Research on Algorithm and Model of Hand Gestures Recognition Based on HMM

Junhui Liu[1,2(✉)], Yun Liao[1,2], Zhenli He[1,2], and Yu Yang[1,2]

[1] School of Software, Yunnan University, Kunming 650091, Yunnan, China
HanksLau@gmail.com
[2] School of Informatics, Yunnan University of Finance and Economics, Kunming 650221, China

Abstract. Human computer interaction is one of the key points in the competition of information industry in the world, all countries in the world put the human-computer interaction as a key technology to study. Butler Lampson, ACM Turing Award winner in 1992 and Microsoft Research Institute chief software engineer pointed out that the computer has three functions in the "21st century computing research" report. The first is simulation; the second is that the computer can help people to communicate; the third is interaction, that is, to communicate with the real world. Human-computer interaction is an important field of computer research, and hand gestures recognition is a key technology in this field. The key of gesture recognition is the feature extraction and the establishment of hand recognition model. It can accurately identify the various kinds of deformation. HMM method has a flexible and efficient training and recognition algorithm, if the system needs to add a new gesture, just need to train the gesture of the sample set can be; If a gesture is not needed, just delete the corresponding HMM algorithm of the gesture, HMM has a strong expansion. Compared with DTW and other methods, HMM in speech recognition, gesture recognition, the recognition effect is better. In this paper, the HMM algorithm is used to identify the typical gestures, got very good recognition effect.

Keywords: Human-computer interaction · Hand gestures recognition · The feature extraction · HMM

1 Introduction

Since the birth of the first computer, how to effectively carry on the human-computer interaction has been the focus of the computer industry and the academic research. Human computer interaction (HCI) is research between the human and the computer through communication mutual understanding and communication, in the maximum extent for people to carry out the information management, service and processing functions, the computer will become the real people work and learning assistant of a science and technology. That enables computer technology to become an assistant to people's work and study. Human-computer interaction technology [1, 2] is a focus of competition in the current information industry, Human-computer interaction is an important field of computer research, and the vision based gesture recognition

J. Wan et al. (Eds.): CloudComp 2016, SPNCE 2016, LNICST 197, pp. 81–90, 2018.
https://doi.org/10.1007/978-3-319-69605-8_8

technology [3, 4] is a key technology in the field of human-computer interaction and has many typical applications, which has become a hot research topic at home and abroad.

In this paper, in the natural human-computer interaction system, the related research is carried out for the human hand gesture recognition, and the relevant algorithms are proposed to extract the feature of the gesture, and the dynamic gesture recognition model is established, to below will be related to this description.

2 Static Hand Gesture Feature Extraction

2.1 Hand Model

Hand model [5] is particularly important for human-computer interaction system based on finger interaction, and the selection hand model is closely related to the task to be processed by the human computer interaction system. Hand models can be very simple, we can use the image gradient direction histogram, on a few simple static hand gestures recognition. The hand model may also be very complex, such as the creation of a complex 3D gesture model as the main way of information interaction in virtual reality human-computer interaction system. 3D gesture model is able to accurately describe the complex state of the hand, at the same time, can identify the vast majority of the hand input information, however, to obtain accurate 3D hand model is not realistic through a single common camera. Although the approximate 3D model of the hand can be obtained to some extent, the type of hand model and the environment have special requirements, so it does not have the practical application value.

Figure 1 shows the 2D model of the hand, model A is composed of two parts of finger and palm. Only take the finger and palm connected joints into consideration, ignore the influence of each joint of the chiral internal finger. Therefore, in this model, the finger only exists, does not exist, and the finger points to several state information, which greatly reduces the complexity of the hand model, and emphasizes the importance of finger pointing. In this paper, the model is improved, and the model B in Fig. 1 is formed. Compared with model A, model B in the palm part in addition to the palm area P description, more focus on the palm of the center of gravity G. In the finger part, the model B emphasizes the position of F, and the position relationship between F and G.

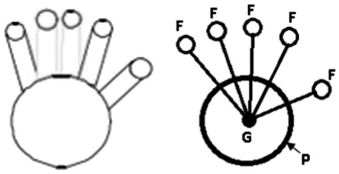

Fig. 1. The 2D model of the hand

According to the model B, The premise of obtaining the position of the fingertip is to find the center of gravity of the palm G. If the center of gravity of the G can be found through the fingertip and the location of the center of gravity G to determine the

location of the fingertip. Therefore find the palm center of gravity G is the use of the premise and the key of the model.

