



SEGMENTACJA OBRAZU W ESTYMACJI MAP GŁĘBI

IMAGE SEGMENTATION IN ESTIMATION OF DEPTH MAPS

Streszczenie: W artykule przedstawiono metodę wykorzystania segmentacji obrazu używając superpikselizacji w estymacji map głębi. Odpowiednie dobieranie współczynnika gładkości na podstawie lokalnej charakterystyki obrazu pozwala na jednoczesne zmniejszenie czasu estymacji map oraz zwiększenie dokładności ich optymalizacji. Pozwala to na uzyskanie dobrej jakości wirtualnych widoków w celach swobodnej nawigacji między kamerami systemu telewizji swobodnego punktu widzenia.

Abstract: In this paper method of depth map estimation using superpixel image segmentation is presented. Optimal selection of smoothing coefficient, based on local image characteristics, simultaneously decreases time of depth map estimation and improves its accuracy. Better quality of depth maps allows to calculate proper virtual views for free navigation in free-viewpoint television systems.

Słowa kluczowe: mapa głębi, segmentacja obrazu, superpiksele, telewizja swobodnego punktu widzenia

Keywords: depth map, image segmentation, superpixels, free-viewpoint television

1. WSTĘP

Podstawową cechą systemów telewizji swobodnego punktu widzenia jest możliwość swobodnej nawigacji wokół zarejestrowanej sceny. Systemy takie składają się zazwyczaj z co najmniej 10 kamer rozmieszczonych na łuku, bądź tak aby ich osie optyczne były równoległe [1]. W celu umożliwienia swobodnej nawigacji konieczna jest synteza widoków wirtualnych znajdujących się w miejscach innych niż pozycje kamer rzeczywistych.

Badania nad systemami FTV stanowią obecnie jeden z najchętniej opisywanych zagadnień dotyczących telewizji, co ma odzwierciedlenie w działalności utworzonej podgrupy MPEG-FTV międzynarodowej grupy standaryzacyjnej MPEG [2]. Najczęściej przedstawiane proponowane zastosowania takich systemów obejmują transmisje z wydarzeń sportowych (stadiony piłkarskie, hale sportowe), czy kulturalnych.

Synteza widoków wirtualnych oparta jest na tak zwanych mapach głębi, które to określają odległość zarejestrowanego punktu obrazu od płaszczyzny przetwornika światłoczułego kamery [3]. Estymacja map głębi jest najbardziej złożonym z etapów przetwarzania

obrazu w telewizji swobodnego punktu widzenia. Dodatkowo, w najbardziej praktycznej z grup tych systemów, w której niewielka liczba kamer jest rozstawiona niezależnie dookoła sceny [4], w wyniku dużych odległości między kamerami estymacja głębi jest znacznie trudniejsza.

Istnieją metody wyznaczania map głębi wykorzystujące aktywne kamery głębi (np. szeroko rozpowszechniony Microsoft Kinect [5]), jednak nie są one wystarczająco uniwersalne, gdyż kamery te nie pozwalają na wykorzystywanie ich na zewnątrz budynków oraz utrudniają synchronizację systemu.

Proponowany algorytm zwiększa jakość map głębi w systemach pasywnych, wykorzystujących jedynie obraz zarejestrowany przez kamery. Dodatkowo, przedstawiona metoda zmniejsza czas optymalizacji map głębi.

2. ESTYMACJA MAP GŁĘBI

2.1. Metoda estymacji map głębi

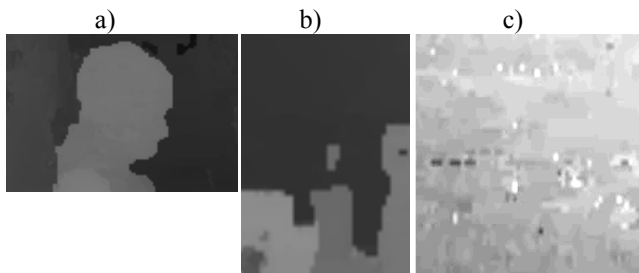
Powszechnie wykorzystywaną techniką estymacji map głębi jest metoda oparta na algorytmie optymalizacyjnym GraphCut [6], która została zawarta w oprogramowaniu odniesienia DERS grupy MPEG [7]. Na potrzeby systemu telewizji swobodnego punktu widzenia metoda ta została rozwinięta tak, aby generowana była głębia spójna międzywidokowo.

