Gesture Recognition based on 2D and 3D Feature by using Kinect Device

Wenkai Xu, Eung Jo Lee

Department of Information & Communications Engineering, Tongmyong University, Busan, Korea
xwk6298@hotmail.com, ejlee@tu.ac.kr

Abstract. Human gesture recognition is a non-verbal part for interaction or movement that can be used to involves real world and virtual world. In this paper, we explain a study in human gesture recognition using RGB color information and depth information by Kinect camera from Microsoft Corporation. To achieve the goal, hand tracking and gesture recognition have no major dependencies of the work environment, lighting or users’ skin color, libraries of particular use for natural interaction and Kinect device, which serves to provide RGB images of the environment and the depth map of the scene were used. An improved Camshift tracking algorithm combined with depth information is used to tracking hand motion, and then an associative method of HMM and FNN is propose for gesture recognition step. The experimental results show out its good performance and it has higher stability and accuracy as well.

Keywords: Camshift, depth information, gesture recognition, Kinect

1 Introduction

Massive technology shift has played a dominant role in all disciplines of science and technology. The use of hand gesture is an active area of research in the vision community, mainly for the purpose of sign language recognition and Human-Computer Interaction (HCI). The history of interaction and interface design is a flow and step from complex interaction to simple interaction between human and computer [1]. The word natural interaction came from Natural User Interface (NUI) that use human body interaction and voice interaction, verbal and non-verbal communication, becoming a one of Human-Computer Interaction (HCI) area. It is an evolution from Graphical User Interface (GUI).

In this paper, An improved Camshift tracking algorithm combined with depth information is used to tracking hand motion by Kinect, and then an associative method of HMM and FNN is propose for gesture recognition step, which combines ability of HMM model for temporal data modeling with that of fuzzy neural network for fuzzy rule modeling and fuzzy inference.
2 Hand Tracking and Recognition

Because the Camshift algorithm is based on color images, tracking error will easily occur when there is similar color in background. Considering the object is usually separated from the surrounding environment in depth, and has fixed moving range, so threshold segmentation in depth map can accurately distinguish the player from the background. According to reference [2], we combined depth information with traditional Camshift tracking algorithm by using Kinect. Gesture Recognition Using Improved HMM Algorithm

Choosing suitable features to recognize the hand gesture path play significant role in the whole system. There are three basic features: location, orientation and velocity. The previous research [3, 4] showed that the orientation feature is the best in term of accuracy results. Therefore, we regard the orientation feature as the main feature during our research process. Based on the research above, a gesture path is spatiotemporal pattern which consists of centroid point \((x, y)\). So, the orientation is determined by the change between two consecutive points from hand gesture path [5].

Fuzzy Neural Network has strong ability for fuzzy rule modeling and fuzzy inference due to its integration of fuzzy set theory and Neural Network together. Since traditional FNN cannot model temporal data and conventional HMM do not own ability for fuzzy inference, we integrate the two models together to represent complex gesture trajectory and perform inference by the integrated HMM-FNN model based on [6] [7] [8], which is shown in Fig.4, for the recognition of dynamic gesture.

HMM-FNN model includes five layers. Its first layer, second layer and HMM layer constitute the fuzzy preprocessing part, third layer and fourth layer constitute fuzzy inference part, fifth layer is the defuzzification part of HMM-FNN and produce distinct output. The following will introduce these five layers in detail.

3 Experimental results

In our experiments, the fuzzy rules, with a total number of characters, are obtained by data clustering combined with human experiences. The initial connecting weights are also set by people’s prior knowledge about dynamic gestures. For each dynamic gesture, we ask each of 5 testers to perform it 5 times, and then get 25 video sections. As a result, there are totally 125 samples for all 5 kinds of gestures, 100 of which are used for model training and the others are for testing. When the error of testing is below the threshold or training times reach its maximum, the model is considered well-trained.

4 Conclusion

In this paper, we propose an automatic system to recognize gestures in real-time. At first, an improved Camshift tracking algorithm combined with depth information is
used to tracking hand motion; Next, HMM-FNN model is proposed for gesture recognition, which combines ability of HMM model for temporal data modeling with that of fuzzy neural network for fuzzy rule modeling and fuzzy inference. The experimental results show out its good performance and it has higher stability and accuracy as well. In the future work, we will study about complex gesture recognition using Kinect for reflecting the methodology we proposed better and making more abundant controlling contents for NUI.

Acknowledgement. This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the IT/SW NHN Program supervised by the NIPA (National IT Industry Promotion Agency)” (NIPA-2011-C1820-1102-0010).

Reference