Hidden-Markov-Model-Based Hand Gesture Recognition Techniques Used for a Human-Robot Interaction System

Chin-Shyurng Fahn and Keng-Yu Chu

Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
Taipei, Taiwan 10607, Republic of China
csfahn@mail.ntust.edu.tw

Abstract. In this paper, we present part of a human-robot interaction system that recognizes meaningful gestures composed of continuous hand motions in real time based on hidden Markov models. This system acting as an interface is used for humans making various kinds of hand gestures to issue specific commands for conducting robots. To accomplish this, we define four basic types of directive gestures made by a single hand, which are moving upward, downward, leftward, and rightward individually. They serve as fundamental conducting gestures. Thus, if another hand is incorporated to making gestures, there are at most twenty-four kinds of compound gestures by the combination of the directive gestures using both hands. At present, we prescribe eight kinds of compound gestures employed in our developed human-robot interaction system, each of which is assigned a motion or functional control command, including moving forward, moving backward, turning left, turning right, stop, robot following, robot waiting, and ready, so that users can easily operate an autonomous robot. Experimental results reveal that our system can achieve an average gesture recognition rate of 96% at least. It is very satisfactory and encouraged.

Keywords: hand gesture recognition, hidden Markov model, human-robot interaction, directive gesture, compound gesture.

1 Introduction

In order to realize a human computer interface that possesses more human nature and is easy to operate, the researches focusing on hand gesture recognition have been rapidly grown up to make humans easier to communicate and interact. Nowadays, the control panel is gradually switched from a keyboard to a simple hand touch, which dramatically changes the communication style between humans and computers. Regarding the hand gesture recognition, it also has many applications such as combining face and hand tracking for sign language recognition [1], using fingers as pointers for selecting options from a menu [2], and interacting with a computer by an easy way for children [3]. Over the last few years, many methods for hand gesture recognition were proposed. Compared with face recognition, hand gesture recognition has different problems to be overcome. First of all, the main features of hand gestures

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 248–258, 2011.

are obtained from the fingers and palms of our hands. Each finger has three joints and the palm of a hand connects with the wrist. Therefore, the hand gestures can produce many great changes on the postures, especially for the different form of changes resulting from varied hand gestures. In addition, the hand gestures can have three-dimensional spatial revolution. Consequently, the hand gesture recognition may not be easier than the face recognition.

Vision-based gesture analysis has been studied for providing an alternative interaction between humans and computers or robots in recent year. In particular, the hand gesture recognition has become a major research topic for a natural human-robot interaction. Users generally exploit arms and hand gestures to give expression of their feelings and notification of their thoughts. And users also employ more simple hand gestures such as pointing gestures or command gestures rather than complex gestures. Kim et al. [4] proposed vision-based gesture analysis for human-robot interaction that includes the detection of human faces and moving hands as well as hand gesture recognition. Their method for detection is resorted to skin colors and motion information, and for recognition is to adopt neural networks. The main disadvantage is the user must wear a long-sleeves shirt that only the skin color of a palm is exposed. Our research is to conduct a robot directly by hand gestures without any auxiliary wearable or special equipment. To implement this, we intend to develop a human-robot interaction system to recognize some gestures defined by users via a PTZ camera capturing color image sequences in real time.

In this paper, part of the human-robot interaction system installed on an autonomous robot is presented. This system consists of four main processing stages: face detection and tracking, hand detection and tracking, feature extraction, and gesture recognition. Prior to the process of gesture recognition, we devise an automatic hand localization method based on skin colors and circle detection for finding palms, which is quite robust and reliable to complete hand detection and tracking in unconstrained environments; thus, the extraction of hand regions is very fast and accurately. Such a hand localization method may not be confined to uncover the lower arm, so users are not required to wear long-sleeves shirts. With the aid of hand localization to achieve hand detection and tracking, the orientation of a hand moving path between two consecutive points that stand for the positions of a palm appearing in a pair of image frames is extracted as an essential feature for gesture recognition. Subsequently, by virtue of the classification scheme adopting hidden Markov models (HMMs) that are widely applied to speech or handwriting recognition, we can effectively acquire the result of hand gesture recognition and issue the corresponding command to conduct the robot as expected.