2.2 Palm Center of Gravity Extraction

The extraction of the palm center of gravity G is the premise and key of the algorithm. Only with relatively accurate the palm center of gravity is possible to obtain accurate fingertip position. The usual practice of obtaining the palm center of gravity is to calculate the center of gravity of the entire hand. And it is known that the approximate location of the center of gravity palm. This method is only applicable to situations that have no fingers or only one finger, as shown in Fig. 2(a). If five fingers are stretched out, the palm center of gravity position obtained by this method will seriously deviate from the true position. As shown in Fig. 2(b).

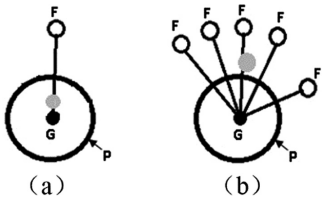

Fig. 2. Wrong position of center of gravity

By setting a search rectangle on the periphery of the hand, then calculate the distance between each pixel in the area of the hand to the search rectangle, and the distance between each pixel and the largest is the center of gravity of the palm. The method relies heavily on the search rectangle of the hand, and it is difficult to accurately locate the center of the palm.

In order to obtain the accurate palm center of gravity, the extended finger must be removed; otherwise the finger will have an impact on the calculation of the palm center of gravity. In this paper, a method of palm center of gravity extraction based on distance transform is proposed.

2.3 Palm Center of Gravity Searching Algorithm Based on Distance Transform

The distance transform of image is defined as a new image, the image pixel value of each output is set to input pixel 0 pixel in the recent distance. According to the different ways of calculating distance, there are two ways of distance transformation. Approximate template method and Euclidean distance method.

The approximate template method was first proposed by Rosenfeld. The basic idea is to use a template to calculate the distance between a point in the image and the last 0 pixels outside the image, each pixel values in the template is the approximate distances from the center of the template of Euclidean distance. As shown in Fig. 3.

4.5	4	3.5	3	3.5	4	4.5
4	3	2.5	2	2.5	3	4
3.5	2.5	1.5	1	1.5	2.5	3.5
3	2	1	0	1	2	3
3.5	2.5	1.5	1	1.5	2.5	3.5
4	3	2.5	2	2.5	3	4
4.5	4	3.5	3	3.5	4	4.5

Fig. 3. Approximate template method

Figure 4 is the result of the distance transform and binary conversion processing. Where the Fig. 4(a) is the original binary conversion image of palm, Fig. 4(b) is the palm part of binary conversion image. After two values of the distance image, not only filter the finger part, but also filter out the edge of the palm. However, because each direction filters edge pixels are basically the same, so this does not affect to the center of the palm computing. After obtaining the binary conversion image of the palm area. Through the 0- and 1-order image moments can calculate the palm center of gravity. The calculation formula is as follows.

$$MM_{00} = \sum_i \sum_j I(i,j) \tag{1}$$

$$MM_{10} = \sum_i \sum_j iI(i,j) \tag{2}$$

$$MM_{01} = \sum_i \sum_j jI(i,j) \tag{3}$$

$$i_c = \frac{MM_{10}}{MM_{00}}, j_c = \frac{MM_{01}}{MM_{00}} \tag{4}$$

where $I(i,j)$ represents the two value of the image coordinates of the pixel value of (i,j), (i_c,j_c) represents the image center of gravity. Palm center of gravity search algorithm based on distance transform described in detail as shown in Table 1.

(a) (b)

Fig. 4. Distance transformation and binary conversion processing result

Table 1. Palm center of gravity search algorithm based on distance transform

Step	Content
Step1	According to color and background color difference detection algorithm, get the binary conversion image of the palm
Step2	The distance transform is performed on the binary conversion image to obtain the result of distance transform
Step3	According to the results of the binary conversion distance transform, the palm area is obtained
Step4	According to: $i_c = \frac{MM_{10}}{MM_{00}}, j_c = \frac{MM_{01}}{MM_{00}}$ calculating the center of gravity of the binary conversion image in the palm region

Figure 5 is the center of gravity according to Table 1 based on distance transform algorithm. The white area in the figure is the area of the hand, the black spot in the white area is the center of the palm of the hand. Figure 5(a) is the center of gravity when the hand is clenched into a fist; Fig. 5(b) is the center of gravity when the index finger extended; Fig. 5(c) is the center of gravity when the five fingers extended. The results show that the palm center-of-gravity position can be measured accurately, and it was no relation with the state of whether to open the palm and how much the finger is opening in this algorithm. Then, the fingertips position can be queried based on position relationship between fingertips and the palm center-of-gravity.

