Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 3739 – 3743
Procedia Engineering
www.elsevier.com/locate/procedia

Advanced in Control Engineeringand Information Science

A Human-Machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion

Chang-Yi Kao[a]* and Chin-Shyurng Fahn

[a]D9515011@mail.ntust.edu.tw, No.43, Sec.4, Keelung Rd., Taipei, 106, Taiwan, R.O.C

Abstract

We have developed an efficient mechanism for real-time hand gesture recognition based on the trajectory of hand motion and the hidden Markov models classifier. In our system, we divide our gestures into single or both hands, one hand have been defined four basic types of directive gesture such as moving upward, downward, leftward, rightward. Then, two hands have twenty-four kinds of combination gesture. However, we apply the most natural and simple way to define eight kinds gestures in our developed human-machine interaction control system so that the users can easily operate the robot. Experimental results reveal that the face tracking rate is more than 97% in general situations and over 94% when the face suffers from temporal occlusion. The efficiency of system execution is very satisfactory, and we are encouraged to commercialize the robot in the near future.

Keywords: hand detection, hand tracking, gesture recognition, hidden Markov model.

1. Introduction

The feature of easy control almost determines the competitive advantages of the product. Our gesture is one of the original and common communication methods in human society. Through the definition of visional identification hand gesture, people can convey their gestures to others. Based on this conception,

* Corresponding author. Tel.: +886986125010;
E-mail address: D9515011@mail.ntust.edu.tw.

1877-7058 © 2011 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2011.08.700

this paper namely applied the visional hand gesture identification to the HCI interface holds controlling usage.

The HCI interface generally includes: the human's face tracking, the hand tracking [1], the face recognition, the hand gesture recognition [2], the body pose recognition [3], and so on. Regarding the hand gesture recognition, it has also many applications such as: sign language recognition, using fingers as pointers for selecting options from menu and interacting with a computer by easy way for children [4]. Over the last few years, there are many methods for hand gesture recognition, have been proposed. Compared with the face recognition, hand gesture recognition also has different relative problems. Shan et al. [1] proposed a new tracking algorithm named mean shift embedded in a particle filter (MSEPF). This method integrates the advantages of the particle filter and mean shift. The MSEPF requires fewer particles to maintain multiple hypotheses than conventional particle filters do, so the computational cost is lower. This method also adopts the skin color model and motion information as its features. K. K. Kim et al. [5] proposed vision based gesture analysis for human-robot interaction that included detection of face and moving hand and gesture recognition. Their method for detection is based on skin color and motion information, and for recognition is used by neural network. M. Elmenzain et al. [6] have proposed a system to recognize the alphabets and numbers in real time based on hidden Markov models.

2. Hand Detection and Tracking

Our proposed face localization method can effectively find human face region. Consequently, we can accomplish the separation of face and hand regions. Then, we continuously process hand region localization (single hand or both hands) since the information of hands region rather than face region are what we want. In the following procedure, we will focus on hand region localization, including detection, tracking, and feature extraction. According our experience, the human body gesture is mostly related to both hands, especially on the palm of hand. So, we will limit the scope to the palm of hand, and its features become our primary interest. After the localization of face is determined, we propose a maximum circle plate mapping to locate the palm of hand from the remaining labeled skin regions.

2.1. Hand detection

In order to extract the images of hands (or palms), we proposed a new method: by mapping a circle plate in the candidate regions of hand and the center point of circle must be also skin color. The mapping is done by employing Bresenham's Midpoint circle scan-conversion algorithm. We can identify the regions that cover the palm. As soon as we detect the matched area in the circle, we add an initial position condition relative to the face to filter other noise regions before starting hand tracking. In the section of hand detection, we use an efficiency approach to process the localization. It is not necessary to fix the palm size beforehand.

