Dynamic Human Activity Recognition Based on Improved FNN Model

Wenkai Xu[†], Eung-Joo Lee[††]

ABSTRACT

In this paper, we propose an automatic system that recognizes dynamic human gestures activity, including Arabic numbers from 0 to 9. We assume the gesture trajectory is almost in a plane that called principal gesture plane, then the Least Squares Method is used to estimate the plane and project the 3-D trajectory model onto the principal. An improved FNN model combined with HMM is proposed for dynamic gesture recognition, which combines ability of HMM model for temporal data modeling with that of fuzzy neural network. The proposed algorithm shows that satisfactory performance and high recognition rate.

Key words: gesture trajectory; gesture detecting; gesture tracking; trajectory recognition; HMM-FNN

1. INTRODUCTION

Tremendous 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).

HCI is emerged as a new field with the motivation to bridge the communication gaps among the humans and computers. Gesture and posture rec-

※ Corresponding Author : Eung-Joo Lee, Address :
Dept. of Information and Communications Engineerings,
535 Yongdang-Dong, Nam-Gu, Busan 608-711, Korea ,
TEL : +82-51-629-1143, FAX : +82-51-629-1129, E-mail
:ejlee@tu.ac.kr
Receipt date : 2011.4.14   Revision date : 2012.1.26
Approval date : 2012.3.31
   [†] Department of Information & Communications Engin-
     eering, Tongmyong University, Busan, Korean
     (E-mail: xwk6298@hotmail.com)
   [††] Department of Information & Communications Engin-
     eering, Tongmyong University, Busan, Korean
     (E-mail: ejlee@tu.ac.kr)

ognition 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. [1] developed a hand gesture system in which combination of location, angle and velocity is used for the recognition. Liu et al. [2] developed a system to recognize 26 alphabets by using different Hidden Markov Model topologies. Hunter et al. [3] used Hidden Markov Model 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 [4], Hidden Markov Model [5] and Fuzzy Systems [6]. Hidden Markov Model is one of the most successful and widely used tools for modeling signals with spatio-temporal variability [7]. 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

Hidden Markov Model methods [5] used in dynamic gesture recognition. In this paper, we also select an improved Hidden Markov Model to recognize the gestures.

In this paper, we propose a novel method to recognize the continuous Arabic numbers (0-9) and alphabets (A-Z) in real-time from color image sequences by the motion trajectory of a single hand using Hidden Markov Model-FNN. Overall, our system is mainly consist of the following parts: we present a new method to get skin-color segmentation based on mixing nonlinear YCbCr elliptic cluster skin-color model and HIS skin-color segmentation model; we estimate the principal gesture plane using Least Squares Method and classifying gestures using Hidden Markov Model-FNN; the likelihood of each Hidden Markov Model to observation sequence is considered as membership value of FNN, and gesture is classified through fuzzy inference of FNN. The gesture recognition process described as the flow chart (Fig. 1). By the propose algorithm we achieve better recognition result for continuous gestures Trajectory.
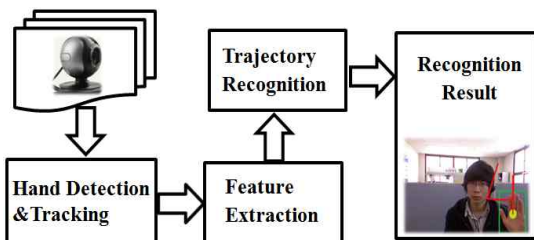

Fig. 1. Flow Chart

2. Hand segmentation and tracking

To be color images, the information of skin-color is very important characteristics for human face. Research shows that: even though of different races, different ages and different gender, the difference in color chrominance is far less than the difference in the brightness. Skin distribution shows clustering distribution in the skin-color space without luminance influence.

In order to reduce the impact of brightness, normally use nonlinear YCbCr elliptic cluster skin-color segmentation model. We ignore Y channel in order to reduce the effect of brightness variation and use only the chrominance channels which fully represent the color information. The research of a large number of color pixels shows that skin-color cluster a very small range of CbCr color space. Normalized chrominance distribution maps, we can find that different skin-color have the same 2D Gaussian model. This method can be accurate to detect skin-color regions. But because of illumination and complex background similar to skin-color effect, this method still may make skin-color region as non-skin color, and make non-skin color as skin-color.

