Send Orders for Reprints to reprints@benthamscience.ae

The Open Cybernetics & Systemics Journal, 2015, *9,* 217-221

Open Access

A Method of Gesture Recognition Based on the Improved Hidden Markov Model

Fu Yan and Ren Li*

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, P.R. China

Abstract: Because the traditional HMM algorithm has three disadvantages: firstly, the output probability of observed features is irrelevant to its history; secondly, continuous multiplication of the probability values can be easy to cause underflow phenomenon in the Viterbi algorithm; thirdly, the observed values of high dimensional vector will bring about a larger computational burden in the training stage, so a new improved HMM algorithm was proposed. At first, we should separate hands from complex backgrounds by using the deep message of kinect, and reduce the dimensionality of the observed value. Next, we use the angel of adjacent point as trajectory feature of gesture and utilize curvature's changing of trajectory as the new HMM Model state numbers. Finally, the improved HMM algorithm is used to train and recognize the gesture. Results show that this method of the improved Hidden Markov Model has a low complexity, high efficiency and accuracy of recognition, which also has a good practicability.

Keywords: Dynamic gesture recognition, improved hidden markov model, the kinect sensor.

1. INTRODUCTION

Human-computer interaction is no longer confined to the keyboard and mouse input, handle, or only touch devices, but the research on human-computer interaction technology interactive suits people's communication habits [1-3]. These researches include face recognition, expressional recognition, lip language, head motion tracking, gesture recognition and pose recognition etc. Gesture is a natural and intuitionistic communication mode, however, as the gesture possesses the diversity, ambiguity, and differences in the space and time, so it is very interesting and challenging to study this direction. Traditional method based on visual gesture recognition contains many algorithms, for example, K. Grobel and M. Assam extracted features from the video by using the technology of HMM to recognise 262 isolated words, and the correct rate is 91.3% [4]. Vogler and Metaxas who combined the data glove with visual gesture recognition adopted a position tracking and introduce three mutually vertical cameras as input devices to recognise the 53 isolated words, the results of recognition rate is 89.9% [5]. Christopher Lee and Xu from CMU used data glove as input equipment, they designed a gesture control system about operating robot [6]. However, the dynamic gesture recognition above based on vision is easily affected by illumination, complex background, the complexity of the algorithm and other factors, so the recognition rate and real-time is limited. I think the main problem of HMM algorithm is the much time consumption when we use the Baum-Welch algorithm to train the training set, and it brings larger computational burden when we use the forward and backward algorithm to evaluate, therefore, we find it difficult to dedicate enough attention to the training efficiency and control precision of the model. This paper focuses on the study of an improved Hidden Markov Model algorithm, it enhances anti-interference ability, reduces the complexity of the algorithm and improves the stability effectively and the robustness of the dynamic gesture recognition when we recognise the dynamic gesture recognition by using the improved Hidden Markov Model algorithm.

2. MAKING USE OF THE KINECT SENSOR TO OBTAIN HAND INFORMATION

2.1. The Introduction of Kinect

Kinect is a motion sensing input device. Kinect sensor interacts the application with software library provided by Kinect For Windows SDK Beta. Kinect For Windows SDK Beta consists of two important API, NUI API (natural user interface) with Audio API respectively. Kinect provides a series of novel and powerful technologies, such as depth sensing, skeletal tracking, speech and body recognition.

NUI API is the core of kinect API, which extracts data from the image sensor, and controls the kinect equipment, and the main function that it provides the access interface of the kinect sensor element connected to the PC. It also offers the image which generated by the kinect imaging sensor and access interface of depth date stream and image which is transmitted and processed. It is used to support the skeleton tracking technology. As shown in Fig. (**1**). Kinect technology is used widely, such as virtual application, 3D Modelling, mechanical controlling, virtual instrument, virtual entertainment, computer related applications, virtual experimenting, game controlling, health training.