2 Feature Extraction of Hand Gestures

2.1 Hand Detection and Tracking

To realize our human-robot interaction system that can automatically recognizes hand gestures consisting of continuous hand motions, hand localization (a single hand or two hands) is a crucial process because the information of hand regions rather than a face region are what we want. The hand localization process includes hand detection, tracking, and feature extraction. Due to the limitation of article length, the detailed description of the hand detection and tracking can refer to our past research [5].

2.2 Feature Extraction

What follows introduces a method to analyze the features of a hand gesture and the ways to extract them before recognition. There is no doubt that appropriate feature selection to classify a hand gesture plays a significant role of improving system performance. Herein, three basic features: location, orientation, and velocity are adopted. From the literature [6, 7], they showed that the orientation feature is the best one to attain high accurate recognition results. Hence, we will treat the orientation as a main feature used in our system to recognize a meaningful gesture composed of continuous hand motions constituting a hand motion trajectory.

The hand motion trajectory is a spatio-temporal pattern that can be represented with a sequence of the centroid points (x_{hand}, y_{hand}) of detected hand regions. Consequently, the orientation between two consecutive points in the hand motion trajectory is determined by Equation (1).

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

where T represents the length of a hand motion trajectory. The orientation is quantized into 12 levels, each of which is separated by 30° in order to generate twelve direction codewords called 1 to 12 as Figure 1 shows. Therefore, a discrete vector is in form of a series of direction codewords and then used as the input to an HMM.

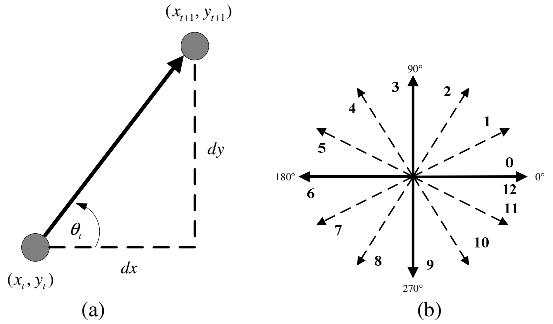


Fig. 1. The orientation and its quantization levels: (a) the orientation between two consecutive points; (b) the twelve quantization levels where direction codeword 0 is equivalent to direction codeword 12

3 Gesture Recognition

Gesture recognition is a challenging task for distinguishing a meaningful gesture from the others. This section depicts our human-robot interaction system that humans make various kinds of hand gestures to issue specific commands for directing robots. To achieve this, we define four basic types of directive gestures made by one hand, which are moving upward, downward, leftward, and rightward individually. They serve as fundamental conducting gestures. Thus, if another hand is incorporated to making gestures, there are at most twenty-four kinds of compound gestures by the combination of directive gestures using both hands. Finally, at the gesture recognition stage in our human-robot interaction system, we will apply a probabilistic approach, the HMM, to classify a hand gesture as shown in Figure 2.

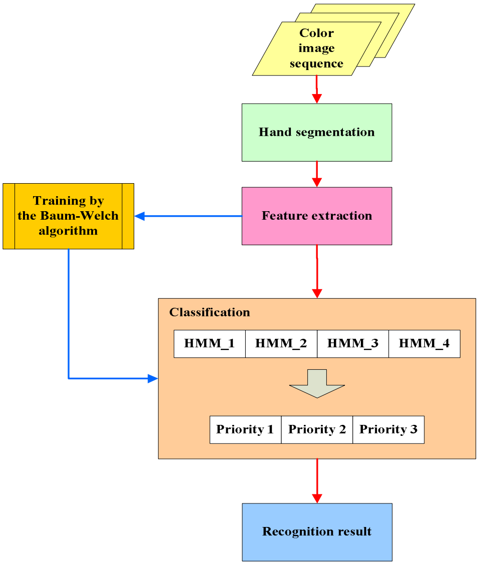


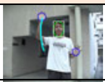
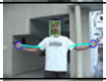
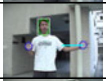
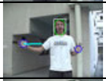
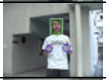
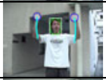
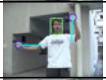
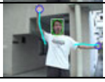
Fig. 2. The flow chart of the gesture recognition procedure