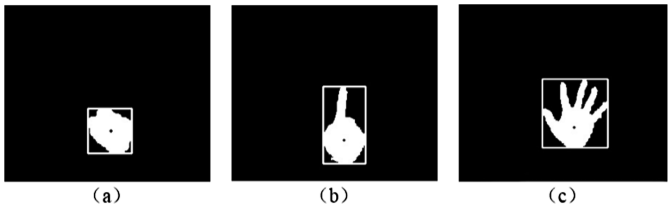

Fig. 5. Calculation results of palm center of gravity

3 Dynamic Gesture Feature Extraction

Dynamic gesture recognition technology [6–8] is mainly divided into three categories: a template-based technology, probabilistic techniques and techniques based on data classification techniques. There are techniques based on probability and statistics, such as hidden Markov model, Dynamic Bayesian network and Conditional Random Fields (CRFs) and other methods; Based on data classification technology with neural network, support vector machine (SVM) and Adaboosts, etc.

Template Matching is the simplest method of gesture recognition, by comparing the input gesture with the pre-stored Template similarity to recognize hand gestures. When dynamic gesture recognition template matching method in a complex background, this method cannot solve the problem of gesture difference in time and cannot accurately achieve real-time multi-gesture recognition. Training with the HMM method gestures

model [9–11], each gesture with a HMM model for training. The advantage of this method is to provide time scale deformation, can more accurately identify gestures of deformation, a flexible and efficient training and recognition algorithm. In this paper, HMM algorithm is adopted to recognize typical gestures, got very good recognition effect.

Dynamic posture gesture not only contains information about each point in time, as well as a gesture of trajectory information [12]. In human-computer interaction, the dynamic gesture can be more effectively and directly pass the user's intent. Gesture is a dynamic process of change in the posture of time sequence, involving time and space context, dynamic gesture recognition not only to eliminate the differences in space, but also to eliminate gestures duration differences.

Gesture trajectory [13] commonly used features such as location, direction angle and the rate of movement and so on. These three features were extracted and their corresponding recognition rates were compared and found that the trajectory identification, direction angle of the greatest contribution to the recognition rate. Direction angle between $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is defined as follows:

$$\varphi(P_1, P_2) = \begin{cases} \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) + \pi \ \textit{if} \ x_2 - x_1 < 0 \\ \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) + 2\pi \ \textit{elseif} \ y_2 - y_1 < 0 \\ \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \ \textit{otherwise} \end{cases} \tag{5}$$

For the purpose of this article preliminary identification of the letters I and J the 26 kinds of gestures, due to the characteristics of sequence similarity is higher, it is difficult to distinguish. In order to make a variety of gestures characteristic differences between the sequences as large as possible, and in order to better distinguish between recognition in full gesture detection below complete sequence and subsequence. In this paper, we choose the direction angle of the center point and the center point as the characteristic of the trajectory.

The direction angle using Eq. (6) is calculated:

$$\varphi 1_t = \varphi(P_c, \ P_t), \ (t = 1, \ 2, \ldots, \ T) \tag{6}$$

Among them, T represents the length of the hand gesture trajectory, $P_t(x_y, y_t)$ corresponding to the X axis and the Y axis coordinates of the T moments, which $P_c(x_c, y_c)$ represent the center point coordinates of the centroid of all gestures in a certain trajectory:

$$(X_c, Y_c) = \frac{1}{T}\left(\sum_{t=1}^{T} x_t, \sum_{t=1}^{T} y_t\right) \tag{7}$$

For the extraction of the direction angle, this paper uses 16 direction chain code to change $\varphi 1_t$ to be quantized to 16 levels, the quantization result is $\frac{\varphi 1_t * 16 + \pi}{2 * \pi} \% 16$, finally gets the discrete characteristic vector, as the HMM input.

4 Establish HMM Model

This paper defines the 26 gestures for identification Arabic Numbers from A to Z which are entered by users. Arabic Numbers from A to Z represent the 26 gestures. In each of gestures, 160 samples are collected. In experimenting, 80 samples are selected as training samples, and the remaining samples as test samples. 26 HMM models, from A-HMM to Z-HMM, will be established for the 26 gestures to test training samples and test samples, respectively.