2.2. Segmentation of Gesture

The segmentation algorithm of gesture is divided into three categories: the segmentation method based on histo-

1874-110X/15

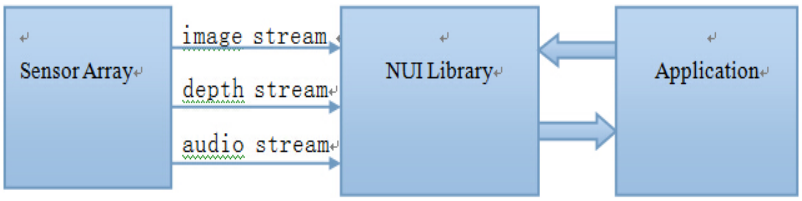

Fig. (1). Kinect interaction.

gram, the segmentation method based on local region information and the segmentation method based on colour or some physical characteristics.

This paper makes use of skeletal identification function provided by Kinect SDK to separate hands from complex background environment. The specific algorithm is as follows: firstly we input deep images and kinect assess the pictures by "pixel", then output target contour; secondly, we scan pixel of deep images point by point in these regions and judge which part belongs to human body. This process includes detecting edge, processing noise threshold and classing the character human target etc; thirdly, the machine learn 32 different parts of human body, then we take different colours to each adjacent part of body. Considering about overlap between different parts of body, we must analyse from the front, profile, and up side by every pixel and judge where is the node; Finally, we confirm the position of joint. It is an evaluation and probability matching process. The main measure is to scan pixel by pixel from the parts to the whole.

2.3. Obtain Information of Gesture Motion's Trajectory and Feature Extraction

Action recognition has a attribute of duration. Gesture is a part of the human body with a continuous action sequence in the space, such as through waving, we can obtain certain feedback information about gesture's speed and direction. Hands glide in the space which could control speed and direction of screen's roll.

In gesture recognition, gesture's translation variation, rotation, size and angle velocity are used as feature vector. In order to avoid the complexity of problem, we decompose complex gestures, 3D dynamic gesture's change of motion in Z axis and a change in image's feature of projection plane according to the feature. The image's change in the projection plane could be further decomposed into hand's change position and shape which were in 2D plane. According to the action recognition with an attribute of duration, we put forward a method of tracking the hand's trajectory in the space and measuring its line speed. The concrete steps as follows: first, we make a judgment on gesture's start and end through the speed of hands, and define the speed threshold. That the initial velocity was 0.02m/s, over speed is 0.02m/s, and the time threshold is 0.24s. As the deep information of the frame rate is 30fps, so it can last up to 8 frames. The advantage of this method is its simple judging condition, which only need to record the velocity at any point. And the disadvantage is the gesture detection can't be stopped and sometimes the hand's stop and start make the recorded gesture incorrect. The most important in gesture recognition is to judge condition of gesture's motion, on the basis of this, for avoiding the cycle start and cycle end of gesture when moving gesture, this paper proposes a method that combines with static gesture to mark the end point and start point. Combining dynamic with static increases the accuracy of program.

This paper processes gesture trajectory by using one-dimensional discrete HHM. The discrete model needs vector quantization for the continuous series of time which will result in some loss of information. Plane trajectory of gesture is composed of a series of location points and its characteristic value is showed as a set of discrete coordinate location. This paper uses the angel of adjacent point as trajectory feature of gesture. In order to get the position of two points by the kinect sensor, we should calculate the angel of two points and disperse the angel 8 directions, and then we can obtain the discrete value about motion direction of the two points.

3. GESTURE TRACK RECOGNITION BASED ON IMPROVED HIDDEN MARKOV MODEL

3.1. The Improved HMM

The Hidden Markov Model is described by a five elements: $\lambda = (N, M, A, B, \pi)$. N: a finite set of states. M: a finite set of observations, A: the state transition probability matrix, B: the observed value probability distribution matrix, π: the initial state probability distribution.

The calculation formulas for forward algorithm are as follows:

$$\alpha_1(i) = \pi_i b_1 \tag{1}$$

$$\alpha_{t+1}(i) = [\sum_{t=1}^{N} \alpha_t(i)\alpha_{ij}]b_j O_{t+1} \tag{2}$$

$$P(o \mid \lambda) = \sum_{j=1}^{N} \alpha_T(j) \tag{3}$$

The calculation formulas for backward algorithm are as follows:

$$\beta_T(i) = 1 \tag{4}$$

$$\beta_T(i) = [\sum_{t=1}^{N} \alpha_{ij} b_j(O_{t+1})]\beta_{t+1}(j) \tag{5}$$

$$P(O \mid \lambda) = \sum_{i=1}^{N} \beta_t(i) \tag{6}$$

The calculation formulas for Viterbi algorithm are as follows:

$$\delta(i) = \pi_i b_i(O_1) \tag{7}$$

$$\varphi(i) = 0 \tag{8}$$

$$\delta_t(i) = \max_{1 \le i \le N}[\delta_{t-1}(i)\alpha_{ij}]b_j(O_t) \tag{9}$$

$$\varphi_t(j) = \max_{1 \le i \le N}[\delta_{t-1}(i)\alpha_{ij}] \tag{10}$$

$$p^* = \max_{1\le i\le N}[\delta_T(i)] \tag{11}$$

$$q_t^* = \max_{1\le i\le N}[\delta_T(i)] \tag{12}$$

$$q_t^* = \varphi_{t+1}(q_{t+1}^*) \tag{13}$$

Formula (13): obtaining the state sequence and getting the optimal state sequence.

The disadvantages of the HMM model are the state transfer hypothesis and hypothesis of the output value of observation. As the state transfer hypothesis only relate to the t time state and it has nothing to do with the previous state when state at time t transfer to at time t+1, in order to compensate for this shortcoming, this paper makes weighting-process in probability of output about the character observed of Hidden Markov Mode and gives different weights according to the frequency of the sequence of the characteristics, as shown in equation (14 and 15). Also, because the continuous multiplication for Viterbi algorithm of the probability values can easily lead to underflow phenomenon, so we take the logarithm of its probability value and then calculate it, as shown in equation (16). This method improves the efficiency of the learning model.

Due to the dynamic gesture recognition has a high-latitude feature, so HMM is good at timing modelling, but observed values with high dimensional vector will bring about a larger computational burden for training and learning model. Therefore, in order to improve the efficiency of the model, we depress the dimension of the observed value.

$$\alpha_{t+1}(j) = \sum_{k=1}^{N} W_k[\sum_{i=1}^{N}\alpha_t(i)\alpha_{ij}]b_j(O_{t+1}), \sum_{k=1}^{N} W_k = 1 \tag{14}$$

$$\beta_t(i) = \sum_{k=1}^{N} W_k[\sum_{j=1}^{N}\alpha_{ij}b_j(O_{t+1})]\beta_{t+1}(j), \sum_{k=1}^{N} W_k = 1 \tag{15}$$

Formula (14): W_k is the component weight.

$$\delta_t(i) = \max_{1\le i\le N}[\delta_{t-1}(i) + \log\alpha_{ij}] + \log b_j(O_t) \tag{16}$$

The HMM is trained by Baum-Welch algorithm. Baum-Welch algorithm is also known as the forward backward algorithm, and it is based on the EM (Expectation-Maximization) algorithm. EM algorithm is an iterative process which consists of two parts of "expectations (E process)" and "maximum likelihood estimation (M process)" by alternately. The condition is the provided incomplete data and the current parameter values. "E procedure" constructs the likelihood function of the data from the conditional expectation, and "M process" makes use of sufficient statistics of parameters to re-estimate the parameters of probability. The final results ensure the log likelihood of the training data to max out. The each iteration of EM algorithm must increase the log likelihood of the training data monotonically, so the iterative processes converge to a local optimal value gradually.

This process can be summarized as follows: firstly, giving a set of sample observation value sequence, then determining the mode $\lambda = (A, B, \pi)$ and ensuring the probability of observation sequence $P(O|\lambda)$ to max out at last. This process should be able to fulfil some optimization rules.

The algorithm steps:

a. Determining the parameters N and M and initializing $\lambda = (A, B, \pi)$.

b. giving the HMM and the observation sequence, then calculating $\xi_t(i,j)$, as shown in formula (17).

$$\xi_t(i,j) = [\alpha_t(i)\alpha_{ij}b_j(O_{i+1})\beta_{i+1}(j)]P(O|\lambda) \tag{17}$$

c. According to λ and $\xi_t(i,j)$, estimating a new set of parameters $\overline{\pi_1}$, $\overline{\alpha_{1j}}$, $\overline{b_{1j}}$, then constructing new model parameters $\overline{\lambda} = (\overline{A} + \overline{B} + \overline{\pi})$, as shown in formula (18-21).

