HMM Static Hand Gesture Recognition Based on Combination of Shape Features and Wavelet Texture Features

Lizhi Zhang[1,4](✉), Yingrui Zhang[2], Lianding Niu[3], Zhijie Zhao[1,4], and Xiaowei Han[1,4]

[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
zhanglizhi58@163.com

[2] School of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

[3] Network and Educational Technology Center, Harbin University of Commerce, Harbin 150028, China

[4] Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin University of Commerce, Harbin 150028, China

Abstract. Gesture recognition is one of the key technologies in the field of computer vision, and hand gesture recognition can be divided into static hand gesture recognition and the dynamic hand gesture recognition. This paper presents a new static gesture recognition algorithm based on hidden markov model. It uses two kinds of new shape features, the specific angle shape entropy feature and the upper side contour feature. They are firstly used for parameters training of hidden makov model, and then identify gesture categories hierarchically. In order to further improve the recognition effect for those small shape differences gesture, this paper adopts wavelet texture energy feature which can reflect the internal details of the gesture image, and makes the final correction estimation based on minimum total error probability. The experimental results show that the method has good recognition effects for gestures no matter the shape differences are big or not, and it has good real time performance as well.

Keywords: Static gesture recognition · Shape entropy feature · Texture energy feature · Minimum total error probability · HMM

1 Introduction

Hand gesture recognition is one of the important technologies in the field of computer vision, it has many applications such as: Motion Simulation Games, Sign Language Communication, Virtual Reality and so on. Hand gesture recognition can be divided into static hand gesture recognition and the dynamic hand gesture recognition, and static hand gesture recognition is the recognition from a single image. So far, there are many works on static hand gesture recognition, and various classification methods have been adopted [1, 2].

M. Jia et al. (Eds.): WiSATS 2019, LNICST 281, pp. 187–197, 2019.
https://doi.org/10.1007/978-3-030-19156-6_18

Ren et al. [3] used depth threshold to remove simple background, extracted finger distance as feature, and used template matching to identify 10 kinds of gestures. Singh et al. [4] used Radon transform to recognize the skeletal representation of gestures, extracted features in the transform domain, and then used K-means clustering algorithm to classify gestures. Jiang et al. [5] set up a standard sample base of gestures based on the model of hand geometric relationship and data glove, and then they implemented the gesture recognition based on BP neural network. Dardas et al. [6] used subtraction, skin color detection and contour comparison to track and detect gestures in complex background, they used HSV space to segment gestures, extracted features, and then used multiclass SVM to recognize gestures.

HMM classifier is a successful classifier for gesture recognition. It has flexible and efficient training and recognition algorithm. Xu et al. [7] established the HMM model for ten Arabic numerals (0–9) gestures, and implemented the real-time recognition system of gesture trajectory. In this paper, we present a HMM static hand gesture recognition method based on combination of shape features and wavelet texture features. The algorithm has good recognition effects for gestures with big shape differences and slight shape differences.

2 Static Hand Gesture Feature Extraction

2.1 Preprocessing

Preprocessing is the first stage for our algorithm which mainly contains ROI (region of interest) detection and denoising. The skin color is firstly detected by histogram statistics and threshold setting, and the denoising can be effectively processed by the mathematical morphology method. Figure 1 shows the preprocessed static hand gesture images.

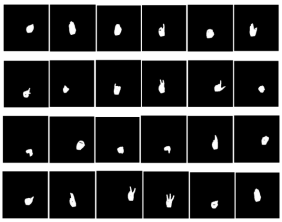

Fig. 1. Preprocessed static hand gesture images.

2.2 The Shape Feature Extraction

The Specific Angle Shape Entropy Feature Extraction of Hand Gesture. For this feature extraction method, the hand gesture edge should be firstly detected. We adopt the Pretty differential operator method to detect the hand gesture edge, and then calculate the centroid pixel coordinate $O(O_X, O_Y)$ of the hand gesture region as shown by the Eq. (1).

$$O_x = \frac{\sum_{i=0}^{sum} x_i}{sum}, O_y = \frac{\sum_{j=0}^{sum} y_i}{sum} \tag{1}$$

Where (x_i, y_i) denotes the coordinate of the pixel on the hand gesture contour, sum denotes total number of the pixels on the hand gesture contour, and the hand gesture region can be divided n parts with uniform angle division as shown Fig. 2.

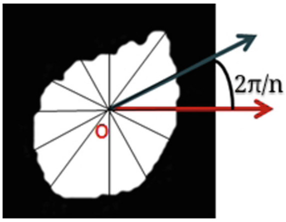

Fig. 2. The specific angle contour illustration.

