

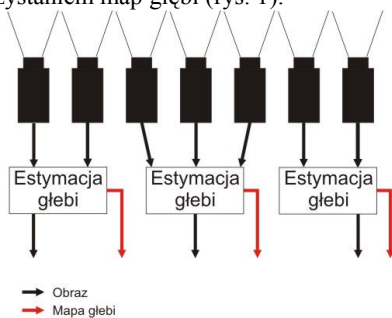
SYSTEM TELEWIZJI STEREOWIZYJNEJ Z WYZNACZANIEM MAPY GŁĘBI

Streszczenie: W artykule przedstawiono propozycję systemu telewizji stereowizyjnej. Omówiono ogólny schemat systemu oraz szczegółowo opisano autorską technikę wyznaczania map rozbieżności wykorzystującą hierarchiczne pasowanie bloków oraz przepływ optyczny. W części dotyczącej wyników przedstawiono wyniki testów przeprowadzonych na wdrożonej przez autorów części systemu.

1. WSTĘP

Na świecie prowadzi się intensywne badania dotyczące telewizji swobodnego punktu widzenia (*Free-view Television – FTV*), która może mieć zastosowanie w systemach umożliwiających nawigację przez użytkownika w scenie trójwymiarowej lub w systemach wykorzystujących monitory autostereoskopowe. Rozważania w tej pracy przeprowadzono pod kątem tego drugiego zastosowania.

Monitor autostereoskopowy umożliwia wyświetlanie obrazów stereoskopowych bez konieczności noszenia jakichkolwiek okularów. Dzięki zastosowaniu pionowo ułożonych pasków walcowych (w kształcie soczewek lenticularnych), pokrywających ekran monitora, do każdego z oczu widza dostarczany jest odpowiedni obraz dający złudzenie przestrzennego obrazu. Monitor taki wymaga dostarczenia wielu obrazów, z których do każdego widza (w zależności od jego położenia) trafi tylko odpowiednia para stereoskopowa. Aby zmniejszyć wymagania transmisyjne usiłuje się przesyłać tylko wybrane widoki, pozostałe resyntezyzując z wykorzystaniem map głębi (rys. 1).



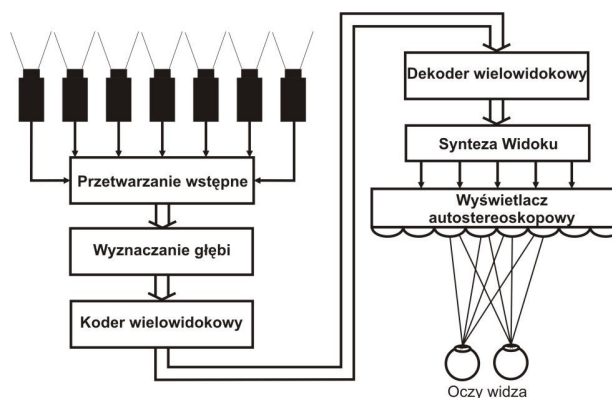
Rys. 1. Redukcja strumienia danych z wykorzystaniem map głębi

2. OPIS SYSTEMU

2.1. System transmisji

Kompletny system telewizji stereowizyjnej składa się z szeregu bloków przedstawionych na rys. 2. Moduł akwizycji zbudowany jest z szeregu kamer dokonujących rejestracji obrazów reprezentujących kolejne widoki.

Obecnie [5] najczęściej przyjmuje się liniowe ułożenie kamer z zachowaniem stałych odstępów pomiędzy obiektywami. Założenie to upraszcza proces estymacji mapy głębi.



Rys. 2. Diagram systemu telewizji stereowizyjnej

Ze względu na trudności lub wręcz niemożliwość uzyskania identycznych parametrów wszystkich kamer (balansu bieli, ostrości, czasu ekspozycji), jak również problemy związane z właściwą kalibracją pozycji kamer, konieczne jest dokonanie przetwarzania wstępnego przed jakąkolwiek dalszą obróbką zebranego materiału. Zazwyczaj dokonywany jest wtedy proces rektyfikacji oraz wyrównania poziomu oświetlenia i wprowadzana jest korekta balansu bieli [6].

Kolejnym krokiem jest wyznaczenie map głębi, określających odległość poszczególnych punktów sceny, obserwowanych przez kamery od ich obiektywów. Mapy te wykorzystywane są w kompresji, dekompresji i resyntezie nowych widoków.