W celu estymacji mapy głębi tworzony jest graf, którego każdy z wierzchołków odpowiada jednemu punktowi jednej kamery systemu. Istnieją dwa rodzaje połączeń wierzchołków: międzykamerowe, które wiążą ze sobą potencjalnie odpowiadające ze sobą punkty w sąsiadujących kamerach, oraz wewnątrzkamerowe, które wiążą ze sobą sąsiadujące punkty w jednym obrazie [4]. Wartość wiązań międzykamerowych zależy od różnicy pomiędzy kolorami punktów, a wiązania wewnątrzkamerowe zależne są od obecnej głębi punktu oraz stałego współczynnika gładkości, określanego dla każdej sekwencji na podstawie dynamiki jej głębi.

2.2. Analiza jakości map głębi

Podejście przedstawione w poprzednim podrozdziale powodować może szereg błędów w wyestymowanej mapie głębi. W zależności od tego, jak zostanie dobrany współczynnik gładkości, sylwetki obiektów mogą

być zbyt duże (rys. 1a), małym obiektom może zostać nadana głęboka tła (rys. 1b), bądź w przypadku niskiego współczynnika gładkości – na gładkich obszarach może pojawić się szum (rys. 1c).



Rys. 1. Najczęstsze błędy estymacji mapy głębi.

Błędy w położeniu krawędzi są szczególnie dokuczliwe podczas swobodnej nawigacji wokół sceny, gdyż powodują pojawianie się małych, nieistniejących w scenie obiektów, przemieszczających się przed kamerą. Konieczna jest zatem taka zmiana sposobu doboru współczynnika gładkości, by zachować istniejące w obrazie krawędzie oraz zachować gładkość obszarów.

3. ADAPTACYJNY WSPÓLCZYNNIK GŁADKOŚCI

W celu poprawy jakości estymowanych map głębi możliwe jest uzależnienie wartości współczynnika gładkości od lokalnej charakterystyki obrazu. Dla dwóch sąsiadujących ze sobą punktów obrazu należących do tej samej gładkiej powierzchni współczynnik gładkości powinien mieć dużą wartość. Jeżeli dwa punkty rozdzielone są krawędzią, współczynnik powinien mieć wartość odpowiednio mniejszą. W ten sposób zapewnia się w wyliczonych mapach głębi jednocześnie gładkość powierzchni oraz zachowanie istotnych krawędzi obrazu.

Powszechnie znane są metody wyznaczania krawędzi w obrazie (np. filtracja Canny’ego lub Sobela [8]), jednak tworzone w ten sposób obszary w obrazie są zbyt duże. Odpowiednim narzędziem do stworzenia małych, percepcyjnie spójnych obszarów jest wykorzystanie tak zwanej superpikselizacji obrazu [9].

3.1. Superpikselizacja obrazu

Metoda segmentacji obrazu zwana superpikselizacją jest oparta na kwantowaniu wektorowym metodą Linde-Buzo-Graya [10]. W pierwszym kroku obraz dzielony jest na równe kwadratowe obszary. Każdy z obszarów k jest definiowany za pomocą wektora C_k złożonego ze średniego koloru (w przestrzeni kolorów YUV) punktów wchodzących w jego skład oraz ich średnich położen w płaszczyźnie x oraz y :

$$C_k = [Y_k, U_k, V_k, x_k, y_k] \quad (1)$$

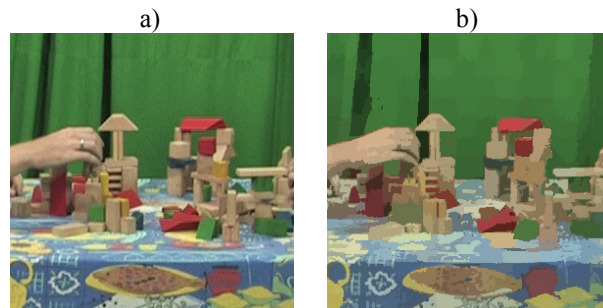
Dla każdego obszaru i obliczana jest odległość D do sąsiednich punktów obrazu j , wyrażona wzorem:

$$D(i, j) = \sqrt{\left(\frac{\sqrt{(y_j - y_i)^2 + (u_j - u_i)^2 + (v_j - v_i)^2}}{m}\right)^2 + \left(\frac{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}{s}\right)^2} \quad (2)$$

gdzie m oznacza współczynnik koloru a S liczbę superpiksela. Następnie sprawdzane jest, do którego

z superpiksela i znajdujących się w sąsiedztwie punktu j obliczona odległość $D(i, j)$ jest najmniejsza. Odnaleziony superpiksel staje się superpikselem punktu j . Przedstawiony proces jest iteracyjny i kończy się gdy żaden z punktów obrazu nie zmieni już swojego przyporządkowania.