3.1 Gesture Definition

Issuing commands to direct robots through only vision without sounds or other senses is similar to conducting a marching band by means of visible gestures. In order to accomplish real-time operations, our system requires simpler body language which is easy to recognize and discriminate from each other. Very many kinds of daily used gestures are characterized by hand motions. In this system, four basic types of directive gestures: moving upward, downward, leftward, and rightward made by one hand are defined to serve as fundamental conducting gestures. By the combination of directive gestures using both hands simultaneously, we will have at most twenty-four kinds of compound gestures. For convenience' sake, a 2-D table is employed to express all the meaningful gestures, and each of them is named a gesture ID respectively. In this manner, it easily symbolizes every gesture and is convenient to add new hand gestures.

In essential, the gesture recognition techniques applied to single right or left hand motions are all the same. For practical use, we must continue to distinguish which one of the two hands is making gestures. At present, eight kinds of meaningful gestures are prescribed for directing robots. The gesture usage contains motion control and functional control. Each of them is assigned a command to conduct robots in our system, which includes moving forward, moving backward, turning left, turning right, stop, robot following, robot waiting, and ready. Table 1 illustrates the above gestures and lists their associated right and left hands features in terms of orientation sequences, respectively. Notice that in order to interact with an autonomous robot smoothly, we specially define one of the eight kinds of gestures, which both hands are put in front of the breast, as a static gesture to make stopping of any robot motion while the robot is operating. According to the features extracted from hand motion trajectories, we feed them to an HMM-based classifier for training and recognizing a given hand gesture seen from the human-robot interaction system.

Table 1. Corresponding Commands of Eight Kinds of Gestures

Command	Gesture	Left hand feature	Right hand feature
Moving forward		Orientation sequence = NULL	Orientation sequence = [10, 10, 10, 10, 10, 10]
Moving backward		Orientation sequence = [1, 1, 1, 1, 1, 1]	Orientation sequence = [7, 7, 7, 7, 7, 7]
Turning left		Orientation sequence = [1, 1, 1, 1, 1, 1]	Orientation sequence = NULL
Turning right		Orientation sequence = NULL	Orientation sequence = [7, 7, 7, 7, 7, 7]
Stop		Orientation sequence = NULL	Orientation sequence = NULL
Following		Orientation sequence = [10, 10, 10, 10, 10, 10]	Orientation sequence = [10, 10, 10, 10, 10, 10]
Waiting		Orientation sequence = [10, 10, 10, 10, 10, 10]	Orientation sequence = [7, 7, 7, 7, 7, 7]
Ready		Orientation sequence = [1, 1, 1, 1, 1, 1]	Orientation sequence = [10, 10, 10, 10, 10, 10]

3.2 The Hidden Markov Model

HMMs are chosen to classify the gestures, and their parameters are learned from the training data. Based on the most likely performance criterion, the gestures can be recognized by evaluating the trained HMMs. However, the HMM is different from the Markov model. The letter is a model with each state corresponding to an observable event and its state transition probability depends on both the current state and predecessor state. And extending from the Markov model, the HMM considers more conditions of the observable event, so it has been excessively used in various fields of applications such as speech recognition [8] and handwriting recognition [9]. Because the HMM is more feasible than the Markov model, we adopt the former to learn and recognize the continuous hand gestures to direct robots.

3.3 Fundamentals of HMMs

An HMM consists of a number of states, each of which is assigned a probability of transition form one state to another state. Additionally, each state at any time depends only on the state at the preceding time. In an HMM, one state is described by two sets of probabilities: one is composed of transition probabilities, and the other is of either discrete output probability distributions or continuous output probability density functions. However, the states of an HMM are not directly observable, but they can be observed through a sequence of observation symbols. Herein, we bring in the formal definition of an HMM [10] which is characterized by three matrices: the state transition probability matrix *A*, symbol output probability matrix *B*, and initial state probability matrix π. They are all determined during the training process and simply expressed in a set of $\lambda = \{\pi, A, B\}$. Figure 3 illustrates a state transition diagram of an HMM with three hidden states and three observations.