Kompresja poprzedza transmisję w systemie i ma na celu redukcję wymaganej przepływności kanału transmisyjnego. Obrazy są rekonstruowane w odbiorniku w module dekompresji i przygotowywane do wyświetlania w module syntezy widoków. Ostatnim blokiem systemu jest wyświetlacz stereowizyjny (najczęściej autostereoskopowy) umożliwiający wyświetlanie widoku żądanego przez widza.

2.2. Hybrydowa technika estymacji głębi

Problematyka wyznaczania map głębi jest znana i szeroko rozważana w literaturze. Wykorzystywane obecnie techniki opierają się m.in. na dopasowywaniu cech [7], pasowaniu bloków [2], po których następuje postprocessing mający na celu poprawę jakości

uzyskanej mapy głębi. W tym celu stosuje się algorytmy „belief propagation” [3], algorytm Viterbiego (lub ogólnie metody programowania dynamicznego [1]) a także teorię grafów [4]. W rozważanym systemie zaproponowano rozwiązanie opierające się na autorskiej technice hybrydowej, wykorzystującej hierarchiczne pasowanie bloków o adaptacyjnym kształcie oraz gradientowy przepływ optyczny. Technika ta jest opisana w punkcie 3 zaś jej pierwotna idea w naszej pracy [8].

2.3. Kompresja obrazów stereowizyjnych

Istnieją dwa schematy kompresji obrazów telewizyjnych: wielowidokowej: polegający na niezależnym kodowaniu poszczególnych widoków oraz wykorzystujący wyznaczone wcześniej mapy głębi.

Na świecie prowadzi się prace dotyczące pierwszego schematu, które w przyszłości mają szansę stać się częścią standardu kompresji obrazów wielowidokowych MVC [10]. W pracach tych wykorzystuje się powszechnie znane mechanizmy kompresji hybrydowej, adaptując mechanizmy skalowalności do zastosowań telewizji wielowidokowej.

Drugi schemat kompresji posiada szereg zalet: między innymi możliwość łatwej syntezy widoków pośrednich, a co za tym idzie brak konieczności transmisji obrazów odpowiadających wszystkim widokom (rys. 1.) i zmniejszenie wysyłanego strumienia danych. Schemat ten jest jeszcze we wczesnej fazie rozwoju i wymaga prowadzenia dalszych badań m.in. dotyczących technik wyznaczania map głębi, które są tematem rozważań niniejszej pracy.

3. HYBRYDOWA TECHNIKA ESTYMACJI GŁĘBI

Dla omawianych zastosowań autorzy zaproponowali oryginalny algorytm wyznaczania map głębi. Polega on na wyznaczeniu mapy rozbieżności (ang. *disparity map*), co pozwala przy znajomości parametrów optycznych kamer na wyznaczenie wartości mapy głębi dla poszczególnych punktów. Przekształcenie to ma charakter odwrotnej proporcjonalności dodatkowo związanej z parametrami optycznymi akwizycji [9].

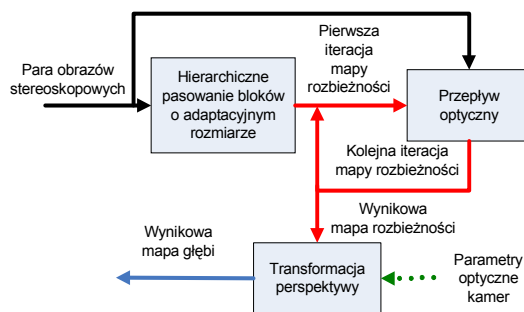
Zaproponowane rozwiązanie jest techniką hybrydową wykorzystującą hierarchiczne pasowanie bloków o zmiennym kształcie oraz gradientowy przepływ optyczny (rys. 3). Algorytm został przystosowany do pracy ze zrektyfikowanymi obrazami pochodzącymi z liniowego ułożenia kamer. W takim przypadku konieczne jest tylko wyszukiwanie rozbieżności wzdłuż osi poziomu kamer.

Pierwsza faza algorytmu polega na wyznaczeniu wstępnej mapy głębi z zastosowaniem hierarchicznego pasowania bloków (rys. 4).

Obrazy z wejściowej stereopary decymowane są wielokrotnie w celu zmniejszenia dłuższego wymiaru obrazów poniżej 32 punktów. Każdorazowa decymacja obrazu związana jest z filtracją dolnoprzepustową usuwającą z obrazów wszelkie nieistotne szczegóły.

