

SZPYTKO Janusz ¹
HYLA Paweł ²
GBYL Michał ³

STEROWANIE DŹWIGNICĄ PRZY POMOCY GESTÓW Z UŻYCIEM HYBRYDOWEGO URZĄDZENIA WEJŚCIA Z TRÓJWYMIAROWYM UKŁADEM ODNIESIENIA

W artykule przedstawiono adaptację urządzenia wejścia z trójwymiarowym układem odniesienia – kontrolera znanego pod nazwą kinect dla potrzeb sterowania laboratoryjną suwnicą pomostową. Przedmiotem pracy jest realizacja koncepcji wkomponowania w układ C-OT-O nowego typu interfejsu komunikacyjnego opartego o niewerbalny bazujący na gestach układ przekazywania informacji do środowiska cyfrowego.

GESTURES HYBRID CRANE CONTROL SYSTEM WITH USE THE NEW THREE DIMENSIONAL INPUT DEVICE

The paper presents the conception of adaptation a new three-dimensional input device under the kinect name. Authors focused their attention on realization the concept of design the new kind of the HMI interface. In this type of interface, communication was based on the non-verbal gestures system. As a reference transport devices was use a laboratory double girder crane.

1. WSTĘP

Historia komputerowych urządzeń wejścia specjalnego przeznaczenia [1, 4] jest ściśle związana z rozwojem segmentu gier komputerowych w których integracja użytkownika z wirtualną rzeczywistością wymaga nowych rozwiązań sprzętowych, a zwłaszcza kontrolerów i interfejsów dla potrzeb integracji użytkownika z aplikacją [11].

¹Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Wydział Inżynierii Mechanicznej i Robotyki; 30-059 Kraków; al. A. Mickiewicza 30,

Tel: +48 12 617 3103, Fax: +48 12 617 3531, e-mail: szpytko@agh.edu.pl

²Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Wydział Inżynierii Mechanicznej i Robotyki; 30-059 Kraków; al. A. Mickiewicza 30,

Tel: +48 12 617 3104, e-mail: hyla@agh.edu.pl

³Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Wydział Inżynierii Mechanicznej i Robotyki; 30-059 Kraków; al. A. Mickiewicza 30,

Tel: +48 12 617 3104, e-mail: mgbyl@agh.edu.pl

Standard dzisiejszej multimedialnej rozrywki dalece wykracza poza programy w których sterowanie przy pomocy podstawowych urządzeń wejścia takich jak klawiatura oraz urządzenie wskazujące (myszy) jest nie tyle możliwe co wysoce nieefektywne [1]. Współcześnie poprzez maksymalnie możliwy do osiągnięcia techniczny poziom odwzorowania rzeczywistości wciąż pojawiają się nowe, nietypowe urządzenia wejścia przeznaczone do kontrolowania gier lub multimedialnych programów edukacyjnych. Można wyróżnić kontrolery w postaci kierownic zintegrowanych z pedałami oraz dżankami zmiany biegów, przeznaczone głównie do gier o charakterze wyścigowym, deski rozdzielcze odzwierciedlające pulpity sterowania pociągami (n.p. Microsoft Train Simulator) lub specjalne maty umożliwiające szerokokorozumiany wirtualny fitness. Mimo wysokiego poziomu realizmu, wymienione kontrolery znalazły do tej pory zastosowanie tylko i wyłącznie w branży gier komputerowych lub nie wykraczają poza obszar innych interaktywnych programów. Duży potencjał oraz szersze zaistnienie na rynku urządzeń wejścia, dotyczy nowego multimedialnego kontrolera firmy Microsoft. Microsoft Kinet [3, 5] pracuje z wykorzystaniem trójwymiarowego układu odniesienia [3] i ma szansę uzyskać taką samą funkcjonalność i popularność jak współczesne urządzenia wejścia i obok klawiatury oraz myszy komputerowej stać się standardowym urządzeniem służącym do komunikacji z komputerem.