HSI color space contains hue(H), saturation(S) and luminance(I). We know that HSI color space can detect the edge information of face completely, which can remedy defects of YCbCr skin-color segmentation. However, by the effects of illumination, skin color region probably is made as non-skin color region, that is unfavorable for skin-color detection and face location.

In this paper, a skin-color region detection method for the color image with complex background is presented, which is a mixed skin-color segmentation model in both YCbCr and HSI color space constructed. The mixed skin-color segmentation model can be used to segment of possible skin regions in the original image. We fuse the results receive from two methods, in other words, we perform a pixel OR on two expressions. Compared with other methods, we can get the better segmentation result.

For the skin segmentation of hands and face in stereo color image sequences an algorithm is used, which calculates the depth value in addition to skin color information. By the given depth information from the camera set-up system [2], the overlapping problem between hands and face is solved since the hand regions are closer to the camera rather

than the face region. In addition, we use blob analysis to determine the boundary area, bounding box and centroid point for each hand region. Consequently, we select a search area in the next frame (Fig. 2(right side)) around the bounding box that is determined from the last frame in order to track the hand and reduce the gesture region of interest. Thereby, the new bounding box is calculated and the centroid point is determined. By iteration of this process, the motion trajectory of the hand so-called gesture path is generated from connecting hand centroid points (Fig. 3).

Fig. 2. Hand detection and Tracking (left: Original, right: Track result)

3. Feature extraction

Using the result of hand detection and tracking previous, we can do the next step to establish gesture coordinate system. In this study, we start tracking and recording dynamic trajectory after the hand detection was confirmed, then trajectory coming with the beginning of hand tracking. Similarly, when the hand motion stop, it means the ending of trajectory record. It is a important step to quantify hand motion trajectory for hand gesture recognition.

3. Estimating Principal Gesture Plane

For a nonlinear gesture, we have found that the trajectory of it is almost in a plane, which we call principal gesture plane [8]. To find the principal gesture plane is the most important step. The gesture coordinate system is established by the gesture trajectory and can be represented by principal gesture plane and its normal vector.

From different viewpoints we get different 3D trajectories while the projections of the trajectories on the principal gesture planes are almost the same. So we use Least Squares Method to estimate principal gesture plane. The principal plane is described in the form of $ax+by+cz+d=0$. The trajectory in 3D space is described as the follow system of homogeneous Equations 1:

$$\begin{cases} ax_1 + by_1 + cz_1 + d = 0 \\ ax_2 + by_2 + cz_2 + d = 0 \\ \quad\cdots \\ ax_n + by_n + cz_n + d = 0 \end{cases} \tag{1}$$

We could describe it in the way as: $Pv=0$ Here, a gesture is defined as a series of points $G(p1,p2,...,p)$ in Cartesian space, where

$$P_n = (x_n, y_n, z_n), \text{ and } v = (a,b,c,d)^T.$$

We can get suitable v by minimizing $E=|Pv|^2 = v^T P^T P v$. The $n*n$ matrix $P^T P$ is by construction symmetric positive semidefinite, and it can be diagonalized in an orthonormal basis of eigenvectors $e_i (i=1,...,n)$ associated with the eigenvalues, $0 \le w_1 \le w_2 \le \cdots \le w_n$ Thus we can write any unit vector as $v = u_1 e_1 + u_2 e_2 + \cdots + u_n e_n$ with $u_1^2 + u_2^2 + \cdots + u_n^2 = 1$. In particular,

$$\begin{aligned} &E(v) - E(e_1) \\ &= v^T P^T P v - e_1^T P^T P e_1 \\ &= u_1^2 u_1^2 + \cdots w_n^2 u_n^2 - w_1^2 \ge 0 \end{aligned} \tag{2}$$

It follows that the unit vector x minimizing E is the eigenvector e_1 associated with the minimum eigenvalue of $P^T P$, and the corresponding minimum value of E is w_1^2 [9].

We choose singular value decomposition to compute the eigenvectors and eigenvalues. Vector $e_1 = (a_1, b_1, c_1, d_1)^T$ is the solution minimizing E and the equation of principal gesture is $a_1 x + b_1 y + c_1 z + d = 0$. Vector $l = (a_1, b_1, c_1)$ is the normal vector of principal gesture plane.