$$\gamma_t(i) = \sum_{j=1}^{N}\varepsilon_t(i,j) \tag{18}$$

$$\overline{\pi_i} = \gamma_1(i) \tag{19}$$

$$\overline{\alpha_{ij}} = \sum_{t=1}^{T-1}\varepsilon_t(i.j) \Big/ \sum_{t=1}^{T-1}\gamma_t(i) \tag{20}$$

$$\overline{b_j(k)} = \sum_{t=1}^{T}\gamma_t(j)\times\delta(O_t, V_k) \Big/ \sum_{t=1}^{T}\gamma_t(j) \tag{21}$$

d. If $\log P(O|\overline{\lambda}) > \log P(O|\overline{\lambda})$ then go to step 2, until the $P(O|\lambda)$ converges, now the $\overline{\lambda}$ is the model parameters to be required.

3.2. The Training and Recognition Model

This paper utilizes curvature's changing of trajectory as the new HMM Model state numbers. If the curvature variation is greater than the number of threshold point, which is used as the trajectory HMM's state numbers.

We can reduce the complexity of the model effectively and improve the efficiency of identification by adjusting threshold size reasonably and controlling the state number in an appropriate range.

General process: first, we separate hands from complex backgrounds by using the deep message of kinect; second, we should get observation sequence and add it to the training database, and we train the training set by using Baum-Welch algorithm and get the certain HMM, and then recognise the gesture; Finally, we use the forward or backward algorithm to evaluate.

Fig. (**2**) A block diagram of dynamic gesture recognition system. Fig. (**3**). The detail flow chart of recognition system.

4. ANALYSING EXPERIMENTAL RESULTS

In order to verify the proposed method of gesture recognition, this paper defines 5 kinds of gesture: left, right, up, down and stop. We experiment 50 times on recognition of gesture by using the traditional HMM and respectively, and compare the recognition rate of the improved HMM with the traditional HMM. From the Table **1**, we can find the improved HMM increase recognition rate that has better recognition ability.

The selection of state number is related to complexity of trajectory, in theory, the more selection about state numbers,

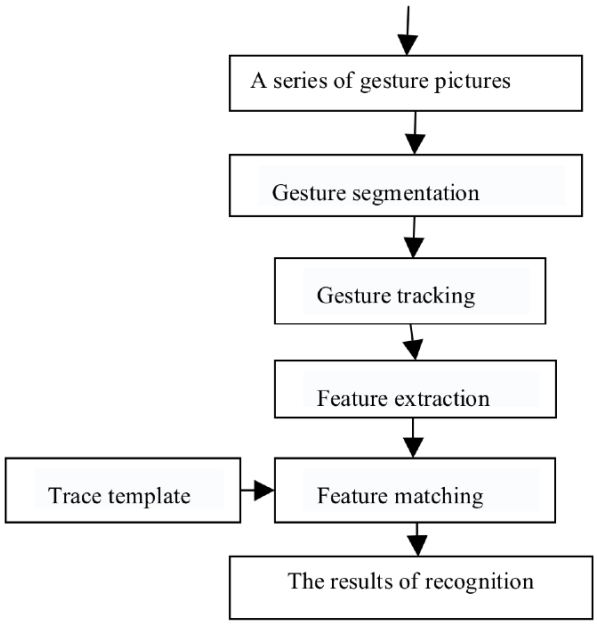

Fig. (2). Block diagram of dynamic gesture recognition system.

the better it is. But along with the increasing of state numbers, the augmentative computational complexity influences the real-time. Fig. (**4**) shows the complexity of different state number. As you can see from Fig. (**4**), the recognition is highest when the state is 4.

The improved HMM distinctly reduces the overhead of computation and training time, and this paper takes the case of left and stop to calculate its average processing time, time unit: ms. As you can see from Fig. (**5**). compared the traditional HMM with the improved HMM, the improved HMM decreases the training time and enhances the recognition efficiency and has better recognition ability.