Ostatni etap to przyporządkowanie każdemu z punktów obrazu średniego koloru superpiksela do którego punkt ten został przyporządkowany. Końcowy efekt superpikselizacji obrazu widoczny jest na rys. 1.



Rys. 2. a) Obraz zarejestrowany kamerą; b) obraz po segmentacji z użyciem superpiksela.

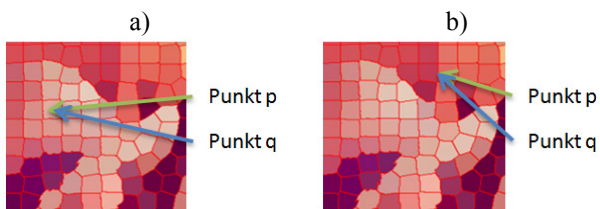
Przedstawiona metoda segmentacji pozwala na określenie ilości superpiksela odpowiedniej do zastosowania (np. dla segmentacji obiektów preferowana jest mała liczba superpiksela) oraz współczynnika koloru określającego wpływ koloru na kształt segmentów (im większa jego wartość, tym bardziej regularny kształt superpiksela). W przedstawianym rozwiązaniu optymalnym wyborem jest użycie dużej liczby segmentów o nieregularnej strukturze. Ten sposób daje najlepsze odwzorowanie krawędzi obrazu, kluczowych podczas estymacji map głębi.

Wyróżnia się wiele metod tworzenia superpiksela, jednak przedstawiona i wybrana do badań metoda zapewnia największą dokładność segmentacji oraz jest metodą najszybszą [9]. Szybkość segmentacji jest kluczową kwestią, gdyż nie powinna ona w znaczący sposób wpływać na czas potrzebnego na cały ciąg przetwarzania obrazu wielowidokowego.

Istnieją już metody zwiększania dokładności map głębi wykorzystujące superpikselizację obrazu [11,12], jednak skupiają się one na poprawie jakości istniejących map głębi (na zasadzie przetwarzania końcowego). Proponowana metoda zwiększa jakość map głębi na etapie ich estymacji, zatem może być wykorzystywana w połączeniu ze wspomnianymi metodami.

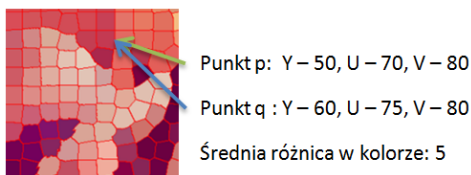
3.2. Algorytm estymacji adaptacyjnego współczynnika gładkości

W celu obliczenia współczynnika gładkości między dwoma punktami w obrazie (wierzchołkami optymalizowanego grafu), w pierwszym kroku przeprowadzana jest segmentacja z użyciem superpiksela na każdym z zarejestrowanych widoków. Jeżeli dwa sąsiadujące ze sobą punkty znajdują się wewnątrz tego samego superpiksela (rys. 3a), to używany jest wejściowy współczynnik gładkości, dostosowany do dynamiki głębi sceny.



Rys. 3. a) Sasiadujące punkty obrazu znajdujące się wewnątrz tego samego superpiksela. b) Punkty na granicy dwóch superpikseli.

Jeżeli dwa punktu rozdzielone są krawędzią (czyli należą do dwóch różnych obszarów – rys. 3b), to współczynnik gładkości dzielony jest przez średnią różnicę w kolorze tych punktów (rys. 4).



Rys. 4. Sposób obliczania średniej różnicy koloru sąsiadujących punktów.

Taki sposób obliczania współczynnika gładkości gwarantuje po pierwsze zmniejszenie wpływu szumu w zarejestrowanej sekwencji na jakość estymowanej głębi. Po drugie, zmniejszenie współczynnika gładkości na krawędziach pozwala na szybsze rozwiązanie problemu optymalizacyjnego za pomocą stworzonego z punktów obrazów grafu.

4. WYNIKI EKSPERYMENTALNE

W celu zbadania poprawy jakości estymowanych map głębi konieczne jest przeprowadzenie syntezy widoku wirtualnego w pozycji kamery rzeczywistej. W ten sposób, wykorzystując obiektywną miarę jakości obrazu PSNR, można określić czy wyliczona głębia odzwierciedla rzeczywistą trójwymiarową strukturę sceny. Syntezy widoku wirtualnego dokonano za pomocą oprogramowania odniesienia VSRS [13].

Do eksperymentu użyto 6 sekwencji wielowidokowych. Dwie z nich (Poznan_Blocks oraz Fencing) zostały zarejestrowane za pomocą systemu telewizji swobodnego punktu widzenia [4]. Pozostałe są sekwencjami syntetycznymi używanymi jako sekwencje testowe podgrupy MPEG-FTV grupy standaryzacyjnej MPEG (BBB Flowers, BBB Butterfly [14] oraz San Miguel [15]). Wykorzystane sekwencje różnią się swoją charakterystyką (dynamiką głębi, liczbą obiektów).

