ICIC Express Letters
Part B: Applications
Volume **8**, Number **5**, May **2017**

ICIC International ©2017 ISSN 2185-2766
pp. 851–857

A REAL TIME HAND GESTURE RECOGNITION METHOD BASED ON THE MULTI-CLASS SVM

Kun Xu, Xusong Chen, Shuxu Jing, Youquan Liu
and Caiqin Kang

School of Information Engineering
Chang'an University
Middle-section of Nan'er Huan Road, Xi'an 710064, P. R. China
xukun_yunjing@hotmail.com

Received October 2016; accepted January 2017

Abstract. *In all gesture recognition methods, the method based on computer vision is most sensitive to environment, such as lightings, skin color likely objects, and complex background. In this paper, we introduce a hand gesture recognition system to recognize continuous gesture before stationary background. Firstly, we apply a new hand region segmentation algorithm based on the adaptive skin color model and the motion information to overcoming the impact of environment. Then we use gradient histogram descriptor to represent the contour of the hand region in order to meet the requirements of real time. After this, we apply the support vector machine (SVM) method to recognizing the input hand gesture. In the experiments, we have tested our system to recognize the 32 different gestures. The recognition rate is above 95% and the processing rate is about 36.64ms per frame.*

Keywords: Hand gesture recognition, Adaptive skin model, HOG, SVM

1. **Introduction.** The user interface (UI) of the personal computer has evolved from a text-based command line to a graphical interface with keyboard and mouse inputs. However, they are inconvenient and unnatural. The use of hand gestures provides an attractive alternative to these cumbersome interface devices for human-computer interaction (HCI).

Baudel and Baudouin-Lafon employed sensors attached to a glove that transduced finger flexions into electrical signals for determining the hand posture [1]. This approach forced the user to carry a load of cables which hinders the ease and naturalness of the user interaction. Vision-based method is more and more popular in gesture recognition field because of its nature and convenience. Gupta used localized contour sequences (LCS) to represent the hand region contour and a nonlinear alignment was formulated to determine the similarity between two LCSs. This method was applied to recognizing the hand gestures static images [2]. Chung et al. proposed a real-time hand gesture recognition system based on Haar-wavelet representation [3]. However, this method has a large amount of calculation; moreover, it cannot effectively eliminate the disturbance of face or other skin color regions. Gabor filters were firstly convolved with images to acquire desirable hand gesture features in [4]. With the reduced Gabor features, support vector machine (SVM) was trained and exploited to perform the hand gesture recognition tasks. The recognition rate of 95.2% can be achieved. A real-time video system for hand gesture recognition was also presented with a processing rate of 0.2s for every frame [4]. In a word, hand gesture recognition algorithm not only has strong robustness against the disturbance caused by the likely skin objects and the light, but also should keep high real-time characteristic.

A new real-time hand gesture algorithm is proposed in this paper and the recognition rate is above 95%, and the processing rate is about 36.64ms per frame. The proposed system consists of four main models: image acquisition, hand region segmentation, feature extraction and hand gesture recognition. Firstly, frames are acquired from the USB camera. Then we introduce an adaptive skin region segmentation method to segment the hand region based on skin color information and motion information. The feature parameters are then extracted and combined into the feature vectors. Finally, the feature vectors of gesture are taken as the input of the trained SVM classifier to classify gesture. A media player based on the hand gesture control is implemented to verify the effectiveness of the proposed algorithm. Figure 1 shows the model framework for our hand gesture recognition system.

