INTERACTIVE MEDIA CONTROL USING NATURAL INTERACTION-BASED KINECT

Madjid Maidi and Marius Preda

Département Artemis - Télécom SudParis
9 Rue Charles Fourier - 91000 Évry, France
{madjid.maidi ; marius.preda}@it-sudparis.eu

ABSTRACT

In this work we present a novel interaction approach based on a gesture recognition system using a Microsoft Kinect sensor. Gestures are defined and interpreted in order to activate controls on a media device. This natural interface enables intuitive interaction with the multimedia content. The depth sensor observes the scene to detect a request for control in the form of a gesture. Then, the application assigns a control for the media system. The application is tested under various scenarios and proved the reliability and the effectiveness of the proposed approach.

Index Terms— Interactive media, Microsoft Kinect, gesture recognition

1. INTRODUCTION

Interactive media is a set of digital content which consists of graphics, images, sound, etc. clustered within an environment enabling to interact with the data and to control applications. This media provides services on digital computer-based systems relying on user actions. Obviously, the most known input and output devices designed to communicate with the system are the keyboard and the mouse. These devices react to the user commands and send signals to the graphical user interface to perform the required task.

To improve human-computer communication and interaction, new media environments have been proposed. The concept is to involve the user as a component of the system by developing a new device-independent approach that provides natural interaction. Advanced user interfaces are based gesture recognition methods. These techniques allow to determine the user motion and intents and interpret gestures into a meaningful input controls while ensuring an intuitive interaction.

The remote sensing is more natural and comfortable because one doesn't need to wear any equipment to interact with the system. In recent years, much efforts were invested to build gesture-based user interfaces. Gesture recognition approaches were interested in hand motions by developing vision-based system to recognize the human gestures [1] [2] [3] [4] [5] [6] [7]. However, vision system can not fulfill the accuracy specifications required by an effective recognition method .

On the other side, the emergence of the Virtual Environment technology enabled to develop new 3D user interfaces and proposed many systems based on motion tracker devices to locate the user within the environment. This approach enhances natural interaction and integrates the human sense in the system for better immersion of users in the virtual world [8] [9] [10] .

The advent in end 2010 of the Kinect sensor opened new prospects for developing new type of applications relying to depth sensing (http://www.xbox.com/en-US/Kinect). This technology was, originally, intended to control video games using depth data to identify the player motions and gestures. The release of the Kinect enabled to exploit depth maps and skeletal joint position data. With this technology, body detection and pose estimation required for gesture recognition were facilitated [7] [11] [12] [13] [14] [15].

Although gesture recognition systems exist for specific applications, mainly games, no application has been developed to interact with digital media, for example: controlling a television or browsing multimedia content using only gestures. Therefore, the aim of this work is to develop a novel application intended to control an interactive media via a gesture recognition system. This application enables browsing pictures of a photo album (figure 1).

The remainder of the paper is as following: section 2 describes the whole system functionalities. Then, we detail our method of hand detection and tracking in section 3. In section 4, we present the gesture recognition technique used to map gestures to the interface controls. Section 5 presents the natural interaction results and we finish with section 6 where we present conclusion and future works.

2. SYSTEM OVERVIEW

The natural interaction considers the human motions and gestures as devices of remote controls. The system receives instructions via recognized gestures which activate the application controls. The objective of this work is to create an application that interacts with an interactive media using gestures.

The system interprets hand motions and gestures using the depth image provided by the infrared sensors of the Kinect (figure 2). The hand gestures have to be mapped into interaction functionalities. The application enables browsing pictures of photo album with hand gestures. The user can view the photos by moving his hand.

Fig. 1. Interactive media control application.

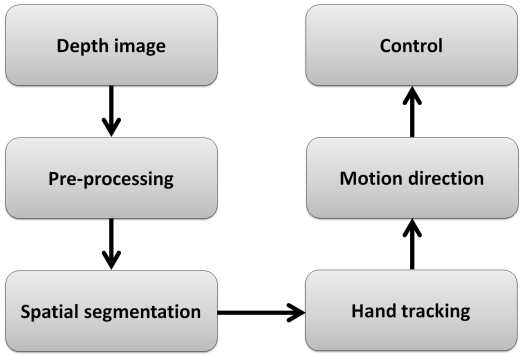

Fig. 2. System overall design.

