniki pokazują, że dla zaprezentowanej metody możliwe jest uzyskanie lepszej jakości estymowanych map głębi dla znacznie krótszego czasu estymacji. Dla sekwencji Poznan Blocks2 aby uzyskać taką samą jakość estymowanej głębi jak dla metody DERS, proponowana metoda wymaga pięciokrotnie mniej czasu. Dla sekwencji Poznan Fencing2 redukcja czasu obliczeń jest jest prawie 40-krotna, co pozwala na uzyskanie dobrej jakości mapy głębi w czasie krótszym niż 30 sekund. Co więcej, eksperymenty przeprowadzono wykorzystując niezoptymalizowaną wersję oprogramowania, nie używającej żadnej metody zrównoleglania obliczeń.

Dla proponowanej metody zapotrzebowanie na pamięć operacyjną wynosi 0,5 GB dla jednoczesnej estymacji map głębi dla 5 widoków. Dla metody DERS wymagane jest 9 GB pamięci dla każdego z widoków.

5. PODSUMOWANIE

W artykule przedstawiono nową metodę estymacji map głębi dla systemów wielokamerowych. Przedstawiony sposób sformułowania optymalizowanej funkcji

energii pozwala na uzyskanie lepszej jakości estymowanych map głębi przy znacznej redukcji złożoności całego procesu.

Przeprowadzone badania eksperymentalne potwierdzają wysoką jakość estymowanej głębi dla proponowanej metody, wyższą niż dla metody odniesienia zaproponowanej przez międzynarodową grupę ekspercką MPEG. Użycie map głębi estymowanych przez proponowaną metodę pozwala na stworzenie wirtualnych widoków o zadowalającej jakości dla widza telewizji swobodnego punktu widzenia.

Znaczące zmniejszenie czasu potrzebnego na estymację map głębi i zapotrzebowania na pamięć operacyjną przybliża nas do zastosowania metod algorytmicznego wyznaczania map głębi na podstawie zarejestrowanego obrazu w praktycznych systemach wizyjnych.

PODZIĘKOWANIA

Praca finansowana ze środków przyznanych przez Ministerstwo Nauki i Szkolnictwa Wyższego na działalność statutową polegającą na prowadzeniu badań naukowych lub prac rozwojowych oraz zadań z nimi związanych, służących rozwojowi młodych naukowców oraz uczestników studiów doktoranckich.

LITERATURA

[1] Domański Marek, Dziembowski Adrian, Mieloch Dawid, Łuczak Adam, Stankiewicz Olgierd, Wegner Krzysztof. 2015. "A Practical Approach to Acquisition and Processing of Free Viewpoint Video". *31st Picture Coding Symposium PCS 2015*: 10-14.

[2] Kovács Peter. 2015. "Big Buck Bunny light-field test sequences". *Dokument ISO/IEC JTC1/SC29/WG11, MPEG M35721.*

[3] Stankiewicz Olgierd, Wegner Krzystof, Domański Marek. 2016. „Depth estimation based on Maximization of A posteriori Probability". *International Conference on Computer Vision and Graphics ICCVG 2016.*

[4] Dziembowski Adrian, Grzelka Adam, Mieloch Dawid, Stankiewicz Olgierd, Domański Marek. 2016. „Depth map upsampling and refinement for FTV systems". *2016 International Conference on Signals and Electronic Systems*.

[5] Stankiewicz Olgierd, Wegner Krzysztof, Tanimoto Masayuki, Domański Marek. 2013. "Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television". *Dokument ISO/IEC JTC1/SC29/WG11 MPEG M31518.*

[6] Stankiewicz Olgierd, Wegner Krzysztof, Tanimoto Masayuki, Domański Marek. 2013. "Enhanced view synthesis reference software (VSRS) for Free-viewpoint Television". *Dokument ISO/IEC JTC 1/SC 29/WG 11, MPEG M31520.*

[7] Fang L., Xiang Y., Cheung N.M., Wu F. 2016. "Estimation of Virtual View Synthesis Distortion Toward Virtual View Position*", IEEE Transactions on Image Processing*, 25 (5): 1961-1976.

[8] Boykov Y., Veksler O., Zabih R. 2001. "Fast approximate energy minimization via graph cuts". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (1): 1222–1239.

[9] Bleyer M., Gelautz M. 2005. "Graph-based surface reconstruction from stereo pairs using image segmentation". *Proceedings of SPIE - The International Society for Optical Engineering* 5665: 288–299.

[10] Zitnick L., Kang S.B., Uyttendaele M., Winder S., Szeliski R. 2004. "High-quality video view interpolation using a layered representation". *ACM SIGGRAPH Conference Proceedings*: 600–608.

[11] Domański Marek, Dziembowski Adrian, Grzelka Adam, Mieloch Dawid, Stankiewicz Olgierd, Wegner Krzysztof. 2016. "Multiview test video sequences for free navigation exploration obtained using pairs of cameras", *Dokument ISO/IEC JTC1/SC29/WG11, MPEG M38247*.

[12] Lafruit G., Domański M., Wegner K., Grajek T., Senoh T., Jung J., Kovács P., Goorts P., Jorissen L., Munteanu A., Ceulemans B., Carballeira P., García S., Tanimoto M. 2016. "New visual coding exploration in MPEG: Super-MultiView and Free Navigation in Free viewpoint TV". *IST Electronic Imaging, Stereoscopic Displays and Applications XXVII:* 14–18.

[13] Domański Marek, Dziembowski Adrian, Kurc Maciej, Łuczak Adam, Mieloch Dawid, Siast Jakub, Stankiewicz Olgierd, Wegner Krzysztof. 2015. "Poznan University of Technology test multiview video sequences acquired with circular camera arrangement – "Poznan Team" and "Poznan Blocks" sequences". *Dokument ISO/IEC JTC1/SC29/WG11, MPEG M35846.*

[14] Zilly F., Riechert C., Muller M., Eisert P., Sikora T., Kauff P. 2014. "Real-time generation of multi-view video plus depth content using mixed narrow and wide baseline". *Journal of Visual Communication and Image Representation*, 25 (4): 632–648.

[15] Mieloch Dawid, Dziembowski Adrian, Grzelka Adam. 2016. „Segmentacja obrazu w estymacji map głębi". *Przegląd Telekomunikacyjny*, 88 (6): 241-244.

[16] Hong L., Chen G. 2004. "Segment-based stereo matching using graph cuts". *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[17] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Susstrunk S. 2012. "SLIC superpixels compared to state-of-the-art superpixel methods". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (11): 2274-2282.

[18] Kolmogorov V., Zabih R. 2004. "What energy functions can be minimized via graph cuts?". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (2): 147–159.