Based on these processing, we define the specific angle shape entropy to describe the uncertainty of the hand gesture contour so as to distinguish the different hand gesture. We firstly define a probability measurement regarding the specific angle shape of the hand gesture contour as the Eq. (2).

$$p_{ij} = \frac{d_{ij}}{\sum_{(i,j) \subset L_C} d_{ij}} \tag{2}$$

Where L_c denotes the set of the pixels on the on the hand gesture contour within the specific angle divided part, and d_{ij} denotes the distance between the pixels on the contour of the specific angle part and the centroid of the whole contour. With this probability measurement, the specific angle shape entropy can be naturally defined as Eq. (3). The specific angle shape entropy of every part of the whole contour can be put together to form an n dimensional feature vector. For instance, we divide the whole hand gesture contour with every 30° in our experiment, we then obtain 12 dimensional feature vector for a hand gesture.

$$H_{L_C} = -\sum_{(i,j) \subset L_c} p_{ij} \log_a p_{ij} \tag{3}$$

Upper Side Contour Feature Extraction of Hand Gesture. The upper side contour of the hand refers to the contour above the horizontal line through the centroid pixel, and experiment shows that the hand gesture difference mainly depends on the upper side contour of the hand. We can calculate the vertical distance between a specific pixel on the upper side contour and the centroid pixel D_y as shown in Fig. 3. Considering the width of the hand contour, a 42-dimensional feature vector can be properly formed.

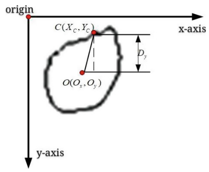

Fig. 3. D_y distance illustration

2.3 Wavelet Texture Energy Feature Extraction

Wavelet Transform. One of the most popular application of wavelets has been to image decomposition. However, images are two-dimensional and decomposition can be illustrated as Fig. 4. The Discrete Wavelet Transform decomposes an image into a multi-resolution expression and it can be shown in Fig. 5. The spatial subbands LL1, HL1, LH1 and HH1 can be got with a two-dimensional image wavelet decomposition. The first subband contains the main information of the image and the other three subbands are the additional information of the image in the horizontal, diagonal and vertical directions. And the spatial subbands LL2, HL2, LH2 and HH2 can be got after a further two-dimensional image wavelet decomposition to LL1. When LL2 is further decomposed, a growing quad-tree can be got [8].

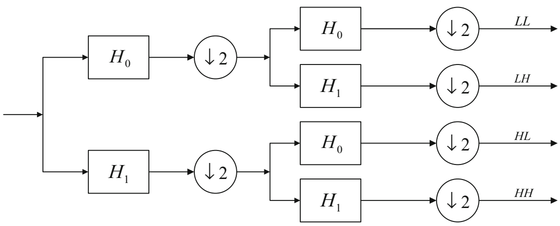

Fig. 4. Subband decomposition of an image.

Wavelet Texture Energy Feature Extraction. Many research works show that the wavelet energy feature is an effective feature in gesture recognition, it can be defined by the coefficients of wavelet decomposition to a cutting gesture image. A cutting gesture image is a part of an image cut off from an image containing a gesture. According to the characteristics of the gesture, Images containing gesture are grouped to several groups

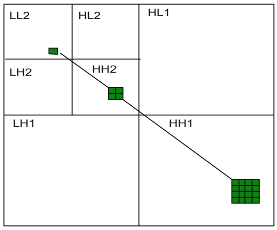

Fig. 5. The subband structure of image.

by the similarity, for instance, gesture g and gesture t, gesture n and gesture q, gesture u and gesture r for editing and cutting, so that most of the resulting cutting gesture image can have differences. The i_{th} level of the wavelet energy in the horizontal, vertical and diagonal directions of a cutting gesture image can be expressed as the Eq. (4) [9, 10]:

$$\begin{aligned} E_i^h &= \sum_{x=1}^{M} \sum_{y=1}^{N} [H_i(x,y)]^2 \\ E_i^v &= \sum_{x=1}^{M} \sum_{y=1}^{N} [V_i(x,y)]^2 \\ E_i^d &= \sum_{x=1}^{M} \sum_{y=1}^{N} [D_i(x,y)]^2 \end{aligned} \tag{4}$$

Where H_i, V_i, D_i is the i_{th} level wavelet decomposition of the cutting gesture image in three directions. Considering the size of the cutting gesture, the cutting gesture is applied to 5-layer haar wavelet transform. In this way, each cutting gesture image can get $1 * 15$ dimensional texture feature vector.

