XXIII Seminarium

ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE' 2013
Oddział Gdański PTETiS
Referat nr 6

ŚLEDZENIE GŁOWY UŻYTKOWNIKA KOMPUTERA Z UŻYCIEM KAMERY TIME OF FLIGHT

Piotr BRATOSZEWSKI, Andrzej CZYŻEWSKI

Politechnika Gdańska, ul. Gabriela Narutowicza 11/12,
80-952 Gdańsk, Wydział Elektroniki, Telekomunikacji i Informatyki, Katedra Systemów Multimedialnych
tel: (58) 347-63-32 e-mail: {bratoszewski, andcz}@sound.eti.pg.gda.pl

Streszczenie: Opisano opracowaną metodę śledzenia położenia głowy użytkownika komputera lub urządzenia mobilnego przy wykorzystaniu kamery mierzącej czas powrotu wiązki promieniowania elektromagnetycznego podczerwonego odbitej od oświetlanego obiektu (ang. *Time Of Flight camera*). Dzięki zastosowaniu odpowiednich metod cyfrowego przetwarzania obrazu pozyskanego z kamery tego typu możliwe jest zlokalizowanie użytkownika w przestrzeni 3D. Znajomość dokładnej lokalizacji głowy może posłużyć tworzeniu nowych interfejsów komunikacji między człowiekiem a maszyną lub tworzeniu aplikacji komputerowych nowego typu.

Słowa kluczowe: śledzenie głowy, obrazowanie 3D, analiza obrazu

1. WPROWADZENIE

Metody śledzenia położenia głowy osoby znajdującej się przed kamerą, zgodnie z aktualną literaturą, oparte są w głównej mierze o lokalizację śledzonej twarzy. Opracowano wiele metod śledzenia twarzy, jednak najbardziej znanymi pozostały metody oparte o obraz pochodzący z kamer RGB. Istnieją metody związane z barwą twarzy [1] traktowaną jako cechę dystynktywną, odróżniającą twarz od koloru tła, czy ubioru użytkownika. Inna, popularna metoda, związana z wartością luminancji pewnych grup pikseli i ich współwystępowaniem w typowej twarzy, została przedstawiona przez Viola i Jones [2]. W tych podejściach skuteczność detekcji oraz śledzenia twarzy w silnym stopniu zależą od ustawienia twarzy względem kamery, a także od oświetlenia danej sceny. Znane są również metody związane z analizą kształtu sylwetki z wykorzystaniem deskryptorów HOG (ang. *Histogram of Oriented Gradients*) [3], metoda ta znajduje swoje zastosowanie przede wszystkim w rozwiązaniach dotyczących inteligentnych systemów monitorowania bezpieczeństwa. Podejście wykorzystujące inną technikę wizyjną niż kamery RGB zaproponowali Shotton, Fitzgibbon i inni, tworząc system rozpoznawania i śledzenia położenia stawów w sylwetkach stojących osób przy użyciu kamery Kinect [4]. Zaprezentowana w niniejszym referacie metoda śledzenia pozycji głowy oparta jest o kamerę mierzącą czas przelotu zmodulowanego promieniowania podczerwonego (ang. *Time Of Flight* (TOF)) [5]. Kamera ta należy do grupy sensorów typu LIDAR (ang. *Light Detection and Ranging*). W polskiej nomenklaturze sensory TOF są również nazywane skanerami pulsacyjnymi. Sensory te należą do grupy tzw. aktywnych skanerów 3D. Poprzez oświetlanie obiektów zmodulowanym promieniowaniem elektromagnetycznym z zakresu bliskiej podczerwieni oraz rejestrację promieniowania odbitego od obiektu możliwe jest określenie odległości oświetlanego obiektu od kamery. Kamery tego typu udostępniają dwa strumienie danych niosące informacje o intensywności odbitego promieniowania podczerwonego od obiektu oraz informacje o odległości obiektu od kamery.