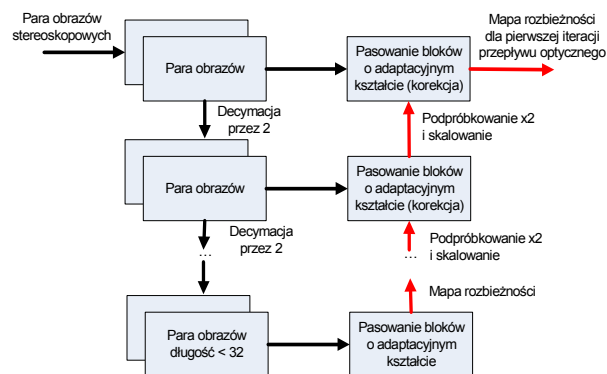
W zdecymowanych obrazach dokonuje się pasowania bloków. W pierwszym kroku wykonywane

jest pasowanie zgrubne (z wykorzystaniem obrazów o najmniejszej rozdzielczości przestrzennej otrzymanej przez decymację).



Rys. 3. Algorytm wyznaczania mapy głębi

Każdy blok pasowania reprezentuje duży obszar w obrazach oryginalnych ze zgrubną informacją o ich zawartości (efekt filtrowania). W kolejnych krokach rozdzielczość obrazu jest dwukrotnie zwiększana (wykorzystywane są obrazy mniej zdecymowane), zaś wyznaczone przez pasowanie bloków wartości rozbieżności otrzymane z poprzedniego poziomu są przeskalowywane oraz iteracyjne poprawiane w celu zwiększenia ich dokładności na podstawie informacji zawartej na aktualnie przetwarzanym poziomie.



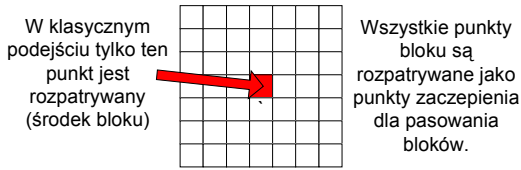
Rys. 4. Hierarchiczne pasowanie bloków

Hierarchiczne pasowanie prowadzi do zmniejszenia złożoności obliczeniowej algorytmu (w porównaniu do pełnego pasowania bloków) oraz pozwala na analizę cech rozproszonych na większej powierzchni obrazu (zwykle niedostępnej dla zwykłego pasowania bloków ze względu na koszty obliczeniowe).

Udoskonalony został także sam proces pasowania bloków przez zastosowanie adaptacyjnego doboru kształtu bloku. Zazwyczaj, rozpatrywane jest symetryczne otoczenie badanego punktu w bloku o danych wymiarach gdzie badany punkt jest umieszczony w środku tego bloku.

W proponowanym algorytmie badane jest każde możliwe najbliższe otoczenie badanego punktu, tak, że może on odpowiadać każdemu położeniu w rozpatrywanym bloku (rys. 5). Prowadzi to do lepszego działania algorytmu w pobliżu krawędzi przestrzennych obiektów występujących w obrazie

(krawędzi na wynikowej mapie rozbieżności a nie krawędzi w obrazie).



Rys. 5. Wybór punktu odniesienia dla pasowania bloków w przykładowym bloku 7×7

W ramach adaptacji kształtu bloków zmienia się również rozmiar bloku (otoczenia badanego punktu) wykorzystywany przy ich pasowaniu. Pasowanie bloków sprawdza rozmiary bloku od 9×9 punktów do rozmiaru 3×3 punktów. Wybierany jest rozmiar dający najmniejszą wartość znormalizowanego kryterium NSAD (*Normalized Sum of Absolute Differences*).

Wynikiem działania hierarchicznego pasowania bloków jest wstępna mapa rozbieżności stanowiąca dane wejściowe dla iteracyjnego przepływu optycznego z kompensacją rozbieżności.

W każdej i-tej iteracji mapa rozbieżności z poprzedniego kroku $d_{i-1}(x, y)$ jest najpierw filtrowana w celu uzyskania wymaganej gładkości oraz dla propagacji informacji o rozbieżności na obszary pozbawione tekstury. Wykorzystano następującą maskę filtru dolnoprzepustowego:

$$\frac{1}{12} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (1)$$

Następnie rozbieżność jest kompensowana przez przesunięcie punktów lewego obrazu $L(x', y')$ do odpowiadających im punktów w prawym obrazie $L(x, y)$ (2).

$$\begin{aligned} x' &= x + d_{i-1}(x, y), \\ y' &= y. \end{aligned} \quad (2)$$

Dla każdego punktu obrazu wyznaczone są: gradient poziomy (3) oraz gradient międzyobrazowy (4):

$$g_x(x, y) = \frac{1}{4\Delta} [R(x + \Delta, y) - R(x - \Delta, y) + L(x' + \Delta, y') - L(x' - \Delta, y')], \quad (3)$$

$$g_t(x, y) = L(x', y') - R(x, y), \quad (4)$$

gdzie Δ jest krokiem aproksymacji gradientu (w eksperymencie $\Delta = 0,01$).