2.2. Hand tracking

We apply the previous hand detection method on the fixed search window region for hand tracking. We can easily calculate the boundary of hand and determine the centroid point of hand region. Through iteration of hand tracking process, we can obtain the motion trajectory of the hand so-called gesture path from connecting hand centroid points set. When we obtained the hand location form hand tracking procedure, it will enter the state of check start point firstly. If the hand is no motion then it takes the next one, else this point is start point and the pure path is begin recorded. Then, it will stay at the state of path recording until the hand location is not moving. Finally, we will enter the state of check end point and obtain the pure gesture path for following recognition. In Fig 1, we outline the procedure of hand tracking.

```
void HandTrack(palmInfo MyHand, int radius, int value)
{
    blockInfo block; /* Set up fixed search block using MyHand */
    MyHand = findMAXCIRCLE (radius, block);
    if (MyHand.valid == true)  /* process the hand motion trajectory's recording */
}
```

Fig. 1. The hand tracking algorithm.

3. Hand Gesture Recognition

3.1. Feature extraction

Generally speaking, in order to indentify the image, we must study the image thoroughly so as to get its features. By putting these features into a classification, we can get good results. In this paper, we will take hand gesture as our major research object. Therefore, we will introduce a method to analyze the hand feature and the ways to extract them before recognition. There is no doubt that selecting good features to recognize the hand gesture path play a significant role in system performance. There are three basic features: location, orientation and velocity. From the research [7], they showed that the orientation feature is the best in term of accuracy results. And, our system is focusing on continuous hand motion of gesture. Hence, we will rely upon orientation as a main feature in our system. A gesture path is spatio-temporal pattern which consists of centroid points (x_{hand}, y_{hand}). So the orientation is determined between two consecutive points form hand gesture path by Equation 1.

$$\theta_t = \arctan\left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t}\right) \quad t = 1, 2, ..., T - 1 \tag{1}$$

where T represents the length of gesture path. The orientation is quantized by dividing it by $30°$ in order to generate the code words from 1 to 12. Thereby, the discrete vector is determined and then is used as input to HMM.

3.2. Gesture definition

Issuing commands to robot through only computer vision without sounds or other media is similar to conducting a marching band by way of visible gestures. In order to achieve real-time operations, our system requires simpler body language which is easy to recognize and differentiate from each other. Our system is mainly to recognize the dynamic hand gesture from continuous hand motion in real-time, and implement on interaction between human and robot. There are very many kinds gesture can be represented by the hand motion. In this system, we describe four types of directive gesture to one hand, which is moving upward, moving downward, and moving leftward and moving rightward separately, for the basic conducting gesture. Thus, if we add one or both hands into gesture invoking, we will have at most twenty-four kinds of meaningful gesture by the permutation combination of both hands. Here, we use a 2-D table to represent the all combination of gesture from both hands and classify each combination into a class which is named gesture's ID individually. By the way is easy to represent every gestures and it is convenient to add new hand gestures.

Table 1. Gestures' ID table

Left hand \ Right hand	No action	Upward	Downward	Leftward	Rightward
No action	0	1	2	3	4
Upward	5	6	7	8	9
Downward	10	11	12	13	14
Leftward	15	16	17	18	19
Rightward	20	21	22	23	24

However, because two hand's recognition is the same, we will only indicate the difference between right hand and left hand. Therefore, in the first item we will define the difference of the single hand gesture recognition for four basic directive gestures. In the second, we further combine left hand and right hand together. So we define the popular eight kinds of hand gestures: (1) the right hand is upward raising vertically; (2) raising both hands horizontally; (3) the left is leftward moving horizontally; (4) the right is rightward moving horizontally; (5) putting both hands on the front of breast; (6) both hands is upward raising simultaneously; (7) raising the right hand horizontally and the left hand up; (8) raising the left hand horizontally and the right hand up.

3.3. The recognition model

We can find the HMM is useful for calculating the probability, and the probability after the HMM compute is quite correct. Therefore, we take the HMM to recognize our hand gesture in this system. We use Baum-Welch algorithm for training the initialized parameters of HMM to provide the trained parameters, as above subsection mentioned. After training procedure, we use the trained parameters and discrete vector, hand motion trajectory, as input to Viterbi algorithm to obtain the best path. By the best path and gesture database, we can recognize the gesture path.