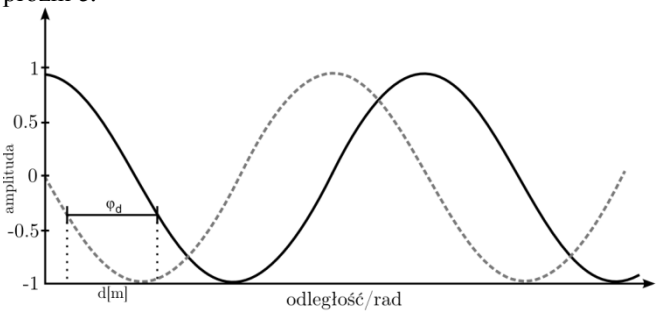
W przedstawionym podejściu wykorzystywany jest strumień danych z kamery TOF obrazujący odległość użytkownika od kamery, tzw. mapa głębi. Następnie zastosowane są liczne przekształcenia obrazu w celu stworzenia systemu śledzącego położenie głowy użytkownika w przestrzeni 3D. Przyjmuje się, że użytkownik ustawiony jest w pozycji siedzącej przed komputerem oraz że jego głowa nie jest przysłaniana przez żadne obiekty.

Referat składa się z następujących części: rozdział 2. zawiera opis wykorzystanego sensora typu TOF, w rozdziale 3. opisana jest architektura opracowanego systemu oraz wykorzystane w nim metody przetwarzania obrazu, w rozdziale 4. zaprezentowane są wyniki eksperymentów. Podsumowanie i wnioski znajdują się w rozdziale 5.

2. KAMERY TIME OF FLIGHT

Istnieje wiele technik obrazowania głębi (ang. *Range Imaging*) danej sceny. Do najbardziej znanych zalicza się techniki wykorzystujące triangulacje obrazu z dwóch lub więcej kamer [6], techniki wykorzystujące strukturalne promieniowanie elektromagnetyczne o fali długości od ok. 380 nm do ok. 760 nm, wykrywalne przez oko ludzkie (promieniowanie optyczne) [7] lub promieniowanie o fali długości od 780 nm do 1 mm [8]. W ramach niniejszej pracy została wykorzystana relatywnie nowa technika obrazowania oparta o kamerę TOF. Wykorzystuje ona zmodulowane promieniowanie elektromagnetyczne o fali długości od 800 nm do 2,5 μm (ang. *near infrared*), do oświetlania obiektu oraz rejestrację promieniowania odbitego. Promieniowanie podczerwone jest modulowane sygnałem o przebiegu kosinusoidalnym, o częstotliwości z zakresu 20-30 MHz. Dzięki modulacji możliwe jest określenie różnic w fazie sygnału nadawanego i odbieranego. Na rys. 1 przedstawione

są zmodulowane sygnały emitowany oraz rejestrowany przez kamerę. Zgodnie ze wzorem 1, możliwe jest określanie odległości *d* obiektu od kamery na podstawie różnicy fazy ϕ_d obu tych sygnałów, znając częstotliwość modulacji f_{mod} oraz prędkość rozchodzenia się fali elektromagnetycznej w próżni *c*.



Rys. 1. Przesunięcie fazowe pomiędzy zmodulowanym promieniowaniem podczerwonym emitowanym oraz odebranym

$$d = \frac{c}{f_{\mathrm{mod}}} \cdot \frac{1}{2} \cdot \frac{\phi_d}{2\pi}, \tag{1}$$

gdzie: *d* – odległość od obiektu, *c* – prędkość światła, f_{mod} – częstotliwość modulacji, ϕ_d – różnica fazy.

Do eksperymentów opisanych w niniejszym referacie wykorzystano kamerę SoftKinetic DS325 [9], wyposażoną w matrycę do pomiaru głębi o rozdzielczości 320x240 pikseli. Dokładność pomiaru głębi dla obiektu odległego od obiektywu o 1 m wynosi mniej niż 1,4 cm. Zasięg roboczy kamery wynosi od 0,15 m do ok. 1,5 m.