W artykule przedstawiono adaptację urządzenia kinect dla potrzeb sterowania laboratoryjną suwnicą pomostową. Autorzy skoncentrowali się na realizacji koncepcji wkomponowania w układ człowiek – obiekt techniczny – otoczenie (C-OT-O) interfejsu telematycznego [13], opartego na gestach i mowie ciała, jako odzwierciedlenie naturalnego dla człowieka systemu przekazywania informacji.

2. URZĄDZENIE WEJŚCIA SPECJALNEGO PRZEZNACZENIA Z TRÓJWYMIAROWYM UKŁADEM ODNIESIENIA

Urządzenie Kinect pierwszy raz zaprezentowano dnia 1 czerwca 2009 roku na międzynarodowych targach E3. Projekt nosił nazwę kodową Natal co z języka angielskiego można tłumaczyć jako *noworodek* lub *nowonarodzony*. Przymiotnik odzwierciedlał ideę twórców, polegającą na narodzinach technologii kolejnej generacji kontrolerów gier, a tym samym nowego typu urządzenia wejścia. Rok później, podczas kolejnej edycji tych samych odbywających się w Kalifornii, Microsoft przedstawił oficjalnie nową nazwę *Projektu Natal* nadając jej nazwę *Kinect*. Etymologicznie słowo kinect stanowi połączenie angielskich wyrazów *kinetic* (kinetyczny) i *connect* (łączyć). W początkowej fazie dystrybucji Kinect (rys. 1) miał być tylko i wyłącznie sensorem ruchu wykorzystywanym przez konsolę do gier XBOX 360. Przy jego pomocy interaktywna rozrywka miała zostać wzbogacona o nowy element w której jako kontroler gry wykorzystywano ruchy, gesty oraz komendy głosowe użytkownika. Przy użyciu kineta wirtualna rozrywka miała być prostsza, przyjemniejsza oraz dostarczać nowego typu wrażeń. Nowatorskie podejście uwolnienia użytkowników od potrzeby użytkowania padów i innych powszechnie do tej pory stosowanych urządzeń manipulacyjnych, stanowiło o sukcesie projektu. Szybko okazało się że technologia Kinect w której kontrolerem jest całe ciało użytkownika nie tylko pogłębia interakcję pomiędzy człowiekiem, a wirtualną rzeczywistością ale też czyni

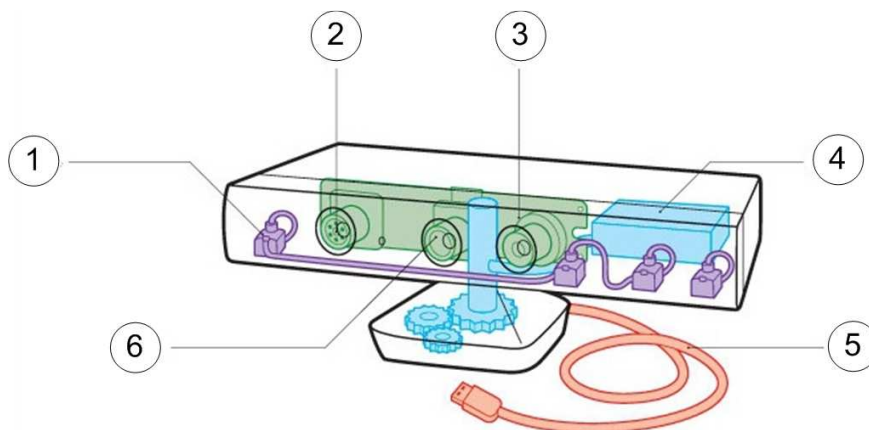
wirtualną rozrywkę łatwiejszą poprzez naturalność sterowania gestami wykonywanymi przez człowieka w życiu codziennym.



