Continuous Gesture Recognition System Using Improved HMM Algorithm Based on 2D and 3D Space

Wenkai Xu and Eung-Joo Lee

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

Abstract

In this paper, we explain a study on natural user interface (NUI) 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: *NUI, depth information, improved HMM, Kinect*

1. Introduction

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).

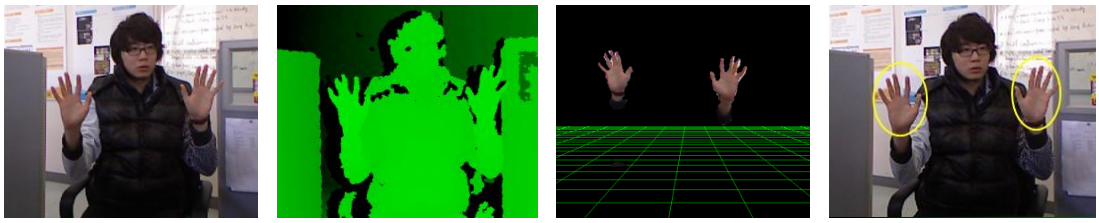
Gesture and posture recognition are application areas in HCI to communicate with computers. A gesture is spatiotemporal pattern which maybe static, dynamic or both. Static morphs of the hands are called postures and hand movements are called gestures. In gesture recognition, Yoon et al. [2] developed a hand gesture system in which combination of location, angle and velocity is used for the recognition. Liu et al. [3] developed a system to recognize 26 alphabets by using different HMM topologies. Hunter et al. [4] used HMM for recognition in their approach where Zernike moments are used as image features for sequence of hand gestures. In the last decade, several methods of potential applications in the advanced gesture interfaces for HCI have been suggested but these differ from one to another in their models. Some of these models are Neural Network [5], Hidden Markov Model (HMM) [6] and Fuzzy Systems [7]. Hidden Markov Model (HMM) is one of the most successful and widely used tools for modeling signals with spatiotemporal variability [8]. It has been successfully applied in the area of speech recognition and is one of the mostly successfully used methods in the research area of dynamic gesture recognition. There are several papers that survey HMM methods [6] used in dynamic gesture recognition.

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 using Improved Camshift Algorithm based on Depth Information

Camshift tracking algorithm based on color performs well in solving the bottom problems of computer vision. Due to its robust and real-time quality, Camshift has become a basic tracking method which can adapt to the continuous variation of the shape and size of the target, compute fast and has strong anti-jamming capability, guaranteeing the stability and real-time of the system. Camshift algorithm is a dynamic change in the distribution of the density function of the gradient estimate of non-parametric methods.

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 [11], we combined depth information with traditional Camshift tracking algorithm by using Kinect. The tracking results are shown as Figure 1.



(a) (b) (c) (d)

Figure 1. (a) Original Image (b) Depth Image (c) Hand Area Extraction (d) Hand Tracking

3. Gesture Recognition using Improved HMM Algorithm

3.1 Feature Extraction

The previous research [9, 10] 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_{hand}, y_{hand}). So, the orientation is determined by the change between two consecutive points from hand gesture path.

The $p_i(x_i, y_i)$ position data of points of the hand trajectory are converted into direction codes representing direction vectors (Fig.3). The moment distance d_i and changing angle θ_i are calculated for each position by the following equations [12]:

$$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (1)$$

$$\theta_i = \tan^{-1}\left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right) \quad (2)$$

The angles are each converted into one of the eight direction codes. The angle ranges of the direction codes have different widths. The feature of 2D motion trajectory of hand gesture is comprised of a series of discrete movement points, and it is represented by a series of discrete movement direction value. For the 2D motion plane, we divide the direction into eight discrete values as shown. Therefore, the trajectory of dynamic gesture ban be described by the sequence of discrete direction value T: $t_1, t_2, t_3, t_4 \ldots t_i (1 \leq t_i \leq 8)$.

3.2 Continuous Gesture Recognition using Improved HMM Algorithm

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], for the recognition of dynamic gesture. HMM-FNN model includes five layers [13]. 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.

The first layer is the input layer of the model and it has three neurons, which correspond to the three movement components of dynamic gesture, i.e. Q_T,Q_Z respectively.