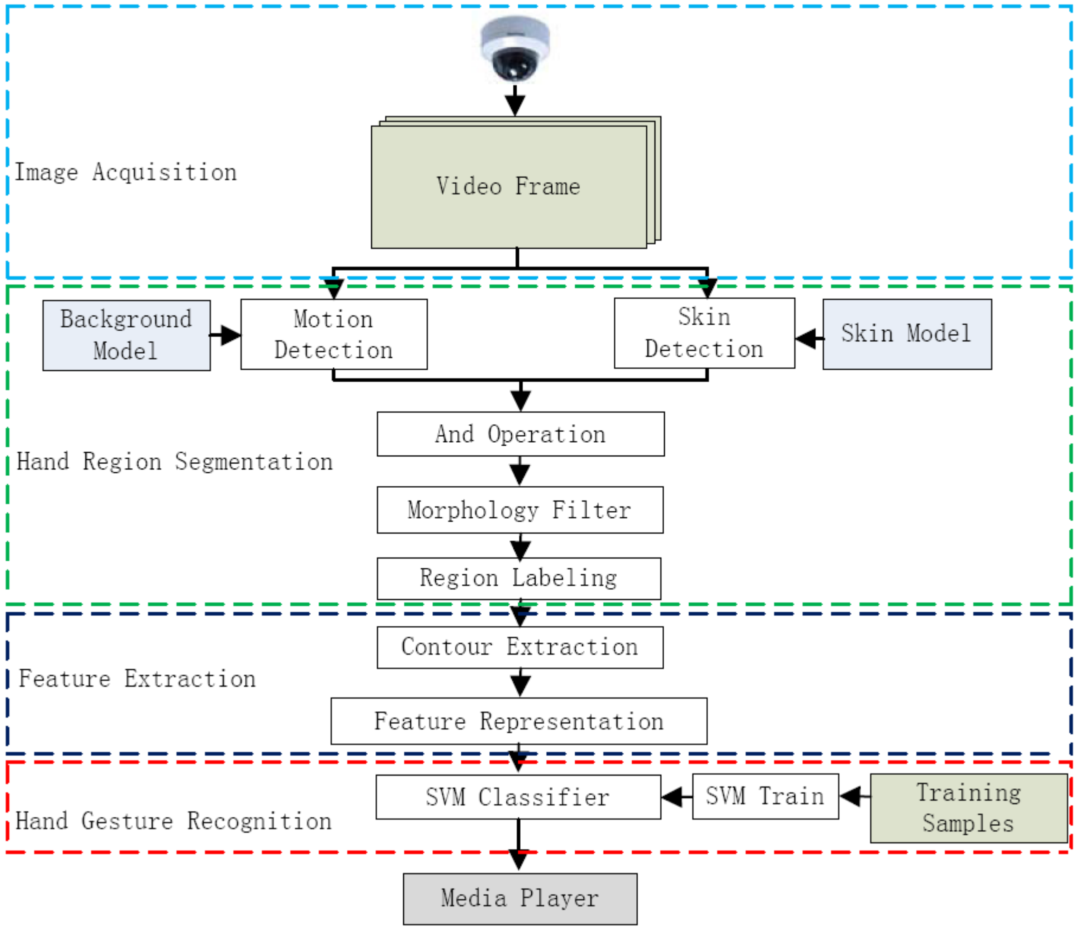


FIGURE 1. Model framework for hand gesture recognition system

The remaining part of this paper is organized as follows. Part 2, Part 3 and Part 4 describe the methods related to our approach in detail, including hand region segmentation, feature representation and hand recognition based on multi-class SVM. The experimental results are described in Part 5. Part 6 concludes our approach and gives some suggestions for our future work.

2. Hand Region Segmentation.

The main procedure of hand segmentation is to detect hand regions in the sequence of hand gesture and separate them from backgrounds.

2.1. Skin color detection.

Hand skin color is relatively concentrated and stable in the color image, and it is not influenced by shape, size and so on.

An effective way to detect hand skin color is to establish the skin color model, but it is difficult to detect skin color more accurately because of different races and various lightings. A new hand skin color model named adaptive Gaussian model $G_i(m_i, C_i)$ in

YCbCr color space is adopted here based on the fact that the color distribution of a single-colored object is not invariant with respect to luminance variations [5].