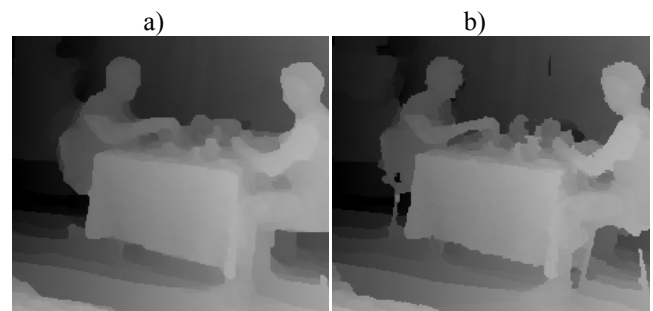
W celu wygenerowania map głębi używano oprogramowania opracowanego przez Katedrę Telekomunikacji i Mikroelektroniki, które zostało dostosowane poprzez implementację opisywanego algorytmu segmentacji obrazu. Dla każdej z sekwencji testowych generowano głębię z użyciem czterech widoków z kamer rzeczywistych. Następnie dokonywano syntezy widoku znajdującego się pośrodku wybranych kamer i porównywano otrzymany przez syntezę widok z obrazem

z kamery rzeczywistej. Otrzymane wyniki przedstawione zostały w tabeli 1.

Tabela 1. Wyniki badań poprawy jakości przy użyciu adaptacyjnego współczynnika gładkości.

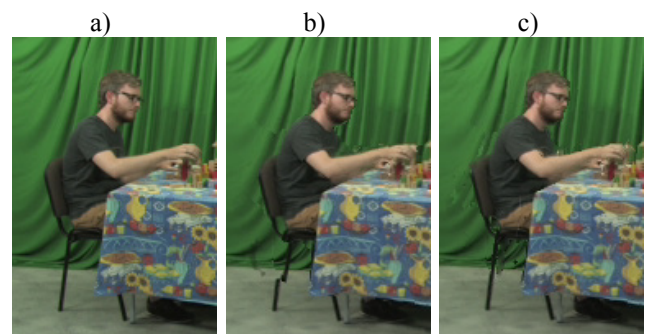
Nazwa sekwencji	PSNR dla stałego współczynnika gładkości [dB]	PSNR dla adaptacyjnego współczynnika gładkości [dB]
BBB Flowers	26,66	27,21
BBB Butterfly	36,01	36,46
San Miguel	28,86	29,15
Poznan_Blocks	32,09	32,42
Fencing	31,36	32,91

Na rys. 5 przedstawiono mapę głębi dla jednego z widoków sekwencji Poznan_Blocks dla dotychczas stosowanego algorytmu estymacji głębi oraz dla algorytmu proponowanego, wykorzystującego segmentację obrazu. Zauważyć można zdecydowanie lepsze odwzorowanie rzeczywistych krawędzi w mapie głębi dla adaptacyjnego współczynnika gładkości, przy jednoczesnym zachowaniu jednolitej głębi dla powierzchni gładkich.



Rys. 5. a) Głębia obliczona ze stałym współczynnikiem gładkości. b) Głębia obliczona z adaptacyjnym współczynnikiem gładkości

Przykładowa różnica w zsyntezowanych widokach wirtualnych dla sekwencji Poznan_Blocks przedstawiona została na rys. 6.



Rys. 6. a) Oryginalny widok z sekwencji Blocks. b) Widok zsyntezowany z użyciem głębi ze stałym współczynnikiem gładkości. c) Widok zsyntezowany z użyciem głębi z adaptacyjnym współczynnikiem gładkości.

Dodatkowo, poza poprawą jakości otrzymanych map głębi, znacznemu przyspieszeniu uległ czas ich estymacji (szczególnie dla sekwencji rzeczywistych), co przedstawiono w tabeli 2.

Tabela 2. Wyniki badań długości estymacji głębi przy użyciu adaptacyjnego współczynnika gładkości.