3. OPIS SYSTEMU ŚLEDZENIA GŁOWY

Stworzony moduł śledzenia głowy wykorzystuje obraz mapy głębi otrzymywany za pomocą kamery TOF oraz szereg następujących po sobie przekształceń cyfrowych, zaprezentowanych na schemacie funkcjonalnym systemu, przedstawionym na rys. 2. Aby możliwe było śledzenie głowy należy wyekstrahować z obrazu, pozyskanego z kamery, binarną sylwetkę użytkownika. W tym celu zaproponowano algorytm składający się z kilku etapów. Pierwszym etapem jest filtracja wstępna składająca się z filtracji medianowej oraz filtracji bilateralnej, dzięki której możliwe jest zachowanie niezmienionych krawędzi w sylwetce użytkownika. Następnym blokiem jest progowanie intensywności pikseli w obrazie. Intensywność pikseli *I*(*x*,*y*) w obrazie mapy głębi jest proporcjonalna do odległości obiektów przed kamerą. Dzięki odpowiedniemu procesowi progowania, zgodnie ze wzorem 2 możliwe jest odrzucenie obiektów o większej odległości od kamery niż zadana wartość *d*.

$$I(x,y) = \begin{cases} 0 & d > 1{,}5\ m \\ I(x,y) & 0 < d \le 1{,}5\ m \end{cases}, \tag{2}$$

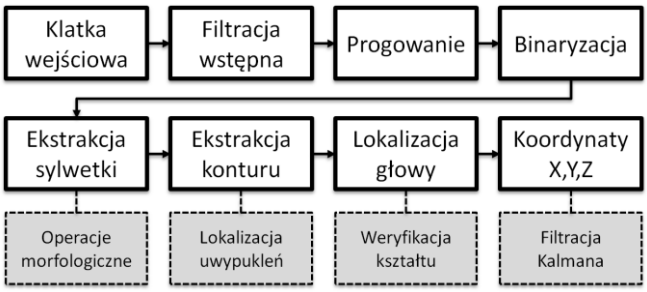
gdzie: *d* – odległość od obiektu, *I*(*x*,*y*) – intensywność pikseli obrazu.

Kolejnym etapem jest binaryzacja sprogowanego obrazu zgodnie ze wzorem 3. Dla danej wartości progu T intensywność obrazu *I*(*x*,*y*) jest transformowana do *f*(*x*,*y*) zgodnie z następującą zależnością:

$$f(x,y) = \begin{cases} 0 & I(x,y) \le T \\ 1 & I(x,y) > T \end{cases}, \tag{3}$$

gdzie: *T* – wartość progu intensywności pikseli, *I*(*x*,*y*) – intensywność pikseli obrazu, *f*(*x*,*y*) – wartość binarnej maski.