Rys. 1. Microsoft Kinect

Rosnąca popularność kinecta zaowocowała zaadoptowaniem urządzenia dla potrzeb obsługi interfejsu MS Windows. Od samego początku istnienia urządzenia powstawały mniej lub bardziej udane pakiety oprogramowania wspierające działanie urządzenia na „Oknach” [7]. Jednak dopiero w styczniu 2011 roku zostały opublikowane jedne z pierwszych oficjalnych sterowników wspierających działanie sensora ruchu w środowisku MS Windows.

Jednak oficjalna premiera sterowników Microsoftu miała miejsce dopiero pół roku później, pod koniec lipca 2011 [6, 10]. Architektura sprzętowa urządzenia jest dość skomplikowana (rys. 2).



Rys.2. Architektura sensora Kinect

W tabeli 1, przedstawiono i scharakteryzowano podstawowe układy wchodzące w skład kontrolera (rys. 2).

Tab.1. Elementy wchodzące w skład budowy urządzenia Kinect (rys.2)

Sensor/Urządzenie/Interfejs	Funkcja/ charakterystyka
Zestaw mikrofonów (1)	Zestaw 4 mikrofonów z zaadoptowaną funkcją rozpoznawania mowy oraz filtrowania zakłóceń.
Emiter podczerwieni (2)	Emituje wiązkę promieni podczerwonych, które odbijając się od powierzchni zniekształcają się i następnie są odczytywane przez kamerę głębokości.
Kamera na podczerwień o rozdzielczości 320x240 (3)	Analizuje zniekształconą przez skanowany obiekt wiązkę promieni podczerwonych.
Automatyczny kontroler nachylenia (4)	Dostosowuje położenie sensora w zależności od pozycji oraz rozmiarów śledzonych obiektów.
Kabel połączeniowy (5)	Kabel przypominający interfejsem gniazdo USB (różnica polega na ściętym narożniku). Do połączenia sensora Kinect z komputerem potrzebny jest dodatkowy zasilacz z wbudowanym kablem USB.
Kolorowa kamera RGB o rozdzielczości 640x480 (6)	Przesyła serie obrazów do komputera z prędkością do 30 klatek na sekundę.

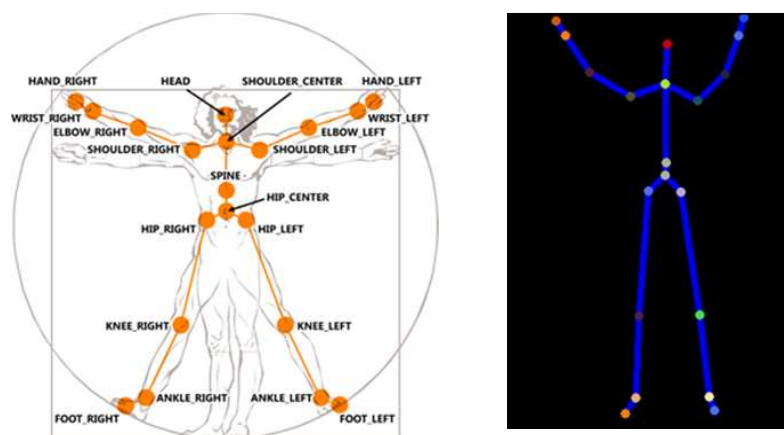
Pomijając zestaw mikrofonów wykorzystywanych do rozpoznawania mowy, najważniejszym wyposażeniem urządzenia stanowi zestaw kamer wraz z emiterym promieni podczerwonych (rys. 2 punkty: 2, 3). Kamera na podczerwień służy do wykrywania głębi, natomiast druga kamera w trybie pracy synchronicznym pobiera kolorowy obraz w przestrzeni barw RGB. Podczas pracy obydwie kamery synchronicznie rejestrują obraz. Z wykorzystaniem podczerwieni z tła obrazu wyodrębniona zostaje sylwetka użytkownika, która w następnym kroku jest podzielona na dwadzieścia (tab. 2), anatomicznych części ciała człowieka (rys. 3).