HMM learning problem is the training process of the HMM model, which is a constantly re-evaluating process of model parameter through given sample observation sequences [14]. The parameters $\lambda = (A, B, \pi)$ of the HMM model will be constantly adjusted to train a most suitable model for sample set when $P(O|\lambda)$, probability of observation sequence O presence, reached its maximum. According to the definition of the forward variable $\alpha_t(i)$ and the backward variable $\beta_t(i)$, the $P(O|\lambda)$ can be easily calculated:

$$\mathrm{P(O|\lambda)} = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i).\alpha_{ij}.b_j(o_{t+1}).\beta_{t+1}(j),\ (1 \le t \le T-1) \tag{8}$$

The optimal model parameters λ^* are obtained when the $P(O|\lambda)$ reaches the maximum value.

Baulm-Welch algorithm is widely used to update the model parameters by repeated iteration calculation in the study of learning problems, and finally make the parameters gradually tend to the optimal value, which is a kind of maximum likelihood estimation process. The specific process of Baulm-Welch algorithm is as follows: First of all, the two variables used the algorithm are defined:

1. The posterior probability function

$$\gamma_t(i) = p(q_t = s_i|o, \lambda) \tag{9}$$

$\gamma_t(i)$ is probability of being state s_j at t moment given observation sequence O and parameters λ, and satisfied with $\sum_{i=1}^{N} \gamma_t(i) = 1$. $\gamma_t(i)$ can be expressed by the forward and backward variables in formula (10).

$$\gamma_t(i) = \frac{\alpha_t(i).\beta_t(i)}{p(o|\lambda)} = \frac{\alpha_t(i).\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i).\beta_t(i)},\ (1 \le i \le N) \tag{10}$$

Probability function:

$$\xi_i(i,j) = p\big(q_t = s_i, q_{t+1} = s_j|o, \lambda\big) = \frac{p\big(q_t = s_i, q_{t+1} = s_j|o, \lambda\big)}{p(o|\lambda)} \tag{11}$$

$\xi_i(i,j)$ is probability of being state S_j at t moment and t + 1 moment given observation sequence O and parameters λ, $\xi_i(i,j)$ can be represented as:

$$\xi_i(i,j) = \frac{\alpha_t(i).\alpha_{ij}.b_j(o_{t+1}).\beta_{t+1}(j)}{p(o|\lambda)} = \frac{\alpha_t(i).\alpha_{ij}.b_j(o_{t+1}).\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i).\alpha_{ij}.b_j(o_{t+1}).\beta_{t+1}(j)} \quad (12)$$

According to the meaning of $\gamma_t(i)$ and $\xi(i,j)$, both relationship is:

$$\gamma_1(i) = \sum_{j=1}^{N}\xi_t(i.j) \quad (13)$$

The specific parameters of the revaluation formula are as follows:

$$\overline{\pi}_i = P(q_1 = s_j) = \gamma_1(i) \quad (14)$$

$$\overline{a}_{ij} = \frac{\sum_{t-1}^{T-1}\xi_t(i.j)}{\sum_{t-1}^{T-1}\gamma_t(i.j)} \quad (15)$$

$$\overline{b}_j(k) = \frac{\sum_{t-1}^{T-1}\gamma_t(j)}{o_t = v_k}\Big/\sum_{t-1}^{T-1}\gamma_t(j) \quad (16)$$

The specific steps to obtain the optimal parameters λ^* of HMM are as follows:

(1) Initialize HMM parameters $\lambda = (A, B, \pi)$;
HMM initial model, $\lambda = (A, B, \pi)$, will affect the final recognition results to some extent, the initial value of each parameter is determined as follows:
   (a) Implicit state
   The number of hidden states of HMM is determined by the complexity of the corresponding gestures, because the recognition rate will be stable when the number of States increase to a certain value. However, If the number of States is overfull, computation amount in the recognition will be increased and it is easy to be over fitted.
   (b) State transition matrix
   The initial value of the state transition matrix A is determined by the following formula.

$$A = \begin{bmatrix} a_{ii} & 1-a_{ii} & & 0 \\ & a_{ii} & 1-a_{ii} & \\ & \cdots & \cdots & \\ & & a_{ii} & 1-a_{ii} \\ 0 & & & 1 \end{bmatrix} \quad (17)$$

   The value of the a_{ii} is related to the duration of the average of each hidden state:

$$a_{ii} = 1 - \frac{1}{d},\ d = \frac{\overline{T}}{N} \tag{18}$$

where T is average value of the length of all training samples for HMM corresponding certain gesture, namely, average sample length, N is the number of implicit state.