Na binarnym obrazie, przedstawionym na rys. 4c, wykonywane są operacje morfologiczne – dylacja oraz erozja, celem uzyskania jednolitej, pozbawionej ubytków binarnej maski sylwetki użytkownika. Na tak przygotowanym obrazie znajdowane są tzw. komponenty połączone (ang. *connected components*). Dzięki wyznaczeniu połączonych komponentów możliwe jest odfiltrowanie obszarów o zbyt małym lub zbyt dużym polu powierzchni. W wyniku przytoczonych przekształceń otrzymano binarną maskę obrazującą sylwetkę użytkownika (rys. 4d), która następnie poddawana jest analizie kształtu. W otrzymanej masce sylwetki znajdowane następuje ekstrakcja konturu [10] oraz znajdowane są uwypuklenia (ang. *convex hull*) [11]. Lokalizacja danego uwypuklenia jest traktowana jako kandydat dla lokalizacji wierzchołka głowy użytkownika. Rys 5. obrazuje iteracyjne sprawdzanie kolejnych kandydatów (uwypukleń sylwetki), podążając od najwyżej do najniżej położonego. Sprawdzane jest, czy poniżej rozpatrywanego kandydata znajduje się obszar, który można traktować jako głowę i ramiona użytkownika. Obszar ten znajdowany jest na podstawie stosunku białych pikseli wewnątrz wzorca głowy i ramion do powierzchni wzorca. Jeśli dane prostokąty we wzorcu posiadają wypełnienie sylwetką użytkownika, a nie tłem, większe od zadanego progu, przyjmuje się, że skonstruowane zostały w lokalizacji uwypuklenia będącego wierzchołkiem głowy użytkownika. W ten sposób określane są koordynaty X oraz Y głowy użytkownika wyrażone w pikselach. Wyznaczone koordynaty X, Y zmieniają się wraz z poruszaniem się użytkownika. Podczas śledzenia mogą występować sytuacje, w których użytkownik w specyficzny sposób ustawia się do kamery lub występują pewne zniekształcenia wyekstrahowanej sylwetki, co może powodować, że wyznaczone uwypuklenie nie jest tożsame czubkiem głowy użytkownika. Dlatego, w celu wygładzenia danych ze śledzenia punktów X, Y, tak, aby nie zmieniały się o zbyt dużą wartość pomiędzy sąsiednimi ramkami, zastosowany jest predykcyjny filtr Kalmana [12].



Rys. 2. Schemat funkcjonalny zaproponowanego systemu śledzenia głowy

W oparciu o stabilne wartości położenia głowy, wyznaczonej z obrazu binarnego, możliwe jest określenie koordynaty Z, czyli odległości głowy użytkownika od kamery korzystając z mapy głębi. W tym celu, wybierany jest obszar poniżej znalezionego uwypuklenia tak, aby obliczana wartość Z znajdowała się na obszarze twarzy. W tym obszarze, o rozmiarze 12x12 pikseli, zliczane są i uśredniane wartości pikseli obrazu głębi. Uśredniona wartość jest wprost proporcjonalna do odległości głowy

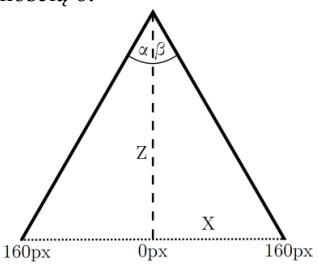
użytkownika od kamery i jest wyskalowana w metrach. Znając wartości X, Y w pikselach i Z w metrach oraz parametry obiektywu kamery, możliwe jest określenie wszystkich koordynat głowy w przestrzeni 3D w metrach. W tym celu wykorzystywana jest zależność trygonometryczna przedstawiona na rys. 3 oraz wykorzystywane są wzory 4, 5 i 6:

$$FOV = \alpha + \beta \tag{4}$$

$$tg\alpha = \frac{X}{Z} \tag{5}$$

$$X = tg\alpha \cdot Z\,, \tag{6}$$

gdzie FOV oznacza szerokość kątową obiektywu w płaszczyźnie horyzontalnej i dla badanej kamery wynosi 74 stopnie. Moduł śledzenia zwraca wartość X i Y w pikselach. Znając parametry matrycy, możliwe jest obliczenie kąta α lub β, o jaki jest odchylona głowa użytkownika od osi kamery. Wychylenie się o 1 piksel jest równe wychyleniu się o 0,23 stopnia. Korzystając z tych danych, możliwe jest wyliczenie wychylenia głowy użytkownika w centymetrach, zgodnie z zależnością 6.