3.2 Hand Trajectory Projection method

The next step after getting the normal vector of the principal gesture plane, we should project the gesture trajectory onto the plane. As frontal view gestures are chosen as training sets, we rotate side principal gesture plane parallel to the frontal ones firstly. We can get the projection result through calculate and the inference reached as follow [8]:

$$p_i^{'} = (x_i^{'}, y_i^{'}) = p_i \times R \times P_r$$
$$= (x_i, y_i, z_i) \times R \times P_r \qquad (3)$$

Here, R is the rotation matrix between side view gesture coordinate system and frontal view gesture coordinate system. We calculate R from R^{-1}, and P_r is the projection matrix,

So we can use the projection result $P_i^{'}$ to simplify the gesture coordinate in 3D space, and then the following recognition part is based on 2D trajectory $G^{'}(P_1^{'}, P_2^{'}, ..., P_n^{'})$.

3.3 Feature extraction

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 [5, 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.

4. Proposed dynamic hand gesture recognition algorithm

The Hidden Markov Model has been widely used in modeling spatiotemporal applications such as speech recognition and motion analysis. There are three types of HMM: fully-connected, left-right and left-right banded.

It's well known that Hidden Markov Model model has strong ability for temporal data modeling, so we apply left-right banded Hidden Markov Model to model gesture trajectory.

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 Fuzzy Neural Network model cannot model temporal data and conventional Hidden Markov Model do not own ability for fuzzy inference, we integrate the two models together to represent complex gesture trajectory and perform inference by the integrated Hidden Markov Model-Fuzzy Neural Network model based on [5,6], for the recognition of dynamic gesture.

Hidden Markov Model-Fuzzy Neural Network model includes five layers. Its first layer, second layer and Hidden Markov Model layer constitute the fuzzy preprocessing part, third layer and fourth layer constitute fuzzy inference part, fifth layer is the defuzzification part of Hidden Markov Model-Fuzzy Neural Network 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 two movement components of dynamic gesture, i.e. QP, QT, respectively.

The second layer and Hidden Markov Model layer compose fuzzification layer. Each Hidden Markov Model 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 Hidden Markov Model, i.e. $p(Q/\lambda)$ is considered as membership value of the corresponding fuzzy class variable.

The third layer is the layer of fuzzy inference, and each neuron represents a fuzzy rule. An exam-

ple of fuzzy rule like this one: As we defined eight direction codes, we can get the first code is "6"(condition s_1), the second code is "1"(condition s_2), then the third code is "3" (condition s_3), the fourth code is "3" (condition s_4), and the fifth code is "7" (condition s_5), the last code is "7" (condition s_6), so the meaning of this type of dynamic gesture is "4." In other words, $s_1 \cap s_2 \cap s_3 \cap s_4 \cap s_5 \cap s_6 \rightarrow S$. The most complex fuzzy rule is for number "8", it has eight conditions. The connecting weights between neurons in second and third layer imply the contribution degree of the antecedent part for this rule.

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.

The fifth layer is the defuzzification layer, the output of which is shown as Equation 4.[11]

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

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

Based on the straightforward structure of Left-Right banded model [12], We choose it as the type of Hidden Markov Model. We employ Gaussian Mixture Model (GMM) as the emission probability of observation, which has the likelihood as described in Equation 5.

$$P\left(\frac{O}{\lambda}\right) = \sum_{i=1}^{M} \omega_i g_i(x) \qquad (5)$$

Where ω_i is the weigh of ith Gaussian component, $g_i(x)$ is one-dimensional Gaussian Function: $g_i(x) \sim N(\mu, \sigma^2)$.

Training of proposed improved FNN model includes two parts: Firstly, the training of HMM model, it is the re-estimation of parameters in state transition matrix, output probability matrix or GMM's expectation and variance [13]. Secondly, the weights of FNN will be tained after HMM training. Back-Propagation (BP) algorithm is chosen for the training of connecting weights. When it reaches maximum iteration number or converge, the training process will stop.

5. Experimental Results

The gesture trajectory recognition system operated by hand gesture is running on the hardware environment of Intel(R) Core(TM) 2 (2.93GHz), a Web camera, and the software environment of Windows 7 and Visual Studio 2008 using OpenCV. Our proposed system showed good results to recognize the numbers in real-time from color image sequences.