Firstly, the skin color regions in the images were transformed from the RGB color space to the YCbCr color space. Then the total range of Y is divided into finite number of intervals N. Pixels whose luminance belongs to the same intensity interval are collected and the covariance matrix C_i and the mean value m_i are calculated with respect to luminance Y.

$$m_i = \left(\overline{Cb_i}, \overline{Cr_i}\right), \ \overline{Cb_i} = \frac{1}{N}\sum_{j=1}^{n_i} Cb_j, \ \overline{Cr_i} = \frac{1}{N}\sum_{j=1}^{n_i} Cr_j \tag{1}$$

$$C_i = \begin{pmatrix} \sigma_{CrCr} & \sigma_{CrCb} \\ \sigma_{CbCr} & \sigma_{CbCb} \end{pmatrix} \tag{2}$$

$i = 1, 2, \ldots, N$, n_i is the number of pixels with respect to the ith interval of Y.

The mathematics model is established through the analysis of the statistics, corresponding to different Gaussian model G_i in different luminance Y_i value. All the Gaussian models are used as the training samples for the BP neural network.

According to the distance between the pixel and the center of Gaussian distribution, the skin's color probability of each pixel is obtained using the following equation.

$$P(x_i) = \exp\left(-0.5(x_i - m_i)^T C_i^{-1}(x_i - m_i)\right) \tag{3}$$

The classification rule of Gaussian model can be determined as follows:

$$S_i(x, y) = \begin{cases} 255 & P(x) \geq T \\ 0 & P(x) < T \end{cases} \tag{4}$$

Threshold value T can be established according to Otsu algorithm. Figure 2 shows the skin color detection result.

(a) Original image (b) Skin color likelihood image (c) Skin color binary image

Figure 2. Skin color detection result

2.2. **Motion detection.** Based on the above method, skin regions including hand, face and likely skin regions in the background are segmented. The motion provides the important and useful information for object localization. Here we suppose that the input hand gesture is non-stationary.

When the hand moves in the spatial-time space, motion detector can track the moving hand by examining the difference between the current frame and the background which is defined as

$$\begin{aligned} \mathit{Diff}_i(x, y) &= F_i(x, y) - B_i(x, y) \\ D_i(x, y) &= \begin{cases} 255 & \mathit{Diff}_i(x, y) > T \\ 0 & \text{else} \end{cases} \end{aligned} \tag{5}$$

where $F_i(x, y)$ and $B_i(x, y)$ are the current frame image and the background image.

The mixed Gaussian model method is adopted to initialize the background image. Based on the old background and the current frame, the new background is calculated as follows

$$B_j(x, y) = (1 - \alpha)B_{i-1}(x, y) + \alpha F_i(x, y) \tag{6}$$

where $B_j(x, y)$ is the ith background, $B_{i-1}(x, y)$ is the $(i - 1)$th background and α is the updating weight. Here α is equal to 0.4. Figure 3 shows the motion detection result.

(a) Original image (b) Background image (c) Motion binary image

Figure 3. Motion detection result

2.3. **Hand region identification.** The hand gestures information consists of movement and skin color information. The logic 'AND' is used here to combine these two types of information, that is

$$C_i(x, y) = S_i(x, y) \wedge D_i(x, y) \tag{7}$$

The results after the logic operation are shown as Figure 4.

(a) Original image (b) Skin color region (c) Motion region (d) Combined region

Figure 4. Hand gesture information

It is hard for the head to maintain stationary when the hand is moving, so there is some other information except the hand information in the combined region. A simple method for region identification is the labeling process. After the labeling process, the small regions can be treated as noise and then be removed. After this, we use the eight neighborhood search algorithm to obtain the contour of the hand region. The results are shown as Figure 5. Regarding the distance between the person and the screen, the binary hand image is normalized to the same size. The size is $64 * 64$.

The proposed hand region segmentation algorithm has the capacity of eliminating the disturbance of facial skin color and some other complex backgrounds.

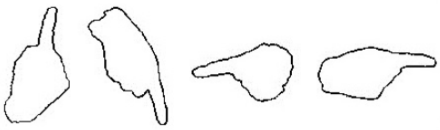
3. **Feature Representation.** Feature representation is an important section for gesture recognition. There are many feature types such as the chain code [6], the local binary pattern (LBP) [7] and the momentum [8].