3. ALGORITHM DESIGN FOR HAND DETECTION AND TRACKING

Our approach relies upon the use of depth images provided by the Kinect sensor. The input image is processed with three successive blocks to accomplish tasks of detection and tracking (figure 3). The Kinect captures the information using the infrared depth sensors which can easily get the location of foreground targets while discarding distant objects. The user interacts with the Kinect with his hands which should be in the front, so the Kinect can extract the hand shape by measuring the depth. The segmentation part is composed of several steps enabling to compute pixel disparity and affecting colors to several distance ranges from the the Kinect position. The hand is, then, properly detected by retrieving the nearest object based on distance information. The hand shape is traced with curvature points defined using a contour detector. The step further is approximating the hand with a polygon to reduce the number of unwanted convexity points and representing the hand shape with main vertices. Finally, the convex hull step enables filtering concave points and grouping neighboring features. Once the hand is detected, the tracking is performed by computing, continuously, the positions of feature points and recovering the motion direction.

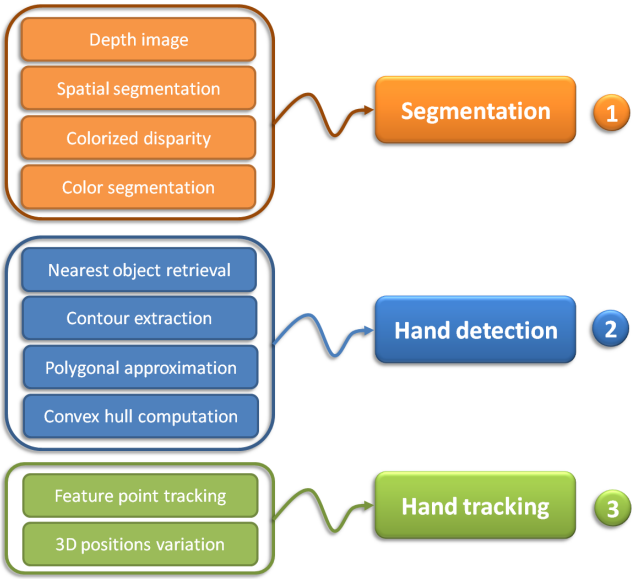

Fig. 3. Hand detection and tracking.

4. GESTURE RECOGNITION USING THE KINECT

To start controlling the device, the user performs a focus gesture in the image center to initialize the hand tracking process. Typically, the four motions we defined are: a click gesture by pushing the hand towards the Kinect sensor. A pull movement to enable returning to the parent level of hierarchy (thumbnail). Moving hand in left or right direction to look around all the photos (figure 4). The hand is detected using the depth map and a spatial segmentation of objects. The gestures are defined with the computation of distance variances along 3D axes.

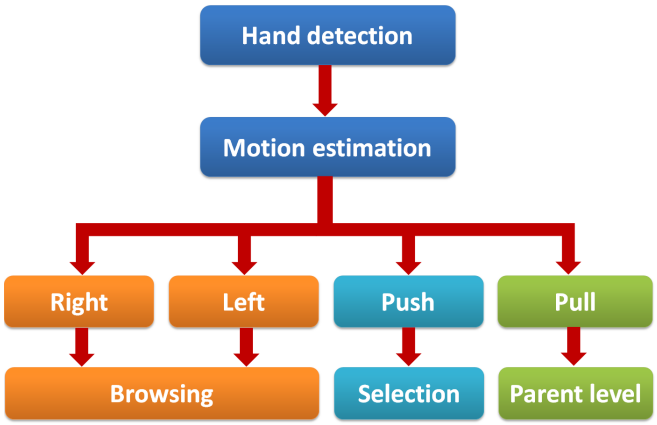

Fig. 4. Gestures used for application control.

5. NATURAL INTERACTION RESULTS

The application we have developed enables the control of a set of photos organized in a image viewer. The user can perform the predefined gestures to activate and control the media. The Kinect captures in real-time the depth images which are presented to the processing module. The motion is analyzed and interpreted as a gesture. The proposed method is based on depth information to segment objects according to their spatial location. A color segmentation step is then applied to assign colors to different regions classified by their distance to the camera. The closest objects are filtered by their color and considered as object of interest (figure 5). Once the hand is detected, the next step aims to determine the movement directions. Hand feature points are tracked consecutively in frames, to identify the gesture, we compute the changes in the depth channel and horizontal and vertical translations.