3 Static Hand Gesture Recognition Algorithm

The Hidden Markov process is a double stochastic process as shown in Fig. 6, one is hidden, and another is observation sequence. The notation $\lambda = (\pi, A, B)$ denotes the parameter set of the model [11], where π indicates the initial probability distribution of the hidden state, A indicates the state transition probability distribution, and B indicates the observation symbol probability distribution at given state. The HMM model can be trained using empirical data, and this leads to the model parameters achievement.

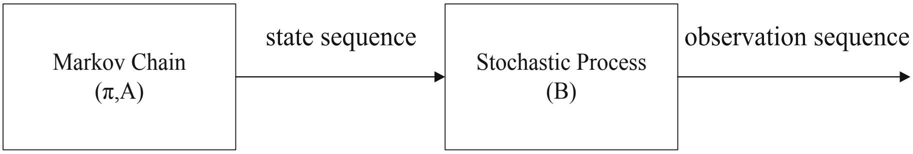

Fig. 6. A hidden Markov process.

4 Gesture Recognition Correction Based on Minimum Total Error Rate

The extraction of wavelet energy features is to distinguish the gesture groups with similar contour. For the two types of gestures that are easily confused with the contour, the weighted city block distance [12] is used to classify the gesture based on the minimum total error rate.

Let V and U be the wavelet energy features of the two gesture images, M be the total levels of wavelet decomposition, and the i_{th} level wavelet energy features of V and U are the weights of the i_{th} level wavelet energy features. Then the definition of wavelet energy features similarity measurement based on the weighted city block distance is given by the formula (5)

$$D(V,U) = \sum_{i=1}^{M}\left(c_i \sum_{j=1}^{3}\left|V_{\substack{i\\(j)}} - U_{\substack{i\\(j)}}\right|\right) \tag{5}$$

The weighted city block distance calculation mainly depends on the influence to the recognition results by the wavelet energy features at each level. The larger the influence of the wavelet energy features is, the larger the corresponding weight c_i is chosen. In the experiment, we extracted the energy features with 5 levels wavelet transform. Figure 7 shows the matching distance of each wavelet transform level of gesture g and gesture t. It can be seen that the 5th-level energy features can most obviously distinguish these two types of gestures, so its weight is set to be the maximum value.

Considering the gesture g and the gesture t, the gesture n and the gesture q, the gesture u and the gesture r, we calculate the reciprocal of the intersecting graphic area

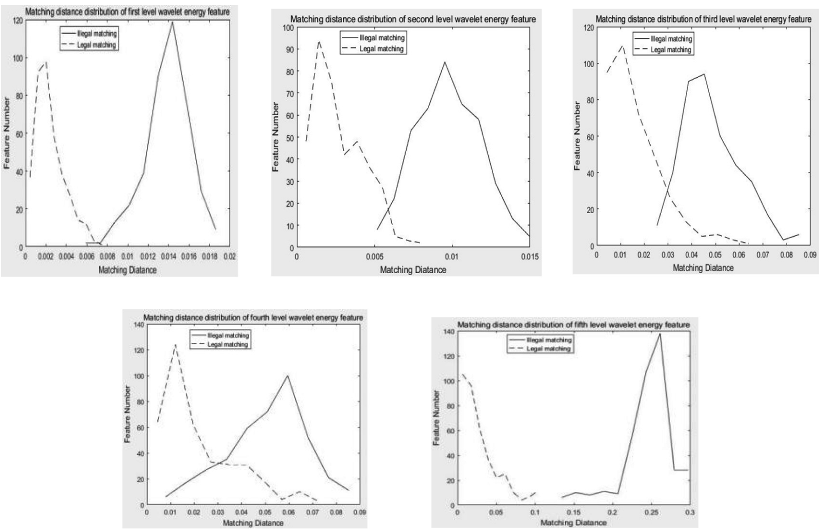

Fig. 7. The matching distance between hand gestures g and hand gestures t

of each wavelet transform levels matching distances curve, and the c_i values can be obtained as shown in Table 1. With the determined weights, we can recognize the gesture in the corresponding gesture group with the following steps. We firstly calculate the wavelet energy features of the gesture, and the matching distance D can be then calculated according to the similarity formula (5). Finally based on the criterion of the minimum overall error rate, the gesture class can be determined as the class with the smaller D value.

Table 1. Gesture group weights at all levels.