HOG descriptor involves calculating a gradient direction and magnitude for every pixel in the image and binning these gradients by their direction with a weight based on their magnitudes. It has the strong robustness to the light, rotation in a small range and

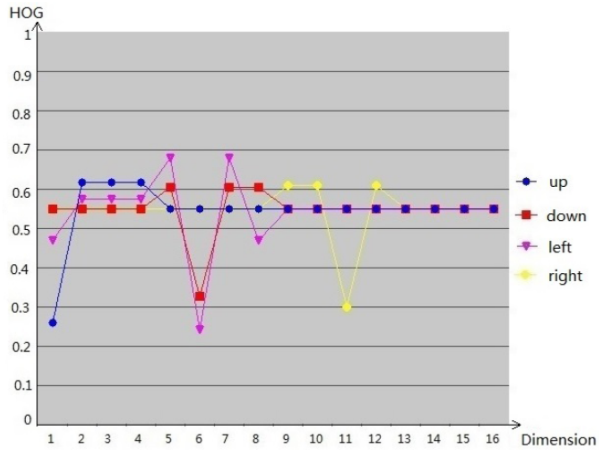
(a) Labeling process result

(b) Hand contour image after size normalized

Figure 5. Hand contour identification results



(a) Four different directions for the same hand gesture



(b) Different HOG features

Figure 6. HOG of the same hand gesture with different directions

the scale change of the object [9]. Except that, different directions of the same hand gesture have different meaning and HOG descriptor has the characteristic of the direction variant, shown as Figure 6. HOG descriptor is selected as the object feature in this paper. Different from [9], the proposed algorithm here is applied into the real-time input video.

4. **Hand Recognition Based on Multi-Class SVM.** There are various approaches for gesture recognition, ranging from template matching [10] or fuzzy-c-mean clustering [11] to artificial neural network [12] or Naive Bayes [13], etc. The classification based on SVM [14,15] is the most commonly used method nowadays due to its simpleness and strong robustness.

Consider the problem of separating the set of training vectors belonging to two separate classes. Suppose a vector in R^D, denoted as $\{x_1, x_2, \ldots, x_n\}$. Each observation x_i is attached with a class label $t_i \in \{-1, +1\}$. We consider a decision function $y(x) = w^T x + b$. $t_i = +1$ when $y(x_i) > 0$ for all i and $t_i = -1$ when $y(x_i) < 0$ for all i. We can combine these requirements by stating,

$$t_i y(x_i) > 0 \ \ \forall i \tag{8}$$

In a linear division environment, SVM can use hyper-plane directly for classification. However, most problems arise from the nonlinear division environments. We should find a mapping function $\varphi(x)$ such that

$$y(x) = w^T \varphi(x) + b \tag{9}$$

The problem can be equivalently understood in terms of projecting the input data into a higher dimensional space where they are separated using the parallel hyper planes. Here the mapping function is radial basis function with Gaussian width σ and error weight C which are set as 8 and 2.0 respectively.

The idea extending it to multi-class problem is to decompose an M-class problem into a series of two-class problems. A method named one versus rest is applied here to recognizing different hand gestures.

5. **Experimental Results.**

The hand recognition system is running on the hardware environment of Intel (R) Core (TM) i5 (2.5GHz), and the software environment of Windows 7 (64bit) and Visual Studio 2008.

In this paper, eight gestures with 4 different directions, shown as Figure 7, have been recognized. There are 250 samples for each direction of every hand gesture. 6400 samples are used for training and 1600 samples are used for testing. Table 1 shows us the experimental results. The recognition rates of training samples and testing samples are 100% and 97.08% respectively. The processing rate is about 0.036 seconds per frame which makes our method can be applied to real-time system.

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 7. Different hand gestures

Table 1. Gesture recognition experiments results

Type	Recognition Rate (%)	
	Training Samples	Testing Samples
(a)	100	100
(b)	100	100
(c)	100	94.3
(d)	100	95.4
(e)	100	90.9
(f)	100	96
(g)	100	100
(h)	100	100
Average Recognition Rate	100	97.075

Table 2. Gesture actions for controlling command

Gesture								
Function	Turn on the volume	Turn down the volume	Fast backward	Fast forward	Stop play	Start play	Full screen play	Exit full screen play

An Intelligent Media Player is implemented based on the method we proposed. Five different gestures above have been integrated into the media player. We define the upwards gesture as "turn on the volume", downwards gesture as "turn down the volume", and so on. The gesture actions for controlling command are shown as Table 2.