Tab.2. Opis segmentacji ciała człowieka sensorem Kinect [6]

Lp.	Angielska nazwa z kodu źródłowego programu	Nazwa polska
1	Head	Głowa
2	Shoulder_center	Kark
3	Hand_left	Lewa dłoń
4	Wrist_left	Lewy nadgarstek
5	Elbow_left	Lewy łokieć
6	Shoulder_left	Lewy bark
7	Hand_right	Prawa dłoń

8	Wrist_right	Prawy nadgarstek
9	Elbow_right	Prawy łokieć
10	Shoulder_right	Prawy bark
11	Spine	Kregłostup
12	Hip_center	Miednica
13	Hip_left	Lewe biodro
14	Knee_left	Lewe kolano
15	Ankle_left	Lewa kostka
16	Foot_left	Lewa stopa
17	Hip_right	Prawe biodro
18	Knee_right	Prawe kolano
19	Ankle_right	Prawa kostka
20	Foot_right	Prawa stopa

Wygenerowanie cyfrowego modelu „szkieletu użytkownika” w oparciu o quasi anatomiczną budowę, umożliwia alokację modelu szkieletu użytkownika w trójwymiarowej przestrzeni bez względu na jego wzrost oraz cechy zewnętrzne (rys. 3b). W pracy urządzenia, kluczową rolę odgrywa zastosowanie podczerwieni. Dzięki podczerwieni poprawne rozpoznawanie i segmentacja ciała jest uniezależniona od jakości i ekspozycji oświetlenia pomieszczenia w którym przebywa użytkownik. Z punktu widzenia przetwarzania i analizy obrazu wyeliminowanie zjawiska refleksyjności przedmiotów w zależności od pory dnia lub warunków oświetlenia sceny ekspozycji, stanowi największą wartość dodaną urządzenia.



Rys. 3. Od lewej: a) szkic sylwetki człowieka widziana sensorem Kinect, b) cyfrowy model reprezentacji sylwetki człowieka (szkielet)

3. STEROWANIE LABORATORYJNĄ SUWNICĄ POMOSTOWĄ PRZY POMOCY URZĄDZENIA KINECT

Przez wiele stuleci rozwijająca się stopniowo wiedza o komunikowaniu się ludzi ograniczała się jedynie do zagadnień związanych z wykorzystywaniem w kontaktach społecznych dźwiękowych systemów językowych. Ten naturalny dla człowieka sposób komunikowania został szeroko zaadoptowany oraz zaimplementowany w szeroko dziś stosowanych interfejsach sterowania głosem. Jednakże ze względu na dotychczasowy brak wiedzy, a przede wszystkim niedoskonałą technologię oraz brak znormalizowanych urządzeń wejścia, technika adaptacji niewerbalnych systemów komunikacji służących do porozumiewania się z komputerem przy pomocy gestów była przedmiotem wyłącznie specjalistycznych projektów [2, 3, 9, 12].

Kinect jest urządzeniem wejścia z trójwymiarowym układem odniesienia, cechującym się powszechną dostępnością i wyposażonym we własne biblioteki sterowników i spójnego pod względem architektury interfejsu wymiany danych (USB). Umożliwia budowę różnorodnych aplikacji ukierunkowanych na wkomponowanie w układ C-OT interfejsu komunikacji opartego o naturalny dla człowieka system przekazywania informacji przy pomocy gestów.

3.1. Segmentacja obrazu dla potrzeb sterowania

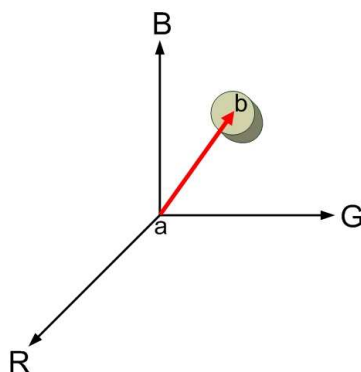
W środowisku Windows, urządzenie Kinect jest rozpoznawane jako szereg urządzeń wykrywanych jako urządzenia audio oraz video (rys. 4).