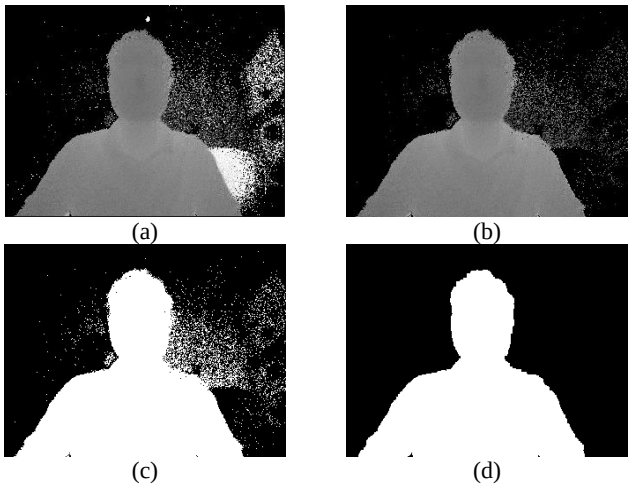


Rys. 3. Zależności pozwalające na znalezienie położenia głowy użytkownikach w jednostkach ze świata rzeczywistego

4. WYNIKI EKSPERYMENTÓW

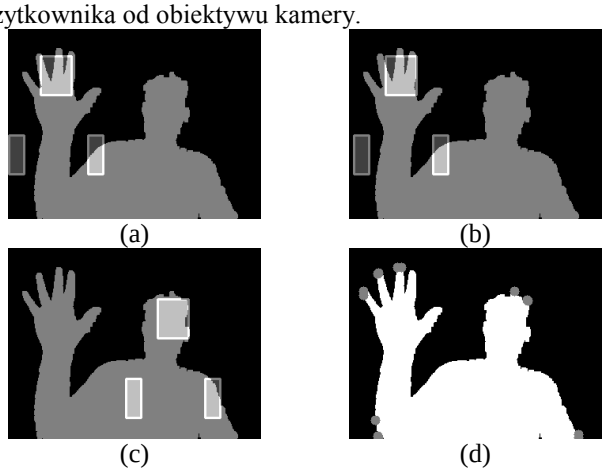
W poniższym rozdziale przedstawione są wyniki przeprowadzonych eksperymentów. Zakłada się, że cała sylwetka użytkownika znajduje się w obszarze roboczym kamery. W przypadku zastosowanej kamery, użytkownik musi znajdować się w odległości od ok. 40 cm do 90 cm od obiektywu, aby zaproponowana metoda działała skutecznie. Nie stosuje się restrykcji dotyczących tła w analizowanej scenie. Tło może być ruchome oraz mogą występować w nim różne obiekty. Zaproponowane rozwiązanie w toku przetwarzania odrzuca obiekty oddalone od kamery dalej niż użytkownik. Rys. 4. przedstawia przebieg obróbki sygnału wizyjnego od momentu jego akwizycji, aż do uzyskania jednolitej sylwetki użytkownika. Rys. 4a przedstawia ramkę obrazu pozyskaną z kamery TOF, gdzie intensywność pikseli reprezentuje odległość obiektów od kamery. Im obiekt znajduje się dalej od obiektywu kamery TOF, tym piksele, z których się składa są jaśniejsze. Na rys. 4b następuje odrzucenie obiektów oddalonych o określoną odległość od kamery, następnie na rys. 4c przedstawiono wynik binaryzacji obrazu. Rys. 4d przedstawia ostateczny wynik przetwarzania obrazu, jakim jest wyznaczenie binarnej sylwetki użytkownika. Powstaje ona z obrazu binarnego po zastosowaniu operacji morfologicznych i wybraniu komponentów połączonych o zadanym polu powierzchni.

W oparciu o zarejestrowaną binarną sylwetkę użytkownika możliwe jest rozpoczęcie śledzenia położenia głowy użytkownika. Rys. 6. przedstawia wyniki znajdowania obszaru głowy w obrazie binarnej sylwetki oraz obliczanie odległości użytkownika od kamery z wykorzystaniem mapy głębi dla dwóch pozycji użytkownika. Rys. 6a oraz 6b przedstawiają wynik poszukiwania uwypukleń (ang. *convexity hulls*) w sylwetce użytkownika. W celu odróżnienia głowy użytkownika od uniesionej ręki lub przedmiotu znajdującego się w obrębie sceny stosuje się prostokątne cechy przedstawione na rys. 6c oraz 6d i opisane w rozdz. 3.



