An Improved Algorithm of Hand Gesture Recognition under Intricate Background

Shuying Zhao, Wenjun Tan, Shiguang Wen, and Yuanyuan Liu

College of Information Science and Engineering, Northeastern University, Shenyang, China, 110004
{zhaoshuying,wenshiguang}@ise.neu.edu.cn, {Jun1983,liuyuanyan}@163.com

Abstract. This paper presents an integrated algorithm of YCbCr-Nrg, Double Color-Spatial Model and Background Model to resolve the problem that single skin-color model is obstructed by near kin color. This segmentation method is realized by the fusion of muti-feature. Based on the good describing ability of Fourier Descriptors algorithm and the good self-learning ability of BP neural network, an improved algorithm of hand recognition is presented and carried out. Results show that this algorithm is robustness for hand gesture recognition under intricate background.

Keywords: gesture recognition; gesture segmentation; Fourier descriptors; neural network.

1 Introduction

Hand gesture, as a natural and intuitive expression of information, has become an important human-computer interactive way. Hand as a computer direct input devices, the use of gesture recognition technology to control machines, the middle media in communication between people and machine needed no longer[1]. Hand is a complex deformation and hand gestures is diversity, polysemy and spatio-temporal difference. Because there are uncertainties in computer vision, the gesture recognition system is a challenging research filed[2].

Gesture segmentation is the basis and one of the key technologies of gesture recognition. At present the majority of the gesture segmentation method is based on complexion segmentation[3-6]. The complete hand segmentation is obtained by comprehensive utilization of the hand movement and color information[3]. However, because light is unstable, many recognition results is mistake by easy interference of similar skin color in prospects and intricate background. Based on the color location and histogram matching method or use of training skin data to establish the find table [7-8], because of skin colors have large change in different light conditions, the skin region is undetected or detected errors.

Gestures feature selection and extraction is also the difficulty of hand gesture recognition. Pixels of the verge of a hand gesture regional characteristics as feature of non-rigid or deformation goal is well identify, but not only tends to suffer more from

C. Xiong et al. (Eds.): ICIRA 2008, Part I, LNAI 5314, pp. 786–794, 2008.

noise interference, but also doesn't identify the rotated and scale hand gestures[9]. The method based on moment invariants is easy to recognize the translation, rotation and scaling invariance of hand gestures. But moment invariants as a gesture can only identify the small number of hand gesture categories.

Aiming at the above problems,, this paper presents a method of the context of real-time gesture segmentation based on dual-complexion and adaptive complex background model, and a method of hand gesture contour extraction based on Fourier descriptors. This paper designed a classifier of gesture recognition based on BP network, so that hand gesture recognition system has strong self-learning ability to improve the recognition rate.

2 Gesture Recognition System

Gesture recognition system is composed of two major parts: gesture segmentation module and feature extraction and recognition module, as shown in Fig 1.

In the image segmentation module, the hand gesture image segmentation attained by image pre-processing and image segmentation will be extracted their color feature in three channels. Channel 1 and Channel 2, respectively extract complexion model in