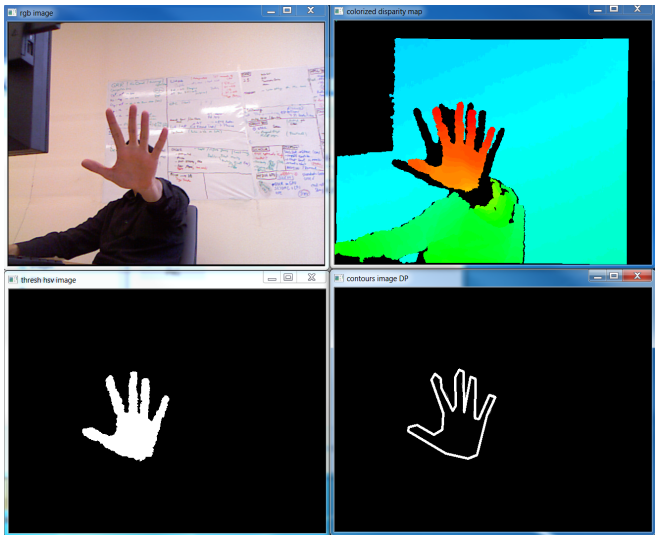

Fig. 5. Hand detection and tracking.

The four gestures (right, left, push and pull) are translated to a data stream via the gesture recognition module and sent to the interactive player by network communications. The advantage of using the interactive media is the ability to interact with the visual contents and sending information via network sockets to launch a control (figure 6).

Experiments were conducted to show the reliability of the gesture recognition method (figure 7). Indeed, we performed a series of gestures that we defined in our application. We have to note that the Kinect sensors should be distant to the user's hand approximately $40cm$ and should not exceed $130cm$ according to our experimental setup. The depth values are available for this distance range only. The blind region represents an area of limited fields of view for the infrared sensors.

Figure 7 illustrates also a tracking process in a real-time experiment. The user can move his hand freely and the recognition method retrieves feature points of the object of interest and computes its spatial displacement. This motion is mapped into interface controls enabling natural interaction with the media.

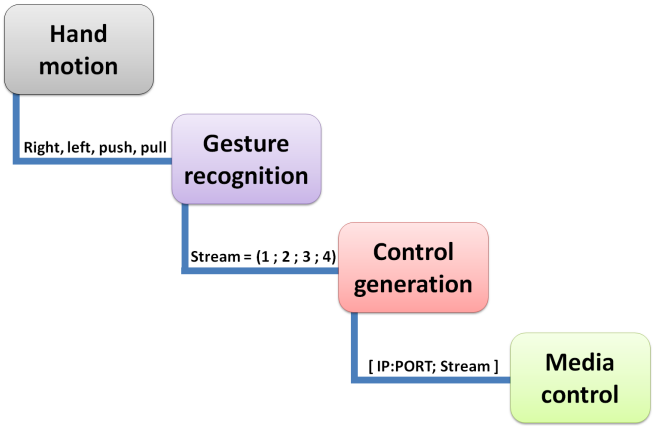

Fig. 6. Interactive media control using gestures.

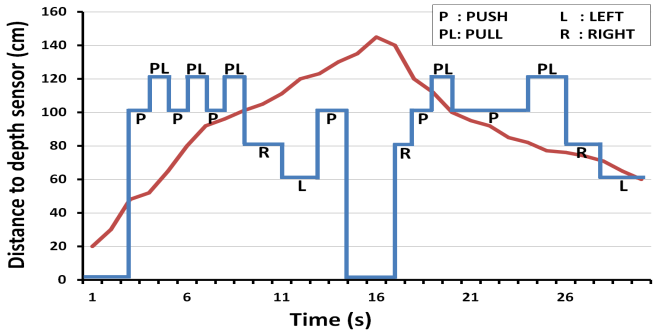

Fig. 7. Gesture identification and tracking.