Gesture group	c_1	c_2	c_3	c_4	c_5
Gesture g & gesture t	0.3764	0.5388	0.0024	0.0020	0.0804
Gesture n & gesture q	0.5787	0.0692	0.0817	0.0933	0.1771
Gesture u & gesture r	0.8977	0.0645	0.0277	0.0036	0.0065

5 Experimental Results and Analysis

The gesture database used in the experiment was derived from the Thomas Moeslund gesture recognition database of the Aalborg University, Denmark. The gesture recognition database images are TIF format, resolution 248 $*$ 256, the gestures are in different scales and different rotation plane, and it contains small contour differences but belongs to different classes of gestures. So, it is suitable for gesture recognition. 24 kinds of original images are shown in Fig. 8. There are 40 images in each class of gesture A–F and 100 images in each class of gesture G–Y. We select the first 20 groups of images in 24 kinds of gestures as training samples, and the rest of the images are used for recognition test. All experiments take matlab.2012a as an experimental platform.

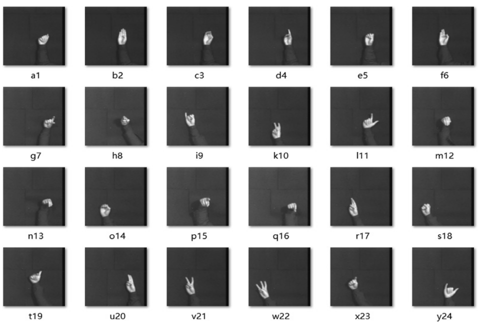

Fig. 8. 24 groups of original gesture images

Table 2. Recognition results using the specific angle shape entropy feature

Gesture group	Recognition rate (%)					
1–6	95	100	100	100	95	85
7–12	80	100	100	100	100	100
13–18	100	100	100	100	85	95
19–24	100	100	85	95	100	100

When HMM statistical model is used as classifier, the computation complexity is its drawback. For the two kinds of shape features extracted before, it is desirable that the specific angle shape entropy features can have an advantage in recognition, or a single feature can get a good recognition result. Therefore, without considering the combination of features, HMM is trained using specific angle contour entropy features and upper contour features separately. Tables 2 and 3 shows the results of the recognition rate respectively, and it illustrates that the special angle contour entropy features outperform the upper contour features in terms of the recognition rate. However, the recognition failure will not occur using the upper contour feature when it fails using the special angle contour entropy features.

So, in order to improve the recognition effect, the shape entropy features and upper side contour features are combined. The former recognition is firstly implemented using shape entropy features, and then we judge whether the recognition rate can be increased further using upper side contour features. Finally, the secondary recognition is implemented if necessary.

Table 3. Recognition results using upper side contour feature.

Gesture group number	Recognition rate (%)					
1–6	95	95	95	100	70	100
7–12	85	100	100	100	80	100
13–18	100	95	85	35	100	85
19–24	100	100	100	100	55	50

After the gesture recognition based on the combination of the shape entropy features and upper side contour features, there are still some wrong recognition gestures as shown in Fig. 9. So, following the multiple features combination strategy, recognition correction based on texture wavelet energy feature was implemented by the steps as shown in Fig. 10. Most wrong recognition gestures are wrongly judged as gestures that are very similar to the outline of gesture classes. For example, gesture *r* and gesture *u* are different only in fingertip overlap and fingertip tiling. In order to distinguish such gestures, the texture wavelet energy features are additionally extracted, and finally the recognition results are corrected according to the similarity of the texture wavelet energy features. The correction results are shown in Table 4, it can be seen that only two gesture recognition failures after the correction based on the principle of minimum total error rate of texture wavelet energy for the gesture n, gesture t and gesture u with

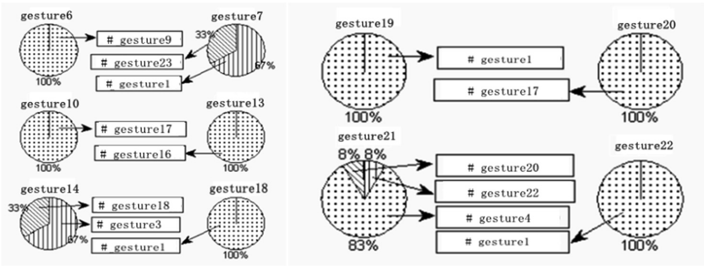

Fig. 9. Wrong gesture recognition distribution (# gesture number means misrecognition as the number).