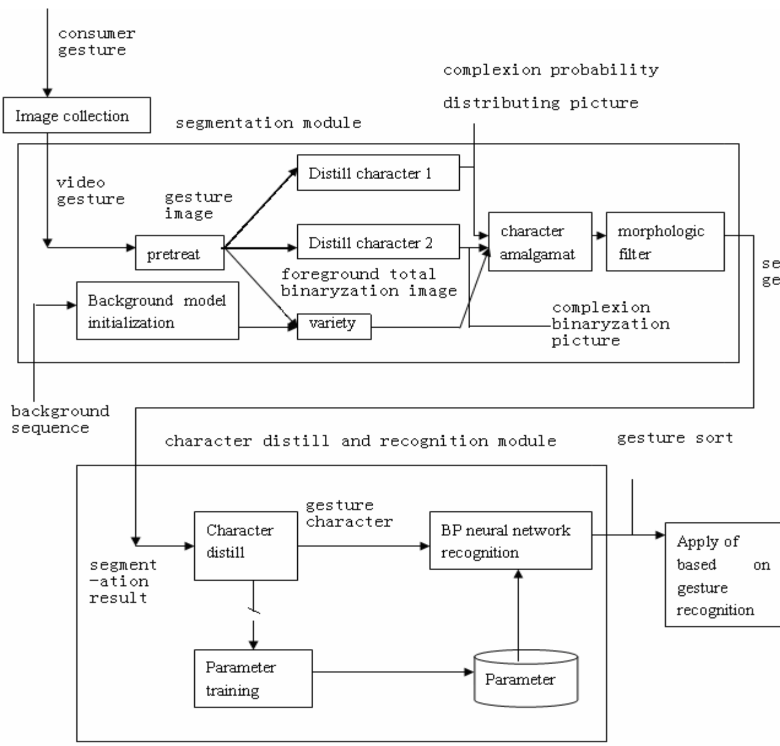


Fig. 1. Structure of hand gesture recognition system

Nrg and YCbCr color space, and Channel 3 extract the foreground feature based on the background of character. Then the hand gestures feature is integrating results of the three-channel.

In feature extraction and recognition module, based on extracting the hand gestures feature, hand gesture is identified by the BP neural network classifier.

3 Hand Gesture Segmentation Based on the Complexion and Background Model

Hand gesture segmentation is separating hand gesture from the background image. In this paper, it is carried out by selecting color space, establishing complexion modeling, adaptive background modeling and other related steps.

3.1 Select Color Space

Based on the statistics of large number of pure complexion samples in HSV, YCbCr, Nrg color space, the cluster analysis results is compared and analyzed. The we select the YCbCr color space in this hand gestures recognition system.

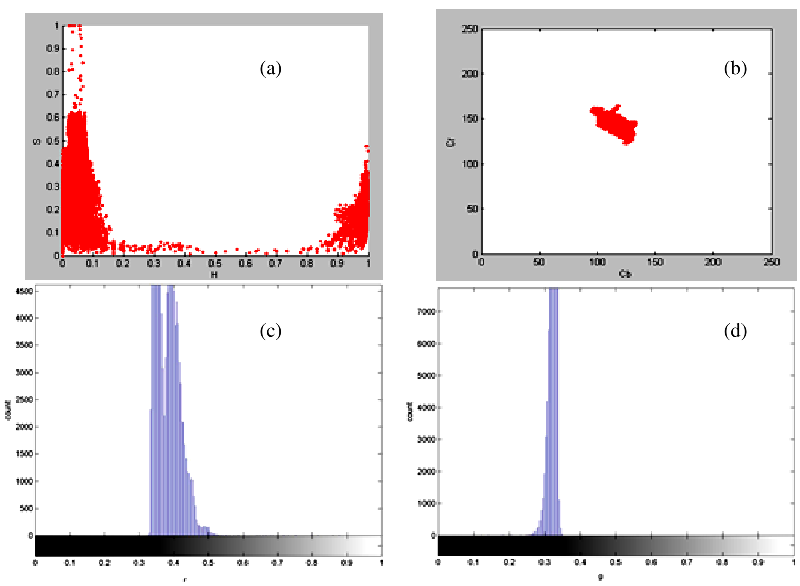


Fig. 2. Layout of color space of complexion:(a) H-S, (b) Cb-Cr , (c) Nrg_r , (d) Nrg_g

3.2 Build Complexion Model

This paper builds the complexion model by using Gaussian distribution in the YCbCr space.

The formula of probability density as follows:

$$f(X)=\frac{1}{\sqrt{2\pi|V|}^{1/2}}\exp[-\frac{1}{2}(X-\mu)V^{-1}(X-\mu)^{T}] \quad (1)$$

where X is defined as $X=(cr,cb)$.

The mean value and covariance matrix as follows:

$$\mu=(\overline{cr},\overline{cb}),\quad V=\begin{bmatrix}\sigma_{cr,cr} & \sigma_{cr,cb}\\ \sigma_{cb,cr} & \sigma_{cb,cb}\end{bmatrix} \quad (2)$$

In a foundation of the complexion sampling, the mean value and covariance matrix of the model is attained by maximum likelihood estimation of the Gaussian model.