6. CONCLUSION

This paper presented a new approach for natural interaction with an interactive media using hand gestures. We used the depth map provided by the Kinect sensor to easily segment the nearest object in the image. The main novelty of this work is the control of a media with hand motion contrarily to other existing applications which are mainly intended to games. This kind of applications is a new way of using human body to perform remote control without need of any device or peripheral. The proposed algorithm segments a moving object in the image, however, we assume that the object should be in the foreground to be segmented correctly. Otherwise, other objects interact with the object of interest and and are considered as part of the moving one. The application was implemented in real-time and proved its effectiveness to generate controls using hand gestures.

Our future work will focus on enhancing the spatial segmentation by reducing detection errors and defining more controls while recognizing advanced complex gestures.

7. REFERENCES

[1] Qing Chen, N.D. Georganas, and E.M. Petriu, "Real-time vision-based hand gesture recognition using haar-like features," in *Proc. of Instrumentation and Measurement Technology Conference (IMTC)*, Warsaw, Poland, 2007, pp. 1–6.

[2] J. Nagi, F. Ducatelle, G.A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L.M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *IEEE Int. Conf. on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, 2011, pp. 342 –347.

[3] M. Manigandan and I. Manju Jackin, "Wireless vision based mobile robot control using hand gesture recognition through perceptual color space," in *Int. Conf. on Advances in Computer Engineering (ACE)*, Taipei, Taiwan, 2010, pp. 95–99.

[4] S. Koceski and N. Koceska, "Vision-based gesture recognition for human-computer interaction and mobile robot's freight ramp control," in *Int. Conf. on Information Technology Interfaces (ITI)*, Croatia, 2010, pp. 289–294.

[5] Chenglong Yu, Xuan Wang, Hejiao Huang, Jianping Shen, and Kun Wu, "Vision-based hand gesture recognition using combinational features," in *Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, Darmstadt, Germany, 2010, pp. 543–546.

[6] Po-Chuan Cho, Chao-Tang Li, and Wen-Hui Chen, "Implementation of low-cost vision-based gesture recognition systems based on fpga approach," in *Int. Symp. on Computer, Consumer and Control (IS3C)*, Taichung, Taiwan, 2012, pp. 329–332.

[7] K.K. Biswas and S.K. Basu, "Gesture recognition using microsoft kinect," in *Proc. of Automation, Robotics and Applications (ICARA)*, Wellington, New Zealand, 2011, pp. 100–103.

[8] D.A. Bowman and R.P. McMahan, "Virtual reality: How much immersion is enough?," *Computer*, vol. 40, no. 7, pp. 36–43, 2007.

[9] R.P. McMahan, A.J.D. Alon, S. Lazem, R.J. Beaton, D. Machaj, M. Schaefer, M.G. Silva, A. Leal, R. Hagan, and D.A. Bowman, "Evaluating natural interaction techniques in video games," in *IEEE Symp. on 3D User Interfaces (3DUI)*, Massachusetts, USA, 2010, pp. 11–14.

[10] R.P. McMahan, D.A. Bowman, D.J. Zielinski, and R.B. Brady, "Evaluating display fidelity and interaction fidelity in a virtual reality game," *IEEE Trans. on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 626–633, 2012.

[11] T. Gonzalez-Sanchez and D. Puig, "Real-time body gesture recognition using depth camera," *Electronics Letters*, vol. 47, no. 12, pp. 697–698, 9 2011.

[12] Zhou Ren, Jingjing Meng, and Junsong Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Int. Conf. on Information, Communications and Signal Processing (ICICS)*, Beijing, China, 2011, pp. 1–5.

[13] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using kinect camera," in *Int. Joint Conf. on Computer Science and Software Engineering (JCSSE)*, Bangkok,n Thailand, 2012, pp. 28–32.

[14] M.K. bin Mohd Sidik, M.S. bin Sunar, I. bin Ismail, M.K. bin Mokhtar, and N.B.M. Jusoh, "A study on natural interaction for human body motion using depth image data," in *Workshop on Digital Media and Digital Content Management (DMDCM)*, Hangzhou, China, 2011, pp. 97–102.

[15] Kam Lai, J. Konrad, and P. Ishwar, "A gesture-driven computer interface using kinect," in *2012 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, New Mexico, USA, 2012, pp. 185–188.