Rys. 4. Widok menadżera urządzeń systemu Windows z zainstalowanym sensorem Kinect

Wyraźne rozdzielanie sygnału współpracujących urządzeń wejścia umożliwia wprowadzenie sygnału wideo do cyfrowego środowiska MatLab, poprzez narzędzie Image Processing; w taki sam sposób jak dla każdego innego cyfrowego źródła sygnału wideo. Różnica polega na konieczności rozdzielania splecionych ze sobą dwóch źródeł sygnału wideo *przechwyconych* w strumieniu wideo z dwóch niezależnych kamer kinecta. Przy

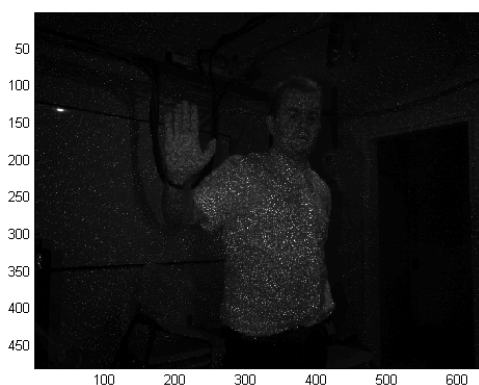
wykorzystaniu funkcji analizy obrazu po odseparowaniu sygnału rejestrowanego poprzez kamerę na podczerwień (rys. 6) zostaje on poddany tak zwanej segmentacji typu RGB [8]. Do segmentacji wykorzystuje się trójkę składowych barw RGB określających położenie wektora w przestrzeni RGB według zależności (1) - rys. 5.



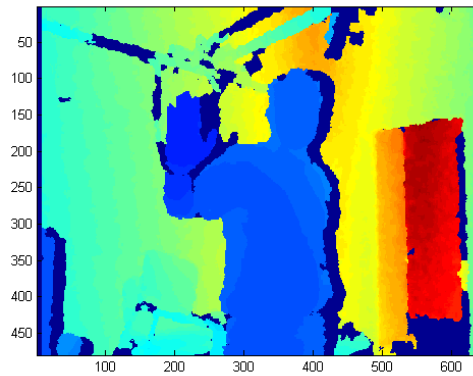
Rys. 5. Wektor głębokości ab w układzie współrzędnych RGB

$$D(a,b) = \sqrt{(a_R - b_R)^2 + (a_G - b_G)^2 + (a_B - b_B)^2}, \quad (1)$$

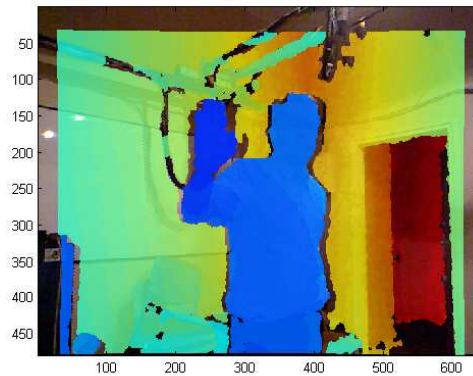
Przy segmentacji RGB, kluczowym problemem polega na określeniu zakresów zmienności składowych RGB oraz sposobu pomiaru podobieństwa między punktami obrazu przestrzeni RGB po odseparowaniu tła (rys. 7). Końcowy etap analizy obrazu polega na sprzęgnięciu obrazu rejestrowanego przez kamerę kolorową wraz z obrazem pochodzącym z segmentacji obrazu widzianego przez kamerę na podczerwień (rys. 8).