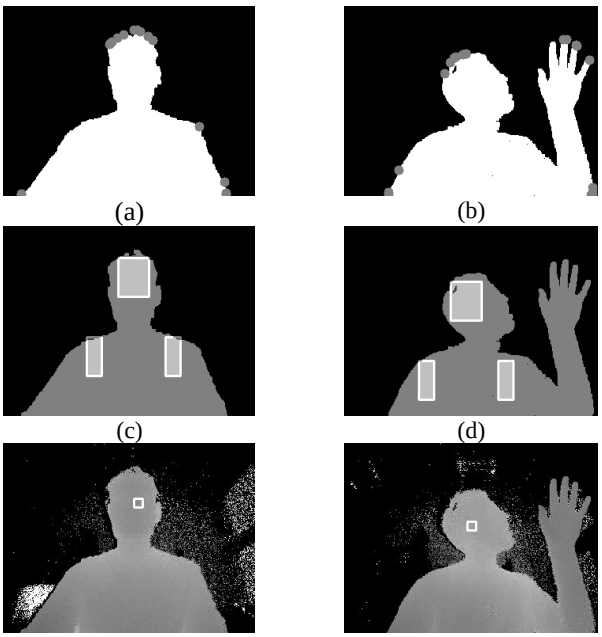
Rys. 4. Kolejne etapy przetwarzania ramki z kamery: a) mapa głębi sceny, b) obraz przycięty do odległości w jakiej znajduje się użytkownik, c) binaryzacja obrazu, d) wyekstrahowana sylwetka

Po jednoznacznym określeniu obszaru, w którym znajduje się głowa definiowany jest obszar w mapie głębi – widoczny biały prostokąt o rozmiarze 12x12 pikseli na rys. 6e i 6f, wewnątrz którego znajdują się piksele określające odległość głowy użytkownika od kamery. Wartości pikseli znajdujące się w prostokącie są sumowane, a następnie dzielone przez znany rozmiar prostokąta, dzięki czemu otrzymuje się stabilną, uśrednioną odległość głowy użytkownika od obiektywu kamery.



Rys. 5. Kolejne iteracje poszukiwania głowy użytkownika: a), b) najwyżej położone uwypuklenia sylwetki, które nie spełniają kryteriów lokalizacji oznaczającej głowę użytkownika, c) lokalizacja spełniająca kryteria głowy, d) wszystkie uwypuklenia sylwetki

Przeprowadzone eksperymenty dowodzą, że opisana metoda pozwala na śledzenie położenia głowy użytkownika w czasie rzeczywistym, przy zachowaniu około 35 ramek na sekundę dla czterordzeniowego procesora taktowanego zegarem 3,4 GHz. Metoda śledzenia jest nieczuła na obrót twarzy użytkownika od osi kamery oraz na zmiany oświetlenia.



Rys. 6. Proces lokalizacji głowy użytkownika dla dwóch póz – na wprost (lewa), z rotacją głowy (prawa): a), b) lokalizacja wszystkich uwypukleń w sylwetce, c), d) weryfikacja, czy uwypuklenie jest wierzchołkiem głowy, e), f) prostokąt, wewnątrz którego obliczana jest odległość

5. PODSUMOWANIE

Przedstawiono opracowaną metodę lokalizacji głowy użytkownika w przestrzeni przed komputerem, opartą na nowoczesnej technice obrazowania za pomocą kamery TOF. Korzystając z szeregu przekształceń obrazu mapy głębi sceny możliwe jest usunięcie tła i wyekstrahowanie jednolitej sylwetki użytkownika, a następnie lokalizacja głowy. Dzięki wykorzystaniu techniki bazującej na promieniowaniu elektromagnetycznemu z zakresu bliskiej podczerwieni stworzono detektor pozwalający na śledzenie głowy z tą samą skutecznością przy różnych warunkach oświetleniowych. Przeprowadzone eksperymenty potwierdzają również umiarkowany koszt obliczeń komputerowych zastosowanej metody oraz płynną pracę detektora przy przetwarzaniu średnio 35 ramek na sekundę.