W każdej iteracji, na podstawie znanych już gradientów (3,4), wyznaczana jest poprawka (5) i dodawana do aktualnej wartości rozbieżności. Równanie (5) zawiera dwa współczynniki: α - współczynniki niepewność związanej z szumem (w eksperymencie $\alpha = 25$), β - krok iteracji (w eksperymencie $\beta = 0,5$).

$$d_i(x, y) \leftarrow d_{i-1}(x, y) - \beta \cdot g_t(x, y) \cdot \frac{g_x(x, y)}{g_x(x, y)^2 + \alpha} \quad (5)$$

Warto zauważyć, iż gdy $\alpha = 0$, równanie (5) sprowadza się do równania (6), które reprezentuje iteracyjną metodę rozwiązania równania w oparciu o gradienty.

$$d_i(y, x) \leftarrow d_{i-1}(y, x) - \beta \cdot g_t(y, x) \cdot \frac{1}{g_x(y, x)} \quad (6)$$

Niezerowe α zostało wprowadzone w celu zredukowania wpływu szumu na obszarach o małym gradiencie (pozbawionym tekstury).

Na koniec każdej iteracji sprawdzane jest czy rozbieżność nie jest ujemna (co wskazywało by na zamianę obrazów wejściowych) i ewentualne dokonywana jest poprawka niwelująca ten efekt, zgodnie ze wzorem (7)

$$d_i(x, y) \leftarrow \frac{1}{2} d_i(x, y) \quad (7)$$

4. WYNIKI

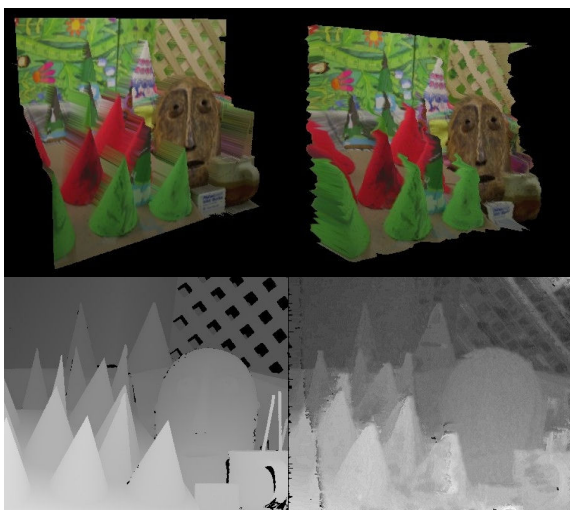
Na rysunkach 6, 7 i 8 przedstawiono porównanie map rozbieżności (wyznaczonych za pomocą algorytmu przedstawionego w punkcie 3) z mapą rozbieżności Ground Truth oraz pokazano resyntezywane na ich podstawie widoki. Wykorzystane obrazy pochodzą z bazy danych Middlebury Stereo Datasets [11].

Przeprowadzono również testy z wykorzystaniem powszechnie stosowanych wielowidokowych sekwencji wizyjnych. Pojedynczą ramkę pochodzącą z jednej z nich przedstawiono na rys. 9.

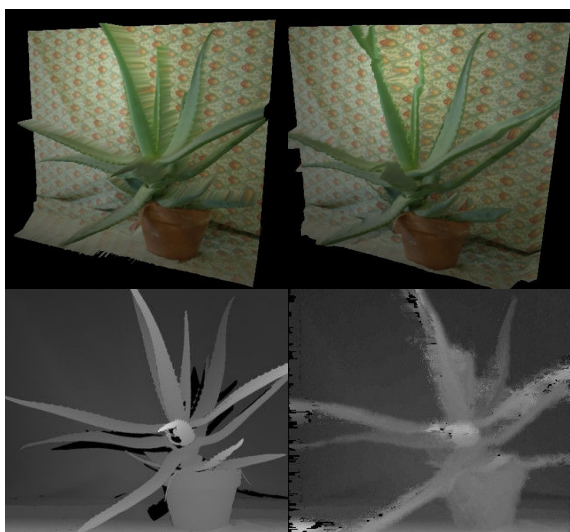
Ponieważ wymogiem stawianym przed poszukiwaną mapą głębi jest uzyskanie możliwie dobrej rekonstrukcji obrazu, dla każdego obrazu testowego dokonywano resyntezy jednego z widoków - z wykorzystaniem mapy rozbieżności uzyskanej za pomocą zaproponowanej techniki oraz drugiego - z wykorzystaniem mapy rozbieżności Ground Truth. Uzyskane w ten sposób widoki posłużyły do porównań z widokiem oryginalnym.