(c) Observation matrix
Assume the same appearing probability of each observation value of each state, the observation of the initial value of matrix B is determined by the following formula:

$$\mathrm{B} = \{b_{jk}\},\ b_{jk} = \frac{1}{M} \tag{19}$$

In formula (19), $j = 1, 2, \ldots, N, j = 1, 2, \ldots, M$. Moreover, as the result of the 16-direction chain code to code characteristic value, M is equal to 16.

(d) Initial state distribution
In this paper, we use the HMM model of the left and right band structure, and the initial state is the first state.

$$\pi = [10\ldots0]^T \tag{20}$$

(2) According to the observation sequence O and the model parameters λ, we can estimate new model parameters $\overline{\lambda}$. In other words, according to the revaluation formula respectively, $\overline{\pi}_i, \overline{a}_{ij}, \overline{b}_j(k)$ can be estimated. Now, a new model parameters $\overline{\lambda} = (\overline{\pi}, \overline{A}, \overline{B})$ are obtained.

(3) The probability $P(O|\lambda)$ and $P(O|\overline{\lambda})$ of observation sequence O in model λ and $\overline{\lambda}$ are calculated respectively by the forward backward algorithm. Then $|logp(O|\overline{\lambda}) - logp(O|\lambda)|$ can be calculated simply. If $|logp(O|\overline{\lambda}) - logp(O|\lambda)| < \varepsilon$, we can deduce that the $P(O|\lambda)$ must converge. Now, $\overline{\lambda}$ obtained by training is closest to HMM of gesture sample. On the contrary, if $|logp(O|\overline{\lambda}) - logp(O|\lambda)| \geq \varepsilon$, we can assume that λ is equal to $\overline{\lambda}$, and the algorithm continue to execute step (b) until $P(O|\overline{\lambda})$ is convergence.

5 Summary

Gesture recognition based on vision technology is the key technology in the natural human computer interaction system. In research of gesture recognition, we conducted a following work: the gesture model of static and dynamic characteristic analysis, distance transform algorithm based on improved the palm center of gravity calculation method, to hand dynamic potential is modeled using the advantages of hidden Markov model and the existing algorithm, and calculate the parameters of the model can learn. In this paper, the results of the study (which based on visual gesture recognition) have a better recognition effect on the more complex gestures.

Acknowledgment. This work is funded by Yunnan Enterprises Key Laboratory of Traffic Engineering Test Center (JTGC-2015-003).

References

1. Ma, C., Ren, L., Teng, D., Wang, H., Dai, G.: Ubiquitous human-computer interaction in cloud manufacturing. Comput. Integr. Manuf. Syst. **17**(03), 504–510 (2011)
2. Fang, Z., Wu, X., Ma, W.: The progress on the study of human computer interaction technology. Comput. Eng. Des. **19**(01), 57–63 (1998)
3. Gao, N.: Research on gesture recognition technology based on vision. Hebei University of Technology (2015)
4. ITU: ITU Internet reports 2005: the Internet of things. ITU, Geneva, Switzerland (2005)
5. Guan R., Xu, X., Luo, Y., Miao, J., Qiu, S.: A computer vision-based gesture detection and recognition technique. Comput. Appl. Softw. (01) (2013)
6. Marcus, A., van Dam, A.: User-interface developments for the nineties. IEEE Comput. **24** (9), 49–57 (1991)
7. Wang, J.: Integration of eye-gaze, voice and manual response in multimodel user interface. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, vol. 15 (1995)
8. Wang, L.: Dynamic gesture tracking and recognition and human computer interaction technology research. Xidian University (2014)
9. Wang, Q., Qi, X., Jiang, Y., Xu, W.: 3D handwriting recognition method based on hidden Markov models. Appl. Res. Comput. **09**, 099 (2012)
10. Fan, Z.: Parallel gesture recognition system based on depth learning in complex background. Xidian University (2014)
11. Qu, Q.Y.: The gesture recognition depth image and dexterous hand based interaction. Shanghai University (2014)
12. Wang, D.: Based on gesture recognition method of human-computer interaction system. Lanzhou University of Technology (2013)
13. Hu, W.: The research of human-computer interaction system based on gesture. Wuhan University of Technology (2010)
14. Tan, D.: Research on vision-based dynamic hand gesture recognition. Harbin Institute of Technology (2014)