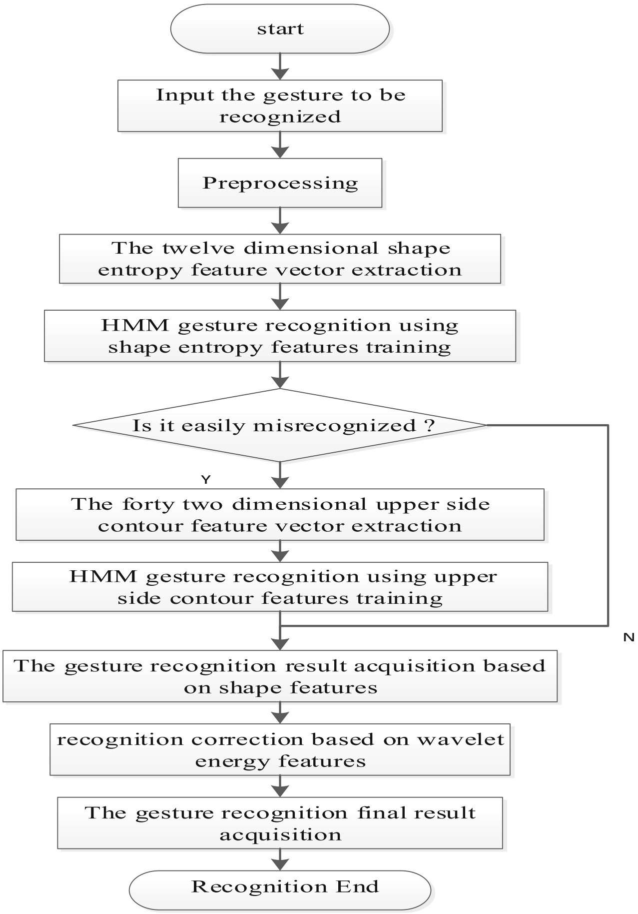

Fig. 10. Static gesture recognition flow chart based on shape features and wavelet texture features.

low recognition rate of shape feature. The recognition rate has been improved from 80.83% to 99.16%, and the overall recognition rate of all gestures reached 98.78%. However, it only increases the use time of 10.018 s, so it shows the performance improvement is significant. It also reflects that the distance between classes of texture is much larger than that between shapes for these specific groups of gesture images.

Table 4. The improvement of the recognition rate after correction.

Gesture group number	Recognition rate before correction	Recognition rate before correction
13%	72.5%	100%
19%	76.25%	97.5%
20%	93.75%	100%

6 Conclusion

Hidden Markov Model trained by static gesture feature has space scale invariance. The algorithm we presented works well even for gestures whose shape difference is not big and provides an independent training for each gesture class which is good for system expansion. Two new features we presented complement each other and recognize gestures step by step. However, it is not good enough for some special gestures who are easily confused in shape. Focusing on those unsatisfied results, this paper adds texture energy feature which can reflect the internal details of the gesture image and make his final correction estimation based on minimum total error probability. The experimental results show that the method has good recognition effect for gestures whose shape differences are small and has good real time performance as well.

Acknowledgement. This work is supported by the Harbin Science and Technology Bureau outstanding subject leader fund project (2017RAXXJ055), Nature Science Foundation of Heilongjiang Province (F2018020).

References

1. Rautaray, S.S., Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. **43**, 1–54 (2015)
2. Pisharady, P.K., Vadakkepat, P., Loh, A.P.: Attention based detection and recognition of hand postures against complex backgrounds. Int. J. Comput. Vis. **10**, 403–419 (2013)
3. Ren, Z., Yuan, J., Zhang, Z.: Robust hand gesture recognition based on finger-earth mover's distance with commodity depth camera. IEEE Trans. Multimed. **15**(5), 1110–1120 (2013)
4. Singh, M., Mandal, M., Basu, A.: Visual gesture recognition for ground air traffic control using the radon transform. In: International Conference on Intelligent Robots and Systems, pp. 2586–2591 (2005)
5. Jiang, L.: Research of Gesture Recognition Based on CAS-Glove. Jiaotong University, Beijing (2006)

6. Dardas, N.H., Georganas, N.D.: Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Trans. Instrum. Meas. **60**(11), 3592–3607 (2011)
7. Xu, X.: Hand Gesture Recognition based on Hidden Markov Module. University of Technology, Guangzhou (2011)
8. Wainwright, M.J., Simoncelli, E.P., Willsky, A.S.: Random cascades on wavelet trees and their use in analyzing and modeling natural images. Appl. Comput. Harmonic Anal. **11**(1), 89–123 (2001)
9. Wu, X., Wang, K., Zhang, D.: Wavelet energy features extraction and matching for palmprint recognition. J. Comput. Sci. Technol. **20**(5), 411–418 (2005)
10. Wu, X., Wang, K., Zhang, D.: Wavelet based palmprint recognition. In: IEEE Proceedings of the International Conference on Machine Learning Cybernetics, USA, pp. 1253–1257 (2002)
11. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**, 257–286 (1989)
12. Wu, X., Zhang, D., Wang, D.: Palmprint recognition. Science Press (2006)