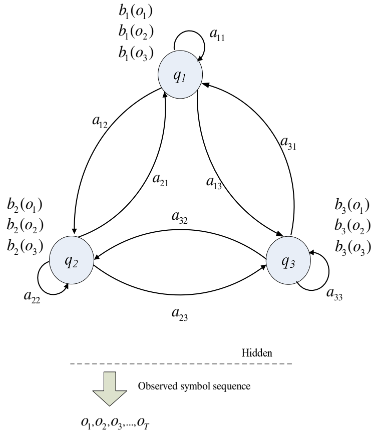


Fig. 3. Illustration of an HMM with three hidden states and three observations

Three states are drawn as circles in this example. Besides, each directed line is a transition from one state to another, where the transition probability from state q_i to state q_j is indicated by a_{ij}. Note that there are also transition paths from states to themselves. These paths provide the HMM with time-scale invariabilities because they allow the HMM to stay in the same state for any duration. Each state of the HMM stochastically outputs an observation symbol. In state q_i, for instance, symbol o_k is the output with a probability of $b_i(o_k)$ at time step *k*. During a period, the state yields a symbol sequence $O = \{o_1, o_2, \ldots, o_T\}$ from time step 1 to *T*. Thus, we can observe the symbol sequences output by the HMM, but we are unable to observe its states. In addition to this, the initial state of the HMM is determined by the initial state probability $\pi = \{\pi_i\}$, $1 \le i \le N$.

In general, three tasks: evaluation, decoding, and training should be accomplished under the framework of an HMM. In practice, they can be commonly solved by the forward-backward algorithm, Viterbi algorithm, and Baum-Welch algorithm, respectively [10].

4 Gesture Recognition

At the gesture recognition stage, the Baum-Welch algorithm is used for training the initialized parameters of an HMM to provide the trained parameters. After the training procedure, both the trained parameters and the discrete vector derived from a hand motion trajectory are fed to the Viterbi algorithm to obtain the best hand moving path. Using this best path and a gesture database, the robot can recognize a given hand gesture made by users. The number of states in our realized HMM is based on the complexity of each hand gesture and is determined by mapping each straight-line segment into one state of the HMM as graphically shown in Figure 4. There are four basic types of directive gestures for interacting with robots, including moving upward, downward, leftward, and rightward individually. In addition, we convert a hand motion trajectory into an orientation sequence of codewords from 1 to 12 acquired at the feature extraction stage depicted in Section 2.2. In the HMM, the number of hidden states is set to 1 and the number of observation symbols is fixed to12.

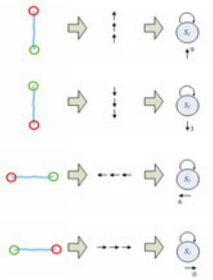


Fig. 4. Each straight-line segment corresponding to a type of directive gestures for establishing the topology of an HMM

Furthermore, a good parameter initialization of $\lambda=\{\pi,A,B\}$ for an HMM will produce better results. First, we determine the initial matrix A is $A=\{a_{11}=1\}$ since for all four types of directive gestures in our system only contain one straight-line segment, each of which needs one state. Second, the parameter matrix B is evaluated by Equation (2). Because the states of an HMM are discrete, all elements of matrix B can be initialized with the same value for each of different states.

$$B=\{b_{im}\} \text{ with } b_{im}=\frac{1}{M} \tag{2}$$

where i is the serial number of states, m is the serial number of observation symbols, and M is the total number of observation symbols. Finally, the initial state probability π_i is defaulted by 1.

Once the parameters of an HMM are initialized, we utilize the Baum-Welch algorithm to perform the training process where the inputs of this algorithm are the initialized parameters and the discrete vector that is obtained from the feature extraction stage. Consequently, we will acquire the new parameters from such a training phase. In our system, the gesture database contains 20 sample videos, 15 of which are used for training and 5 for testing, including the four types of directive gestures. And we take 10 discrete vectors for each type of directive gestures. While the training process for each video is finished, we will take the discrete vector and the new parameters of the HMM as the inputs for the Viterbi algorithm. Therefore, we can get the best hand moving path that is corresponding to the maximal likelihood of the four types of directive gestures. In the sequel, we compare and choose the higher priority form the gesture database, then output the result of recognition.