6. **Conclusion.** In this paper, a real-time algorithm based on the adaptive skin model segmentation and SVM has been proposed to recognize hand gestures. One advantage of the proposed approach is that it is strongly robust against disturbance of the face and other skin color regions. In addition, the proposed recognition system is glove-free, accurate and fast. The accuracy is 97.75% in average and the recognition time is about 36.64ms per frame. The results imply that the proposed hand gesture recognition system is both effective and real-time, even in the condition of just using a cheap USB camera as input device. There are several studies that should be carried on in the future work. Dynamic hand gesture should be integrated to our system and how to apply our methods to depth image is a new challenge.

Acknowledgments. This work is supported by Special Fund for Basic Scientific Research of Central Colleges, Chang'an University (CHD2013G1241108).

REFERENCES

[1] T. Baudel and M. Baudouin-Lafon, CHARADE: Remote control of objects using free-hand gestures, *Comm. ACM*, vol.36, no.7, pp.28-35, 1993.
[2] L. Gupta and S. Ma, Gesture-based interaction and communication: Automated classification of hand gesture contours, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.31, no.1, pp.114-120, 2001.
[3] W. K. Chung, X. Wu and Y. Xu, A realtime hand gesture recognition based on Haar wavelet representation, *Proc. of IEEE International Conference on ROBIO 2008*, Sanya, China, pp.336-341, 2008.
[4] D.-Y. Huang, W.-C. Hu and S.-H. Chang, Vision-based hand gesture recognition using PCA+Gabor filters and SVM, *The 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kyoto, Japan, pp.1-4, 2009.
[5] G. Yang, H. Li, L. Zhang and Y. Cao, Research on a skin color detection algorithm based on self-adaptive skin color model, *International Conference on ICCIIS 2010*, Nanning, China, pp.266-270, 2010.
[6] W. Xu and E.-J. Lee, A new NUI method for hand tracking and gesture recognition based on user experience, *International Journal of Security and Its Applications*, vol.7, no.2, pp.149-158, 2013.
[7] Y. Ding, H. Pang and X. Wu, Static hand-gesture recognition using HOG and improved LBP features, *International Journal of Digital Content Technology and Its Application*, vol.5, no.11, pp.236-243, 2011.
[8] J. Li, L. Zheng, Y. Chen et al., A real time hand gesture recognition system based on the prior facial knowledge and SVM, *Journal of Convergence Information Technology*, vol.8, no.11, pp.185-193, 2013.
[9] Sheenu, G. Joshi and R. Vig, Histograms of orientation gradient investigation for static hand gestures, *International Conference on ICCCA 2015*, Greater Noida, India, pp.1100-1103, 2015.
[10] M. Matsuda, T. Nonaka and T. Hase, Input method by hand motion for a portable information appliance, *Proc. of IEEE International Conference on SMC 2006*, Taipei, Taiwan, pp.157-161, 2006.
[11] R. Verma and A. Dev, Vision based hand gesture recognition using finite state machines and fuzzy logic, *International Conference on ICUMT*, St. Petersburg, Russia, pp.1-6, 2009.
[12] H.-I. Lin and Y. P. Chiang, Understanding human hand gestures for learning robot pick-and-place tasks, *International Journal of Advanced Robotic Systems*, vol.12, no.5, pp.1-14, 2015.
[13] A. Christoph, K. Thomas, B. Jonas et al., Advancing muscle-computer interfaces with high-density electromyography, *Proc. of the 33rd Annual Conference on Human Factors in Computing Systems*, New York, USA, pp.929-938, 2015.
[14] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
[15] A. I. Maqueda, C. R. Del-Blanco, F. Jaureguizar et al., Human-computer interaction based on visual hand-gesture recognition using volumetric spatiograms of local binary patterns, *Computer Vision and Image Understanding*, vol.141, no.8, pp.126-137, 2015.