Rys.6. Obraz z kamery na podczerwień

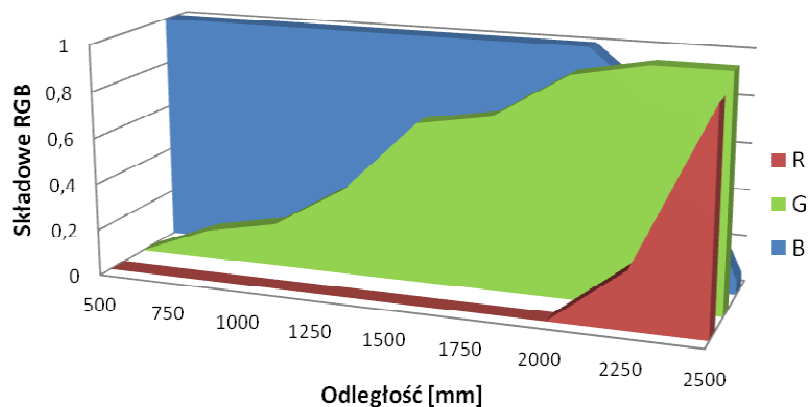


Rys.7. Segmentacja RGB obrazu z kamery na podczerwień



Rys. 8. Obraz hybrydowy powstały poprzez nałożenie segmentowanego strumienia wideo oraz obrazu kolorowego RGB

Uzyskanie obrazu hybrydowego poprzez nałożenie dwóch niezależnie rejestrowanych i przetwarzanych w czasie rzeczywistym obrazów umożliwia odseparowanie sylwetki operatora od tła pomieszczenia wraz z jednoczesnym mapowaniem jego położenia względem urządzenia rejestrującego obraz uwidacznianym poprzez intensyfikację koloru niebieskiego (rys. 8). Na rysunku 9 przedstawiono rozkład kolorów przestrzeni barw RGB rejestrowanych w zależności od odległości operatora względem urządzenia Kinect.







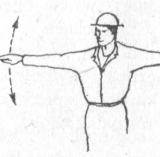
Rys. 9. Rozkład składowych RGB widzianych sensorem Kinect w zależności odległości od sensora

W rezultacie analizy wykresu można stwierdzić, że skuteczna segmentacja RGB jest możliwa dopiero gdy sylwetka operatora znajduje się w odległości większej niż 50 cm od urządzenia. W przedziale od 50 do 200 [cm] segmentacja wykorzystuje rejestrowanie zmiany wartości progu koloru zielonego przy maksymalnym wysyceniu obrazu kolorem niebieskim i nierejestrowanym kolorem czerwonym. Powyżej 200 [cm] stopniowo zaczyna spadać wartość progu dla koloru niebieskiego. Miejsce koloru niebieskiego stopniowo zajmuje kolor czerwony który wraz z większą odległością przybiera na wartości.