5 Experimental Results

5.1 Tests on the Gesture Recognition

Figure 5 shows some training samples of hand gestures which our system can recognize. For a convenience, Roman numerals (I to IX) are assigned to symbolize each kind of gestures. We can understand the reason why the accuracy rates of recognizing Gestures I, III, IV, and V are higher than those of recognizing other kinds of gestures. Especially for Gesture V, its accuracy rate of recognition is the highest, because the feature values of putting both hands in front of the breast are easier to discriminate from each other than those of stretching two hands simultaneously. The accuracy rate of recognizing Gesture II is not desirable, on account of the influence of luminance and different motion speeds with both hands. Since we make all training samples not only out of doors but also within doors, sometimes the hands are too close to a fluorescent light when we raise them up.

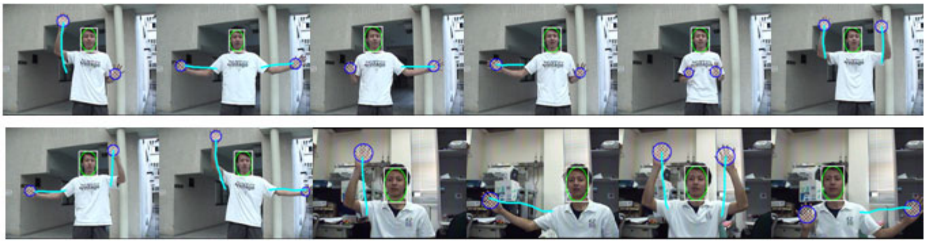


Fig. 5. Some training samples of gestures which our system can recognize

After calculating the necessary terms, we apply the recall and precision rates to measure the performance of our HMM-based hand gesture classifier. Table 2 lists the precision and recall rates of the experimental results. We can see that the recall rate measures how the proportion of a certain kind of subjects is classified into the correct kind of gestures, whereas the precision rate responds to the result of the

Table 2. The Recall and Precision Rates of the HMM-Based Hand Gesture Classifier

System performance Kind of gestures	Hidden Markov Models	
	Recall rate	Precision rate
Moving forward (I)	99.8%	97.1%
Moving backward (II)	97.0%	99.8%
Turning left (III)	99.4%	97.6%
Turning right (IV)	99.6%	97.8%
Stop (V)	100%	100%
Following (VI)	98.2%	100%
Waiting (VII)	98.8%	100%
Ready (VIII)	97.8%	100%
Others (IX)	98.8%	97.1%
Average	**98.8%**	**98.8%**

misclassification of some kinds of gestures. In other words, the higher precision rate means that the other kinds of gestures are rarely misclassified to the kind of the target gestures.

Table 3 shows the average accuracy rate of recognizing each kind of gestures using the HMM-based classifier. We can observe that the average accuracy rate of recognizing Gesture V is the highest by use of this classifier. The average accuracy rates of recognizing Gestures VI, VII, and VIII are not desirable, whose common characteristic is to raise one or two hands up. It can be inferred that the feature values of hands are not stable when they are raised up. We must find other feature values to enhance the stability to solve this problem. Most of gestures grouped into the kind of “Others” are to make both hands hang down naturally or cross around the torso and chest. Since the average accuracy rate of recognizing Gesture I is not bad, we consider adding new kinds of gestures in a similar way.