The second layer and HMM layer compose fuzzification layer. Each HMM model is related to a neuron in second layer, which represent a fuzzy class to which the input observation possibly belongs. The likelihood of input observation sequence Q to each HMM, i.e. p (Q/λ) is considered as membership value of the corresponding fuzzy class variable. At the same time, the neurons in second layer constitute the antecedent part (conditional part) of fuzzy rule. The number of neurons of this layer is m1+m2, where m1, m2 are the class numbers of 2D trajectory and movement in the Z-axis direction respectively.

The third layer is the layer of fuzzy inference, and each neuron represents a fuzzy rule. The connecting weights between neurons in second and third layer imply the contribution degree of the antecedent part for this rule. The output of neuron in third layer is calculated as shown in Eq.3.

$$O^{(3)} = b = \sum_{i=0}^{m} \omega_i I_i^{(3)} = \sum_{i=0}^{m} \omega_i \, p\left(\frac{Q}{\lambda}\right), \text{and } \sum \omega_i = 1 \tag{3}$$

The fourth layer is normalization layer, the neuron number of which is equal to that of third layer.

In order to speed up convergence of the network during training, the output of third layer is normalized to assure the sum of them is equal to 1. Output of its neuron is shown as Eq.4.

$$O_i^{(4)} = I_i^{(4)} / \sum_{i=0}^{N} I_i^{(4)} \tag{4}$$

The fifth layer is the defuzzification layer, the output of which is shown as Eq.5.

$$O^{(5)} = \sum_{j=1}^{N} \omega_j \, O_j^{(4)}, \text{and } \sum_{j=1}^{N} \omega_j = 1 \tag{5}$$

Where ω_j implies the importance of each rule for the final classification output, N is the total number of fuzzy rules.

We choose left-right banded model as the type of HMM model due to its straightforward structure. Corresponding to the features' type, the type of HMM models for posture changing and movement in Z-axis direction are one-dimensional continuous HMM models, while that of 2D trajectory is a one-dimensional discrete one. As for continuous HMM model, we employ Gaussian Mixture Model (GMM) as the emission probability of observation, which has the likelihood as described in Equ.6:

$$p\left(\frac{O}{\lambda}\right) = \sum_{i=1}^{M} \omega_i g_i(x) \tag{6}$$

Where ω_j is the weight of i^{th} Gaussian component.

In our system, we defined four steps of movement in the Z-axis direction, including moving towards, moving away from Kinect, moving towards and then moving away, and keeping constant. When the hand moving towards and then moving away from Kinect in Z-axis direction, we regard it as "Click" function.

Suppose that complex gesture trajectory has already been decomposed into three independent parts during hand tracking. The three feature sequences are considered as input of HMM-FNN model, and calculate the likelihood of HMM model p (Q/λ) according to forward probability method. The final output of HMM-FNN model indicates the class type to which the input gesture belongs, such as the output of trajectory A is between the range (α, β] and trajectory B is between (β, γ] and so on. The continuous gestures paths are recognized by its discrete vector and HMM Forward algorithm corresponding to maximal gesture models over the Viterbi best path. Moreover, BW algorithm is used to do a full training for the initialized HMM parameters to construct gestures database.