Hand gesture segmentation only using this model will produce some interference by similar complexion regional. For example, there are some check errors of the wristband and the fingertips in Fig3 (c). This paper uses Nrg space under the complexion model and background color model to compensate above problems. Custering regional of Nrg color space as follows:

$$r\in[0.33,0.51],\ g\in[0.28,0.35],\ even.r>g$$

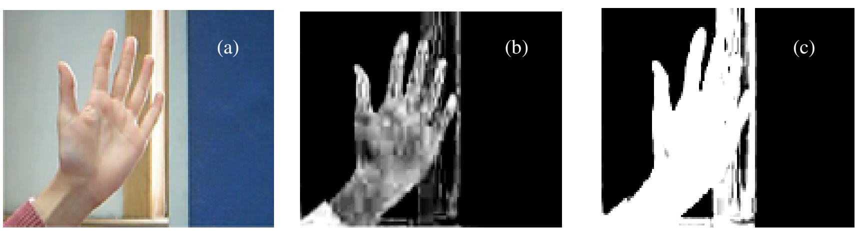


Fig. 3. Complexion segment of YCbCr model: (a) original picture, (b) probability distributing picture, (c) binaryzation picture

3.3 Build Background Model

The adaptation environment ability of the system is improved by building the initial background model and automatically updating to acquire a background model parameters. According to continuous background image, the initial statistical background model is established.

$$\mu_i=\frac{1}{n}\sum_{t=1}^{n}\mu_{it} \quad (3)$$

$$\sigma_i^2=\frac{1}{n}\sum_{t=1}^{n}(\mu_{it}-\mu_i)^2 \quad (4)$$

where n is the number of the continuous image frames, i is The ordinal number of points in the background image, μ_i is the expected value of color variance of this point, σ_i^2 is the variance of distribution of color variance of this point.

The system can adapt to the environment changing by constantly updating background model. Suppose $\mu_i(t)$ and $\sigma_i^2(t)$ is the expected value and the variance of distribution of color variance of the point i at time t, $y_i(t)$ is the color value of the point i at time t. So $\mu_i(t)$ and $\sigma_i^2(t)$ at time $t+1$ as follows:

$$\mu_i(t+1) = \begin{cases} (1-\alpha)\mu_i(t) + \alpha y_i(t), & (D_i = 0) \\ \mu_i(t), & (D_i = 1) \end{cases} \tag{5}$$

$$\sigma_i^2(t+1) = \begin{cases} (1-\alpha)\sigma_t^2(t) + \alpha(y_i(t) - \mu_i(t))^2, & D_i = 0 \\ \sigma_t^2(t), & D_i = 1 \end{cases} \tag{6}$$

3.4 Improve Segmentation Results

Fig4 is the improved segmentation results by using above method. The results show that this method can effectively eliminate interference of similar complexion in the intricate background.

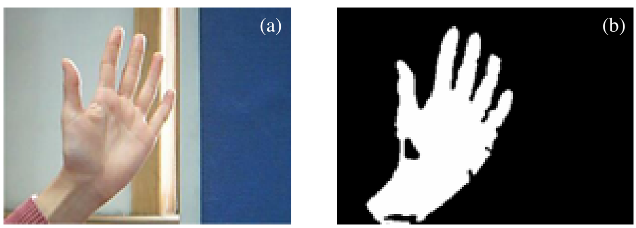


Fig. 4. Improved segmentation result: (a) original picture, (b) segmentation result

4 Fourier Descriptors of Hand Gesture Feature

Fourier descriptor is a good descriptor of the contours, and is rotational invariance, translation invariance and scale invariance. Fourier descriptors are attained by calculating Fourier factors of the sequence of gestures border point. The method describing the gesture feature has nothing to do with the starting point in border and identifies hand gesture fast.

Using a particular point on the border of segmentation gesture as the starting point, coordinate sequence of the border is attained by counter-clockwise.

$$Z(k) = [x(k), y(k)], \qquad k = 0, 1, \cdots, n-1 \tag{7}$$

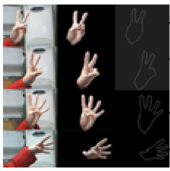
The plural form is as follow:

$$p(l) = x(l) + jy(l), (l = 0, 1, \cdots, n-1), \ j = \sqrt{-1} \tag{8}$$

Two-dimensional problem will be simplified into a one-dimensional problem. The border of one-dimensional Discrete Fourier coefficient sequence is defined as:

$$z(k) = \frac{1}{n}\sum_{l=0}^{n-1} p(l)\exp(-j\frac{2\pi lk}{n}), k = 0,1\cdots,n-1 \tag{9}$$

Then a part of Fourier descriptors of gesture outline is attained as fig5 by normalizing the correlation coefficient.