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 clips. As a result, there are totally 250 training samples for all 10 kinds of gestures, 150 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. Orientation dynamic features are obtained from spatio-temporal trajectories and then quantized to generate its code words. The algorithm we present have the best performance and achieves average rate recognition 95.6% (as show in Table 1).

In this proposed method, high-dimensional gesture feature is transformed into several low-dimensional features, so computational complexity is reduced.

Table 1. Number recognition rate

Number	0	1	2	3	4
Recognition Rate (%)	96.6	96.5	95.4	95.3	95.8
Number	5	6	7	8	9
Recognition Rate (%)	95.3	95.7	96.2	95.1	95

6. Conclusion

In this paper, we propose a dynamic human activity algorithm to recognizes hand trajectory gestures. At first, a skin-color region detection method for the color image with complex background is presented, which is a mixed skin-color segmentation model in both YCbCr and HSI color space constructed. Then we assume the gesture trajectory is almost in a plane that called principal gesture plane, after that the Least Squares Method is used to estimate the plane and project the 3-D trajectory model onto the principal. Next, an improved FNN model is proposed for gesture recognition, which combines ability of Hidden Markov Model model for temporal data modeling with that of fuzzy neural network for fuzzy rule modeling and fuzzy inference.

In this proposed method, high-dimensional gesture feature is transformed into several low-dimensional features, so computational complexity is reduced. Orientation dynamic features are obtained from spatio-temporal trajectories and then quantized to generate its code words.

REFERENCE

[1] 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.

[2] N. Liu, B. Lovel, and P. Kootsookos, "Evaluation of Hmm Training Algorithms for Letter Hand Gesture Recognition," *IEEE Int'l Symp. on Signal Processing and Information Technology*, pp. 648-651, 2003.

[3] E. Hunter, J. Schlenzig, and R. Jain, "Posture Estimation in Reduced-Model Gesture Input Systems," *International Workshop on Automatic Face-and Gesture- Recognition*, pp. 290-295, 1995.

[4] X. Deyou. "A Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG," *International Conference on Pattern Recognition*, pp. 519-522, 2006.

[5] 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,"*Journal of W.S.C.G.,* Vol.16, No.1, pp. 65-72, 2008.

[6] E. Holden, R. Owens, and G. Roy. "Hand Movement Classification Using Adaptive Fuzzy Expert System," *Journal* of *Expert System,,* Vol.9, No.4, pp. 465-480, 1996.

[7] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of the IEEE,* Vol.77, No.2, pp. 257-286, 1989.

[8] Jun Cheng, Pengcheng Li, Guang Chen, Can Xie, and Qi Xie, "View Invariant Hand Gesture Recognition using 3D Trajectory," *Proceeding of the World Congress on Intelligent Control and Automation,* pp. 6315-6320, 2010.

[9] Forsyth D. A. and Ponce J., *Computer Vision: A Modern Approach*, Prentice hall, New Jersey, USA, 2002.

[10] L. Nianjun, C. L. Brian, J. K. Peter, and A. D. Richard. "Model Structure Selection & Training Algorithms for a HMM Gesture Recognition System," *International Workshop on Frontiers in Handwriting Recognition*, pp. 100-106, 2004.

[11] Wenkai Xu and Eung-Joo Lee, "Continuous Gesture Trajectory Recognition System Based on Computer Vision," *An International Journal of Applied Mathematics & Information Sciences*, Vol. 6, No.2S, pp.339S-346S, 2012.

[12] Lawrence R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol.77, No.2, pp. 257-286, 1989.

[13] Wenkai Xu and Eung-Joo Lee, "Hand Gesture Recognition Using Improved Hidden Markov Models," *Journal of Korea Multimedia Society*, Vol.14, No.7, pp. 866-871, 2011.

Wenkai Xu

received his B. S. at Dalian Polytechnic University in China (2006-2010). From 2010 to now he is a M. S student of Tong-myong University, Korea. His main research tops are image processing, computer vision, bio-metrics and hand recognition.

Eung-Joo Lee

received his B. S. , M. S. and Ph. D. in Electronic Engineering from Kyungpook National Uni-versity, 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 pro-fessor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests includes biometrics, image processing, and computer vision.