Table 3. The Average Accuracy Rate of Recognizing Each Kind of Gestures

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

5.2 Tests on the Human-Robot Interaction System

In this experiment, we equip our HMM-based hand gesture classifier on an autonomous robot actually. Through the eight kinds of gestures defined by us using both hands, we can interact with the robot easily. Such a kind of gestures is corresponding to a motion or functional control command; for example, we can raise the right hand up vertically, extend both hands horizontally, spread the left hand horizontally, and spread the right hand horizontally to conduct the robot moving forward, backward, leftward, and rightward, respectively. And, we can easily stop the current action of the robot by putting both hands in front of the breast. Besides, we devise another interaction mode that the robot follows a user. When the user raises his/her both hands upward simultaneously, the robot will begin to follow the user till

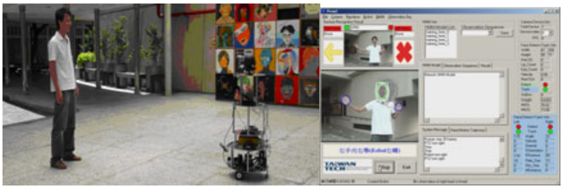


Fig. 6. A real situation of a user interacting with the robot

he/she gives the robot a waiting command, spreading the right hand horizontally and raising the left hand up. If the user tries to stop the robot following action and return to a ready status, he/she simply spreads the left hand horizontally and raises the right hand up. Figure 6 demonstrates the user standing in front of the robot at a proper work distance in an outdoor environment. In the right part of this figure, we also illustrate the user interface of the HMM-based hand gesture classifier installed on the robot.

6 Conclusions

In this paper, we have accomplished the recognition of pre-defined meaningful gestures consisting of continuous hand motions in an image sequence captured form a PTZ camera in real time, and an autonomous robot is directed by the meaningful gestures to complete some responses under an unconstraint environment. Four basic types of directive gestures made by a single hand have been defined such as moving forward, downward, leftward, and rightward. It results in twenty-four kinds of compound gestures from the combination of the directive gestures made by two hands. At present, we apply the most natural and simple way to select eight kinds of compound gestures employed in our developed human-robot interaction system, so that users can operate the robot effortlessly. From the experimental outcomes, the human-robot interaction system works by 7 frames per second on an average, and the resolution of captured images is 320×240 pixels using the PTZ camera. The average gesture recognition rate is more than 96%, which is quite satisfactory and encouraged to extend this system used in varied areas.

Acknowledgement. The authors thank the National Science Council of Taiwan for supporting this work in part under Grant NSC99-2221-E-011-132.

References

[1] Soontranon, N., Aramith, S., Chalidabhongse, T.H.: Improved face and hand tracking for sign language recognition. In: Proc. of the Int. Conf. on Information Technology: Coding and Computing, Bangkok, Thailand, vol. 2, pp. 141–146 (2005)

[2] Zhu, X., Yang, J., Waibel, A.: Segmenting hands of arbitrary color. In: Proc. of the IEEE Int. Conf. on Automatic Face & Gesture Recognition, Pittsburgh, Pennsylvania, pp. 446–453 (2000)

[3] Mitra, S., Acharya, T.: Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews 37(3), 311–324 (2007)

[4] Kim, K.K., Kwak, K.C., Chi, S.Y.: Gesture analysis for human-robot interaction. In: Proc. of the Int. Congress on Anti Cancer Treatment, Phoenix, Arizona, pp. 20–22 (2006)

[5] Fahn, C.S., Chu, K.Y.: Real-time hand detection and tracking techniques for human-robot interaction. In: Proc. of the IADIS Interfaces & Human Computer Interaction Conference, Freiburg, Germany, pp. 179–186 (2010)

[6] Liu, N., Lovell, B.C., Kootsookos, P.J., Davis, R.I.A.: Model structure selection and training algorithm for an HMM gesture recognition system. In: Proc. of the Int. Workshop in Frontiers of Handwriting Recognition, Brisbane, Australia, pp. 100–106 (2004)

[7] Yoon, H., Soh, J., Bae, Y.J., Yang, H.S.: Hand gesture recognition using combined features of location, angle and velocity. Pattern Recognition 34(70), 1491–1501 (2001)

[8] Korgh, A., Brown, M., Mian, I.S., Sjolander, K., Haussler, D.: Hidden Markov models in computational biology: Applications to protein modeling. Journal of Molecular Biology 235(5), 1501–1531 (1994)

[9] Mohamed, M.A., Gader, P.D.: Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation-based dynamic programming techniques. IEEE Transactions on Pattern Analysis & Machine Intelligence 18(5), 548–554 (1996)

[10] Lawrence, R.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. of the IEEE 77(2), 257–286 (1989)