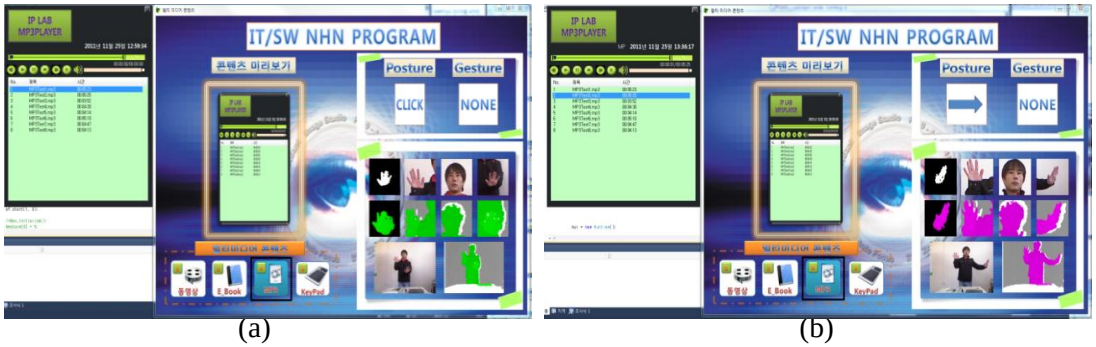
4. Experimental Results

Based on the method we proposed, we implement an Intelligent Media Controlling System based on computer vision, including MP3 player, Movie player, E-Book. As five types have been defined, four of which have been demonstrated: upwards, downwards, leftwards and rightwards. Besides, there is another one type of movement in the Z-axis direction defined in our system: moving towards or away from the camera. In the controlling system we presented, we regard these five types gesture as different kinds of meanings for control signal. We defined the upwards gesture as "Turn on the volume", downwards gesture as "Turn down the volume", leftwards gesture as "Previous" and rightwards gesture as "Next", the type of movement in the Z-axis direction is defined as the signal of "Play Or Cancel". The system GUI and experimental results are shown as Figure 2 and Table 1.

5. 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 HCI.



(a) (b)

Figure 2. (a) Click Function for Starting MP3 Player (b) "Rightwards" Gesture for Next Music

Table 1. Gesture Recognition Results

	Num	Percent
Leftwards COR	23	92%
Rightwards COR	23	92%
Upwards COR	24	96%
Downwards COR	24	96%
Click COR	22	88%
Total COR	116	92.8%
Total Test	125	100%

Acknowledgements

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).

References

[1] A. Valli, "The design of natural interaction", Multimedia Tools Appl. 38(3), pp. 295-305 **(2008)**.
[2] H. S. Yoon, J. Soh, Y. J. Bae and H. S. Yang, "Hand gesture recognition using combined features of location, angle and velocity", Pattern Recognition, vol. 34, pp. 1491-1501 **(2001)**.
[3] N. Liu, B. Lovel and P. Kootsookos, "Evaluation of hmm training algorithms for letter hand gesture recognition", Proc. IEEE Int'l Symp. On Signal Processing and Information Technology, pp. 648-651 **(2003)**.
[4] E. Hunter, J. Schlenzig and R. Jain, "Posture estimation in reduced-model gesture input systems", Proceedings of International Workshop on Automatic Face-and Gesture-Recognition (**1995**), pp. 290-295.
[5] X. Deyou, "A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG", ICPR Conference (**2006**), pp. 519-522.
[6] M. Elmezain, A. Al-Hamadi and B. Michaelis, "Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences", W S C G Journal, Vol.16(1), pp. 65-72 **(2008)**.
[7] E. Holden, R. Owens and G. Roy, "Hand Movement Classification Using Adaptive Fuzzy Expert System", Expert Systems Journal, Vol. 9(4), pp. 465-480 **(1996)**.
[8] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proc. IEEE, vol.77, no. 2, pp.257-286 **(1989)**.

[9] M. Elmezain, A. Al-Hamadi and B. Michaelis, "Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences", W S C G Journal, Vol. 16(1), pp. 65-72 **(2008)**.
[10] L. Nianjun, C. L. Brian, J. K. Peter and A. D. Richard, "Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System", in: International IWFHR, pp.100-106 **(2004)**.
[11] H. Hai, L. Bin, H. BenXiong and C. Yi, "Interaction System of Treadmill Games based on depth maps and CAM-Shift", IEEE 3rd International Communication Software and Networks **(2011)**, pp.219-222.
[12] W. Xu and E.-J. Lee, "Continuous Gesture Trajectory Recognition System Based on Computer Vision", Appl. Math. Inf. Sci., Vol. 6, No. 5S **(2012)**.
[13] W. Xu and E.-J. Lee, "Gesture Trajectory Recognition System Based on Improved HMM Algorithm", Journal of Korea Multimedia Society, Vol. 15, No. 4 **(2012)**.

Authors

Wenkai Xu

Wenkai Xu received his B. S. at Dalian Polytechnic University in China (2006-2010). Currently, he is studying in Department of Information and Communications Engineering Tongmyong University, Korea for master degree. His main research areas are image processing, computer vision, biometrics and hand recognition.

Eung-Joo Lee

Eung-Joo Lee received his B. S. , M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997 he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of Digital Net Bank Inc. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests include biometrics, image processing, and computer vision.