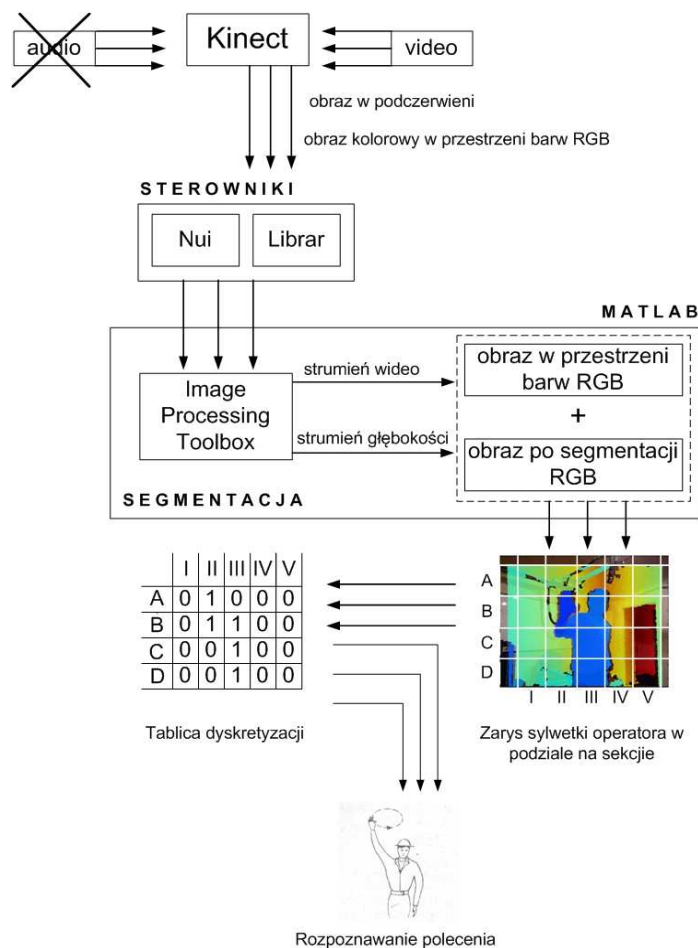
4. ALGORYTM PRZETWARZANIA OBRAZU

Sterowanie dźwignicą z użyciem urządzenia Kinect wykorzystuje znormalizowane sygnały (gesty) porozumiewawcze przekazywane między hakiem oraz operatorem dźwignicy - tab. 3 [14]. Ze względu na ich niepowtarzalny ich charakter, stanowią one bazę do zaimplementowania w urządzenie projektowane dla celów sterowania dźwignicą przy pomocy gestów.

Tab. 3. Wybrane sygnały porozumiewawcze pomiędzy hakowym a operatorem dźwigni

Lp.	Szkic sytuacyjny	Znaczenie/ Czynność	Opis szkicu
1		podnosić	Prawa ręka uniesiona do góry zatacza powolnie okręgi
2		opuszczać	Prawa ręka opuszczona do dołu zatacza powolnie okręgi
3		przemieszczać w poziomie	Ręką uniesioną do boku wskazać kierunek przesunięcia ładunku
4		STOP	Ręka uniesiona do boku wykonuje zdecydowane ruchy wahadłowe w poziomie
5		ALARM	Obie ręce zataczają półokręgi wahadłowe w płaszczyźnie pionowej

Każdy z wyselekcjonowanych gestów jest niepowtarzalny i posiada charakterystyczną cechę, którą łatwo można odnaleźć na obrazie powstałym na skutek nałożenia na siebie obrazu segmentowanego uzyskanego z kamery na podczerwień oraz obrazu rejestrowanego w przestrzeni barw RGB. Na rys. 10 przedstawiono schemat blokowy algorytmu pozyskiwania, segmentacji oraz selekcji znacznika RGB w zależności od sekcji zdjęcia.



Rys. 10. Schemat blokowy algorytmu przetwarzania obrazu dla potrzeb interfejsu sterowania gestami

W przedstawionym algorytmie sygnał wideo pobierany za pomocą urządzenia Kinect oraz bazy bibliotek sterowników NUI oraz Librar jest przekazywany bezpośrednio do środowiska MatLab poprzez środowisko Image Processing Toolbox. Po obróbce sygnału wideo uzyskuje się: obraz rejestrowany w podczerwieni, obraz kolorowy rejestrowany w przestrzeni barw RGB, segmentowany obraz z mapą głębokości RGB oraz obraz hybrydowy powstały na wskutek nałożenia obrazu kolorowego z segmentowaną mapą głębokości. W algorytmie wyróżniono także tak zwane tablice dyskretyzacji, które w zależności od zdefiniowanego zestawu znaczników pojawiających się w odpowiednich obszarach zdjęcia są bezpośrednio odpowiedzialne za rozpoznawanie i generowanie poleceń na podstawie posiadanych wzorców.

5. PODSUMOWANIE

Zmieniające się uwarunkowania wszechobecnej *wirtualnej rzeczywistości* zmuszają do podejmowania działań mających na celu poszukiwanie nowych interfejsów komunikacyjnych na linii człowiek – komputer (rozumiany jako maszyna) ułatwiających komunikację. Komunikacja pomiędzy człowiekiem, a środowiskiem komputerowym w ostatnim czasie ulega istotnym zmianom i przeobrażeniom. Na wzór wykształconych w procesie ewolucji zmysłów człowieka w których to wzrok dostarcza 90% informacji o otaczającym świecie, dodatkowo adaptując inne zmysły naturalne w procesie komunikacji między ludźmi. Dziś już nie tylko słuch (interfejsy sterowane głosem), dotyk (ekrany dotykowe) ale również umiejętność rozpoznawania kształtu dzięki urządzeniu Kinect ma szansę stać się standardem i obok klawiatury, myszy komputerowej znaleźć swoje stałe miejsce na biurku przeciętnego użytkownika ułatwiając codzienną pracę.