0.153	0.128	0.102	0.008	0.006	0.001
0.534	0.296	0.101	0.006	0.004	0.001
1.748	0.894	0.389	1.616	0.186	0.015
1.565	0.716	0.974	1.334	0.144	0.121

Fig. 5. Segmentation of test sample and result of contour extraction

Because the curvature of the finger outline is large, the low-frequency Fourier descriptors can not describe the essential character of the outline. 11 contour features were attained by using the K-L transform to compress the features of high-frequency Fourier descriptors. Tracking results of gesture outline as fig6.

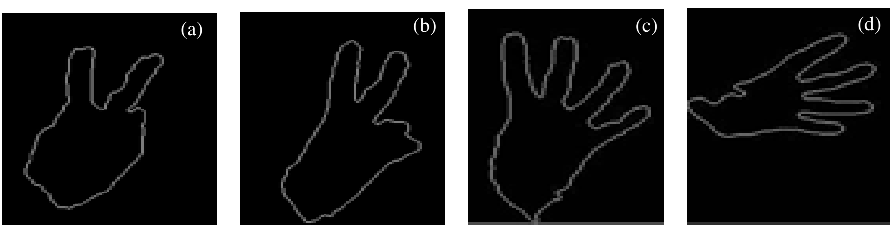


Fig. 6. Contour of gesture: (a) gesture 1, (b) gesture 2, (c) gesture 3, (d) gesture4

5 Gesture Recognition Based on BP Neural Network

The low recognition rate to hand gestures is weakness of the normalized Euclidian distance. To solve the problem, this paper recognizes the hand gesture based on BP neural network.

For distance-based Continental classification of the problem of low recognition rate, based on the design of BP neural network classifiers gesture recognition.

Gesture recognition of the BP network as shown in Figure 7, the network from the input layer, hidden layer and output layer structure, have training and identification of the two working condition.

Input layer: the network of neurons enter the number is 11, corresponding gestures of the 11 characteristics of the outline. Output layer: the output of the network for four neurons. With four binary can be a sign that 0-9.

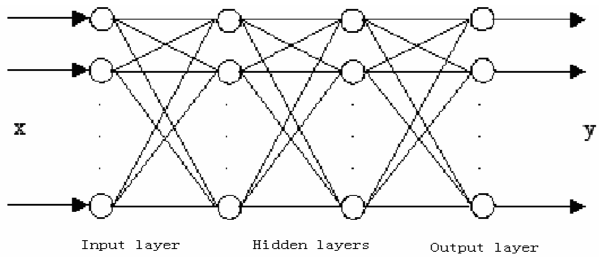


Fig. 7. Structure of three-layer neural network mode

Hidden layers: the hidden network of neurons is 12, can achieve a high recognition rate.

The importation of samples are $X_k = \left[X_{k1}, X_{k2}, \cdots, X_{kM}\right]$, the value vector between the input layer and the hidden layer are $W_{MJ}(n)$, the input (signals) and output (signals) from the input layer to the hidden layer are:

$$u_j^J = \sum_{m=1}^{M} W_{mj} X_{km} \tag{10}$$

$$v_i^J = f\left(\sum_{m-1}^{M} W_{mj} X_{km} + \theta_i^I\right), i = 1, 2, \cdots \tag{11}$$

The input (signals) and output (signals) from the hidden layer to the output layer are:

$$u_p^P = \sum_{j=1}^{J} W_{jp} v_j^J \tag{12}$$

Hidden layers to the output of the input signal and output signals:

$$y_{kp} = v_p^P = f\left(\sum_{j=1}^{J} W_{jp} v_j^J + \theta_i^I\right), p = 1, 2, \cdots, P \tag{13}$$