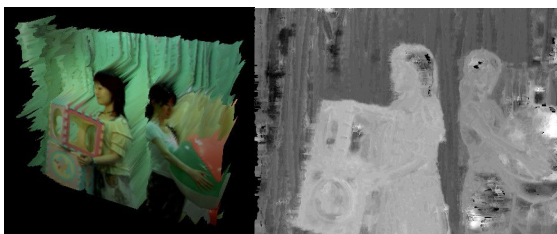
Rys. 6. Porównanie otrzymanej mapy rozbieżności (po prawej) z Ground Truth (po lewej) dla obrazu Tsukuba



Rys. 7. Porównanie otrzymanej mapy rozbieżności (po prawej) z Ground Truth (po lewej) dla obrazka Cones



Rys. 8. Porównanie otrzymanej mapy rozbieżności (po prawej) z Ground Truth (po lewej) dla obrazka Aloe



Rys. 9. Otrzymana mapa rozbieżności dla pojedynczej klatki z sekwencji Akko&Kayo

Przedstawione obrazy widoków są subiektywnie nierozróżnialne względem uzyskanych za pomocą mapy rozbieżności Ground Truth. Świadczy to o małym wpływie błędów popełnianych przez proponowaną estymację map rozbieżności na proces resyntezy.

Wyniki testów obiektywnych w postaci wskaźnika PSNR zebrano w tabeli 1.

Tab. 1. Wyniki porównania resyntezy widoków

Para obrazów	Ground Truth PSNR [dB]	Proponowany algorytm PSNR [dB]
Tsukuba	32,867	27,623
Cones	27,205	26,284
Aloe	33,975	32,149

5. PODSUMOWANIE

Na tle zastosowań telewizji swobodnego punktu widzenia przedstawiono oryginalną koncepcję techniki estymacji głębi. Technika ta została wdrożona i przetestowana przez autorów, a także jest wykorzystywana w ramach prac grupy MPEG dotyczących FTV [8]. Wyniki wskazują, że jakość obrazów resyntezyowanych na podstawie estymowanej mapy głębi obrazów jest wystarczająca do przewidzianych zastosowań. Zaproponowany algorytm może konkurować z innymi tego typu technikami zarówno pod względem jakości jak i prędkości działania. Dalsze prace będą się skupiać na realizacji pozostałych bloków rozważanego systemu.

SPIS LITERATURY

- [1] O. Veksler, "Stereo correspondence by dynamic programming on a tree", CVPR 2005
- [2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", IJCV 2002.
- [3] J. Sun, H. Y. Shum, and N. N. Zheng, "Stereo matching using belief propagation", ECCV 2002.
- [4] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts", ECCV 2002.
- [5] "Call for Contributions on FTV Test Material", MPEG 2007/N9468, Shenzhen, China, October 2007
- [6] I. Lee, Y. Lung, B. Kim and S. KO. "An Advanced Video "Camera System with Robust AF, AE, and AWB Control", IEEE Trans on Consumer Electronics. vol. 47, no. 3. pp. 694-699, Aug. 2001.
- [7] T. Sato,; N. Yokoya,; "Multi-baseline stereo by maximizing total number of interest points" SICE, 2007. Annual Conference, 17-20 Sept. 2007 Pages:1471 – 1477
- [8] "Depth Map Estimation Software", MPEG 2008/M1517, Antalya, Turkey, January 2008
- [9] "Description of Exploration Experiments in 3D Video", MPEG 2008/N9596, Antalya, Turkey, January 2008 – wzór (1)
- [10] K. Mueller P. Merkle A.Smolc T.Wiegand "Multiview Coding using AVC, ISO/IEC JTC1/SC29/WG11 documnet M12945 Bangkok, Thailand, January 2006
- [11] "Middlebury Stereo Datasets" <http://vision.middlebury.edu/stereo/data/>