Zaproponowana metoda może znaleźć zastosowanie w interfejsach człowiek-maszyna, gdy na podstawie ustawienia głowy względem monitora możliwe jest przedstawianie użytkownikowi innej perspektywy obrazu (rzeczywistość rozszerzona, grafika komputerowa) lub w rozwiązaniach wymagających adaptacji do aktualnego położenia głowy użytkownika, na przykład w systemach projekcji dźwięku.

6. BIBLIOGRAFIA

1. Mohamed A., Weng Y., Ipson S., Jiang J.: Face Detection based on Skin Color in Image by Neural Networks, International ICIAS Conference on Intelligent and Advanced Systems, November 2007, Kuala Lumpur, s. 779-783, ISBN 978-1-4244-1355-3.
2. Viola P., Jones M.: Rapid Object Detection using a Boosted Cascade of Simple Features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision And Pattern Recognition, 2001, s. 511-518, vol.1, ISBN 0-7695-1272-0.
3. Li M., Zhang Z., Huang K., Tan T.: Rapid and Robust Human Detection and Tracking Based on Omega-Shape Features, 16th International IEEE Conference on Image Processing, November 2009, Cairo, s. 2545-2548, ISBN 978-1-4244-5653-6.
4. Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kimpan A., Blake A.: Real-Time Human Pose Recognition in Parts from Single Depth Images, International IEEE Conference on Computer Vision and Pattern Recognition, June 2011, Providence, s. 1297-1304, ISBN 978-1-4577-0394-2.
5. Hansard M., Lee S., Choi O., Horaud R.: Time-of-Flight Cameras: Principles, Methods and Application, Springer 2012, ISBN 978-1447146575.
6. Kanade, T., Yoshida A., Oda K., Kano H., Tanaka M.: A Video-Rate Stereo Machine and Its New Applications, IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, June 1996, s.196-202, ISBN 0-8186-7259-5.
7. Caspi D., Kiryati N., Shamir J.: Range Imaging With Adaptive Color Structured Light, IEEE Transactions on Pattern Analysis and Machine Intelligence, May 1998, s. 470-480, vol. 20, nr 5.
8. Wang J., Zhang Ch., Zhu W., Zhang Z., Xiong Z., Chou P.: 3D Scene Reconstruction by Multiple Structured-Light Based Commodity Depth Cameras, International IEEE Conference on Acoustics, Speech and Signal Processing, March 2012, Kyoto, s. 5429-5432, ISBN 978-1-4673-0045-2.
9. SoftKinetic DS325, specyfikacja kamery, dostęp: http://www.softkinetic.com, [dostęp: 18.11.2013].
10. Suzuki S., Abe K.: Topological Structural Analysis of Digitized Binary Images by Border Following, CVGIP, 1985, s. 32-46, vol.30, nr. 1
11. Sklansky, J.: Finding the Convex Hull of a Simple Polygon, PRL, 1982, s. 79-83, vol. 1982
12. Szwoch G., Dalka P., Czyżewski A.: Resolving conflicts in object tracking for automatic detection of events in video, Elektronika: konstrukcje, technologie, zastosowania, s. 52-54, vol. 52, nr 1, ISSN 0033-2089.

HEAD TRACKING USING THE TIME OF FLIGHT CAMERA

Key-words: head tracking, depth image, image processing

A depth image based real-time head tracking system is described. The proposed system utilizes the Time of Flight camera and digital image processing techniques in order to track user's head position in the real world coordinates. The detection of head position is based on the shape features of the silhouette. Tracking of head location is enhanced and smoothed by the usage of the Kalman filtering. The developed application runs in real-time and is resistant to different lighting conditions.