If the input spreads positively, the output is not the expected the result, then turns to spread reversely, to put the error signals return along the original link, according to the steepest decline in law to amend repeatedly the coefficient of all levels of neurons, errors are made to minimize $d_k = \left[d_{k1}, d_{k2}, \cdots, d_{kp}\right]$ represents (said) the expected output $Y_k(n) = \left[Y_{k1}(n), Y_{k2}(n), \cdots, Y_{kp}(n)\right]$ is the actual output of the network in the course of the Nth iteration, the rate of learning is η. Firstly, to calculate reversely local gradient of every neuron, it is δ. And the local gradient of the output and the hidden layers:

$$\delta_p^P(n) = y_p(n)(1-y_p(n))(d_p(n)-y_p(n)),\ p=1,2,\cdots,P \tag{14}$$

The local gradient of the hidden and input layer:

$$\delta_j^J(n) = f'(u_j^J(n))\sum_p^P \delta_p^P(n) w_{jp}(n),\ j=1,2,\cdots J \tag{15}$$

To amend all of the weight vector by calculating local gradient, the formulas are as follows:

$$\Delta W_{jp}(n) = \eta \delta_p^P(n) v_j^J(n) \tag{16}$$

The amended weight vector between the hidden layers and the output layers:

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n),\ j=1,2\cdots J,\ p=1,2,\cdots,P \tag{17}$$

The amended weight vector between the input layers and the hidden layers:

$$\Delta W_{mj}(n) = \eta \delta_j^J(n) x_{km}(n) \tag{18}$$

$$w_{mj}(n+1) = w_{mj}(n) + \Delta w_{mj}(n), m=1,2\cdots M; j=1,2,\cdots,J \tag{19}$$

Through repeated iteration to make the actual output approach gradually the expected output until meet the required.

Network identification process is described as follows: put the features of test samples as the input of the network, from the input layer the information starts to spread positively in layers after the role of the function, followed by the all nodes of the hidden layer, reached finally the nodes of the output layer , at last, it gives the output.

6 Experimental Results

20 gestures are identified in this paper, through the experiments on samples, to statistics the rates of recognition and time spent in identification in two methods, as shown in table 1, we can see that BP neural network identification method is superior to Continental Distance method.

Table 1. Comparative result of two algorithm recognition rate

classifiers	rates of recognition	time of identification
Continental Distance	83.33%	4.5610
Continental Distance	97.78%	4.5760

7 Conclusions

In the paper, we propose a method that combines Two-complexion model and background model, to a certain extent, resolve the problem of gesture segmentation in the intricate background. But in high light and shadow, the result of gesture segmentation is still not perfect. Combining with Fourier descriptors of the rotational, translation and scale invariance features, the outline of the features is extracted availably. Using BP network as a classification resolved the problem of inconsistencies in space issues and improved the recognition rate.

References

1. Zhu, Y.X., Xu, G.Y., David, J.S.: A real-time approach to the spotting, representation, and recognition of hand gestures for human–computer interaction. Computer Vision and Image Understanding 85, 189–208 (2002)
2. Attila, L., Tamas, S.S.: User-adaptive hand gesture recognition system with interactive training. Image and Vision Computing 23, 1102–1114 (2005)
3. Ren, H.B., Zhu, Y.X., Xu, G.S.: Hand gesture segmentation and recognition with complex backgrounds. Acta Automatica Sinica 10, 256–261 (2002)
4. Kaustubh, S.P., Sumantra, D.R.S.: Hand gesture modelling and recognition involving changing shapes and trajectories, using a Predictive EigenTracker. Pattern Recognition Letters 28, 329–334 (2007)
5. Chen, D.S., Liu, Z.K.S.: A survey of skin color detection. Chinese Journal of Computers 24, 65–69 (2006)
6. Phung, S.L., Bouzerdoum, A., Chai, D.S.: Skin segmentation using color pixel classification: analysis and comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 148–154 (2005)
7. Bradski, G., Boon, Y., Minerva, M.S.: Gesture for video content navigation. In: SPIE IS&T/SPIE Conf., vol. 36, pp. 230–242 (1999)
8. Starner, T., Weaver, J.S.: Real-time American sign language recognition using desk and wearable computer based video. IEEE Transactions Pattern Analysis and Machine Intelligence 20, 1371–1375 (1998)
9. Zhang, L.G., Wu, J.Q., Gao, W.S.: Hand Gesture Recognition Based on Hausdorff Distance. Journal of Image and Graphics 8, 56–57 (2003)