After the HMM parameters are initialized, we utilize Baum-Welch algorithm to perform the HMM training where the inputs of this algorithm are the initialized parameters and the discrete vector. Then, we will get the new HMM parameters from the training stage. In our system, our gesture database contains 20 sample videos, which 15 video sequences for training and 5 video for testing, for each isolated gesture, upward, downward, leftward, and rightward. And we take 10 sequences of discrete vector for each isolated gesture. While the training process for each video sequences is finished, we will take the discrete vector and the new HMM parameters as input for the Viterbi algorithm. Thereby, we can get the best gesture path which is corresponding to the maximal likelihood of four gestures HMM. Finally, we compare and choose the higher priority form the gestures database and output the result of recognition.

1. Initialize $\lambda = \{\pi, A, B\}$.
2. Compute $\alpha_t(i)$, $\beta_t(i)$ and $\gamma_t(i)$, and $\xi_t(i, j)$.
3. Re-estimate the model $\bar{\lambda} = \{\bar{\pi}, \bar{A}, \bar{B}\}$.
4. If $P(O \mid \bar{\lambda}) \geq P(O \mid \lambda)$, go to 2.

Fig. 2. The hand tracking algorithm.

4. Experience result

We define eight gestures from I to VIII: Forward, Backward, Turn left, Turn right, Stop, Following, Waiting, Ready, and the IX are the other gestures. Table 2 shows the average accuracy rate of recognizing each gesture using the three classifiers. We can observe that the average accuracy rate of recognizing Posture V is the highest in our recognition system. The average accuracy rates of recognizing Postures II, VI, and VIII are not desirable, whose common characteristic is only extending hands up without extending hands horizontally. We can summarize that the feature values of hands are not stable when extending them up. We must find other feature values to enhance the stability to solve this problem. Most of postures in the “Other” samples are to make both hands hang down naturally or cross around the torso and chest. Since the average accuracy rate of Posture I is not bad, we consider adding new kind postures in this way.

Table 2. The Average Accuracy Rate of Recognizing Each Kind of Hand Gestures Using the Hidden Markov Models.

Posture / Measurement	I	II	III	IV	V	VI	VII	VIII	IX
Average accuracy rate	98.96%	96.07%	98.53%	98.76%	99.07%	98.40%	97.83%	96.93%	97.83%

5. Conclusion

During tracking face, we can use fewer samples to maintain multiple hypotheses than conventional particle filters do, so that the computational cost is greatly decreased. Followed by hand region localization procedure, we propose a new method of mapping circle plates to locate hand. The mapping is done by employing Bressenham’s midpoint circle scan-conversion algorithm with second-order differences. By the way, we can clearly obtain the palm’s position. In the feature extraction procedure, we take the orientation between two consecutive points which are on motion trajectory of a single hand as main feature in our system. As for the outcome of average gesture recognition rate is more than 96%, which is very accuracy and satisfactory.

References

[1] C. Shan, Y. Wei, T. Tan, and F. Ojardias, “Real time hand tracking by combining particle filtering and mean shift,” in Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, South Korea, pp. 669-674, May 2004.

[2] X. Liu and K. Fujimura, “Hand gesture recognition using depth data,” in Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, pp. 529-534 May, 2004.

[3] H. Jiang, Z. N. Li, and M. S. Drew, “Human posture recognition with convex programming,” in Proceedings of the IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands, pp. 574-577, July 2005.

[4] S. Mitra and T. Acharya, “Gesture recognition: A survey,” in Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics, Kolkata, India, pp. 311-324, May 2007.

[5] K. K. Kim, K. C. Kwak and S. Y. Chi, “Gesture Analysis for Human-Robot Interaction,” International Congress on Anti Cancer Treatment, pp. 20-22, 2006.

[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,” The Journal of W S C G’08, vol. 16, no. 1, pp. 65-72, 2008.

[7] N. Liu, B. C. Lovell, P. J. Kootsookos, and R. I. A. Davis, “Model structure selection and training algorithm for a HMM gesture recognition system,” to appear in Proceedings of the International Workshop in Frontiers of Handwriting Recognition, Brisbane, Australia, pp. 100-106, 2004.