W artykule przedstawiono realizację koncepcję wkomponowania w układ C-OT-O nowego typu interfejsu komunikacyjnego opartego o niewerbalny, bazujący na gestach, układ przekazywania informacji do środowiska cyfrowego. Eksperyment zakończony wynikiem akceptowalnym, rokującym na dalszy rozwój zrealizowano na laboratoryjnej suwnicy pomostowej.

Podziękowanie

Praca badawcza sfinansowana ze środków budżetowych na naukę w latach 2008-2011.

6. LITERATURA

- [1] Baecker R.M, Grudin J., Buxton W.A., Greenberg S. (Eds.): *Human Computer Interaction: Toward the Year 2000*. Morgan Kaufmann, San Francisco, 1995 (ISBN 1-558-60246-1).
- [2] Chang Y.J., Chen S.F., Chuang A.F.: *A gesture recognition system to transition autonomously through vocational tasks for individuals with cognitive impairments*. Research in Developmental Disabilities, v. 32, pp. 2064–2068, 2011.
- [3] Dutta T.: *Evaluation of the Kinect_ sensor for 3-D kinematic measurement in the workplace*. Article in Press, Applied Ergonomics, 2011.
- [4] Helander M.G., Landauer T.K., Prabhu P.U. (Eds.): *Handbook of Human-Computer Interaction*. Elsevier Science, Amsterdam, 1997 (ISBN 0-444-81876-6).
- [5] Kean S., Hall J., Phoenix Perry P.: *Meet the Kinect: Programming and Scripting Natural User Interfaces*. Apress L.P., 2011 (ISBN 978-1-4302-3888-1).
- [6] Kowalczyk T.: *Kinect SDK – Wprowadzenie*. Artykuł dostępny on-line: <http://msdn.microsoft.com/pl-pl/library/kinect-sdk--wprowadzenie.aspx>
- [7] Kramer J.: *Hacking the Kinect*. Apress L.P., 2011 (ISBN 978-1430-2386-76).
- [8] Kristensen F., Nilssen P., Öwall V.: *Background Segmentation Beyond RGB*. Conference Proceedings of 7th Asian Conference of Computer Vision, Hyderabad, India 2006, pp.602-623, Part II, Springer-Verlag, Berlin Heidelberg, 2006.
- [9] Semeraro F.: *Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality improvement*. Article in Press, Resuscitation, 2011.

-
- [10] Stevens T.: *Kinect for Windows SDK beta launches, wants PC users to get a move on*. Artykuł dostępny on-line: <http://www.engadget.com/2011/06/16/microsoft-launches-kinect-for-windows-sdk-beta-wants-pc-users-t/>
- [11] Szpytko J., Hyla P.: *Ontologia i semantyka w inżynierii procesów decyzyjnych eksploatowanych środków transportu*. Materiały VI Konferencji LogiTRANS: Logistyka, Systemy Transportowe, Bezpieczeństwo w Transporcie, Dokument elektroniczny 12 stron, Szczyrk, 15-17 kwietnia 2009.
- [12] Szpytko J., Hyla P.: *Work space supervising for material handling devices with machine vision assistance*. Journal of KONBiN, Safety and reliability systems, no 3-4, pp.7-16, Warsaw, 2009 (ISSN 1895-8281).
- [13] Szpytko J., Kocerba A.: *Telematics in supervision of manufacture transport device*. ZN Politechniki Śląskiej, Transport, nr 1657, z.55, s.417-422, Gliwice, 2004.
- [14] PN-78/M-45980; Sygnały porozumiewawcze podawane przez hakowego.