5. CONCLUSION

According to the shortcomings of the traditional HMM algorithm, this paper proposes a new improved HMM

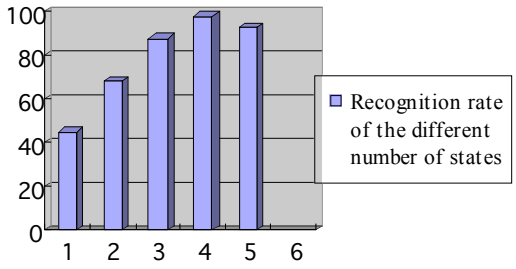

Fig. (4). The different recognition rate of state numbers.

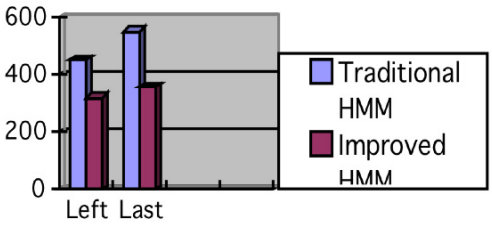

Fig. (5). Comparison of time-consuming between the improved HMM and traditional HMM.

algorithm. As for these questions, we weighted the probability of the observed value's output through the frequency of feature sequence. Then, we calculate the probability of logarithmic values. At last, the dimensionality of the observed value should be reduced. Compared the traditional HMM with the improved HMM, the results show that the improved algorithm depresses the computational burden and the complexity of the model, besides, it decreases the training time. On the whole, the improved HMM algorithm enhances the efficiency and the accuracy of gesture recognition to a certain extent.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

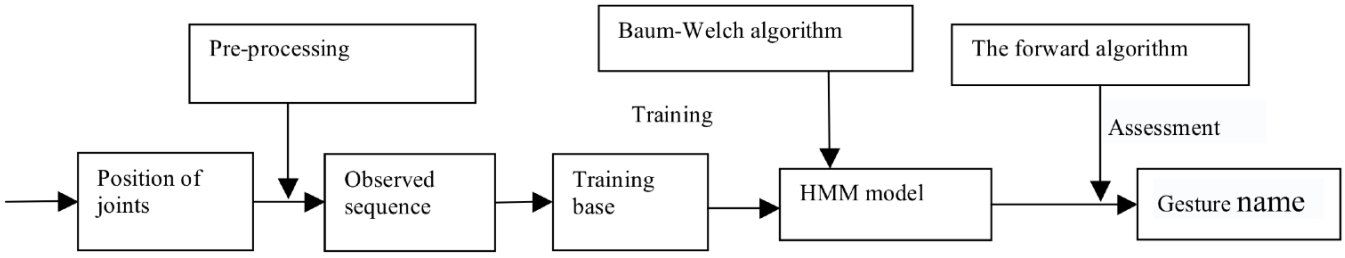

Fig. (3). The detail flow chart of gesture recognition.

Table 1. Comparison of recognition rate between the improved HMM and traditional HMM.

Gesture Track	Up	Down	Left	Right	Stop
Traditional HMM recognition rate (%)	90%	93%	93%	95%	94%
Improved HMM recognition rate (%)	95%	96%	97%	98%	96%

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] M. Fleming, and G. Cottrell. "Categorization of faces using unsupervised feature extraction," In: *International Joint Conference on Networks*, Neural Networks: San Diego, CA, USA, vol. 2, 1990, pp. 65-70.

[2] T. Kansde, "*Picture processing by computer complex and recognition of human faces*," Technical Report, Kyoto University. Department of Information Science, 1973.

[3] A.O. Toole, A.J. Mistilin, and A.J. Chitty, "A physical system approach to recognition memory for spatially transformed faces," *Neural Network*, vol. 1, no. 3, pp. 179-199, 1988.

[4] K. Grobel, and M. Assam, "Isolated sign language recognition using hidden Markov models," In: *Proceedings of the IEEE International Conference on Man and Cybernetics,* Orlando, 1997, pp. 162-167.

[5] C. Vogler, and D. Metaxas. "Adapting Hidden Markov Models for ASL recognition by using three dimensional computer vision methods," In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Orlando, 1997, pp. 156-161.

[6] C. Lee, and Y. Xu, "Online interactive learning of gestures interfaces," In: *Proceeding of IEEE International Conference on Robotics and Automation*, vol. 3, no. 1, 1996, pp. 30-42.

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014