Nazwa sekwencji	Czas estymacji dla stałego współczynnika gładkości [s]	Czas estymacji dla adaptacyjnego współczynnika gładkości [s]
BBB Flowers	800	770
BBB Butterfly	896	748
San Miguel	605	601
Poznan_Blocks	899	721
Fencing	951	798

5. PODSUMOWANIE

W artykule przedstawiono metodę estymacji map głębi dla systemów telewizji swobodnego punktu widzenia wykorzystującą segmentację obrazu superpikselami. Zaprezentowany sposób obliczenia adaptacyjnego współczynnika gładkości używanego podczas estymacji pozwolił na poprawę dokładności map głębi. Tym sposobem, z wirtualnych widoków wyeliminowane zostały najbardziej dokuczliwe błędy, które zmniejszały znacząco jakość obrazu przy swobodnej nawigacji dookoła sceny. Poprawa jakości wystąpiła dla wszystkich użytych sekwencji testowych, różniących się od siebie wartością oraz dynamiką głębi, co podkreśla uniwersalność przedstawionego rozwiązania.

Proponowana metoda jest całkowicie niezależna od istniejących metod poprawy jakości map głębi poprzez przetwarzanie końcowe. Co więcej, znaczącemu skróceniu (nawet o 15 – 20% dla sekwencji rzeczywistych) uległ czas estymacji map głębi, będący najbardziej złożonym obliczeniowo etapem przetwarzania sekwencji wielowidokowych. Zmniejszenie czasu potrzebnego na estymację jest w tej chwili kluczowym zagadnieniem w dalszych pracach nad systemami telewizji swobodnego punktu widzenia.

PODZIĘKOWANIA

Praca finansowana ze środków przyznanych przez Ministerstwo Nauki i Szkolnictwa Wzszego 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] Tanimoto M. 2006. "Overview of Free Viewpoint Television". *Signal Processing: Image Communication*, 21(6) :454-461.
- [2] Tanimoto M., Senoh T., Naito S., Shimizu S., Horimai H., Domański M., Vetro A., Preda M., Mueller K. "Proposal on a new activity for the third phase of FTV". *Dokument ISO/IEC JTC 1/SC 29/WG 11*, numer MPEG 2013/M30229/M30232.
- [3] Zinger S., Do L., de With P. 2010. "Free-viewpoint depth image based rendering". *Journal Visual Communication and Image Representation*, 21 : 533-541.
- [4] Domański M., Dziembowski A., Mieloch D., Łuczak A., Stankiewicz, O., Wegner K. 2015. „A practical approach to acquisition and processing of free viewpoint video”. *IEEE Conference Publications: Picture Coding Symposium 2015* : 10-14.
- [5] Yang R., Chan Y. H., Gong R., Nguyen M., Strozzi A., Delmas P., Gimel'farb G., Ababou R. 2013. "Multi-Kinect scene reconstruction: Calibration and depth inconsistencies". *28th International Conference of Image and Vision Computing New Zealand*: 47-52.
- [6] Kolmogorov V. 2004. "What energy functions can be minimized via graph cuts". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (2) : 147-159.
- [7] Stankiewicz O., Wegner K., Tanimoto M., Domański M. 2013. "Enhanced depth estimation reference software (DERS) for Free-viewpoint Television". *Dokument ISO/IEC JTC 1/SC 29/WG 11*, numer MPEG2013/M31518.
- [8] Shrivakshan G.T., Chandrasekar C. 2012. "A Comparison of various Edge Detection Techniques used in Image Processing". *IJCSI International Journal of Computer Science Issues*, 9 (5) : 269-276.
- [9] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Süsstrunk S. 2012, „SLIC Superpixels Compared to State-of-the-art Superpixel Methods”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11) : 2274-2282.
- [10] Linde Y., Buzo A., Gray R. 1980. „An Algorithm for Vector Quantizer Design”. *IEEE Transactions on Communications*, 28(1) : 84-95.
- [11] Schmeing M., Jiang X. 2014. "Superpixel-based disocclusion filling in depth image based rendering". *Proceedings of 22nd International Conference on Pattern Recognition* : 1073-1078.
- [12] Van den Bergh M., Carton D., Van Gool L. 2013. "Depth SEEDS: Recovering incomplete depth data using superpixels". *IEEE Workshop on Applications of Computer Vision* : 363 – 368.
- [13] Stankiewicz O., Wegner K., Tanimoto M., Domański M. 2013. "Enhanced view synthesis reference software (VSRS) for Free-viewpoint Television". *Dokument ISO/IEC JTC 1/SC 29/WG 11*, numer MPEG2013/M31520.
- [14] Kovacs, P. 2015. "[FTV AHG] Big Buck Bunny light-field test sequences". *Dokument ISO/IEC JTC1/SC29/WG11*, numer MPEG2015/M35721.
- [15] Goorts P., Javadi M., Rogmans S., Lafruit G. 2014. "San miguel test images with depth ground truth". *Dokument ISO/IEC JTC1/SC29/WG11*, numer MPEG2014/M33163.