

# Application of the Support Vector Machine Algorithm based Gesture Recognition in Human-computer Interaction

Wangcheng Cao

School of Computer and Information Technology, Mudanjiang Normal University

Mudanjiang, Heilongjiang, 157011, China

Corresponding address: Mudanjiang Normal University

No.191, Cultural Street, Aimin District, Mudanjiang, Heilongjiang 157011, China

E-mail: wangchengc80@126.com

**Keywords:** support, human-computer interaction, gesture recognition, image segmentation

**Received:** November 29, 2018

*A gesture recognition technology is an important part of the human-computer interaction. This study focuses on the application of the support vector machine (SVM) in gesture recognition. The gesture image is segmented by YCgCr color space based skin color segmentation method. Then four Hu invariant moments and the ratio of area to circumference of gesture are taken as eigenvalues to extract gesture features. Finally, SVM is used for recognition. It was discovered that the proposed method has a good performance in the gesture recognition and can segment the collected images accurately. The recognition rate of Hu invariant moments based SVM algorithm reaches 99.2% in the recognition of the six gestures designed in this study, which is 9.2% higher than that of the HMM algorithm. The proposed method is reliable and feasible and can contribute to the simple man-machine interaction.*

*Povzetek: Opisana je aplikacija algoritma SVM za prepoznavanje gest pri komunikaciji z računalnikom.*

## 1 Introduction

With the development of science and technology, human-computer interaction has gradually become a part of people's lives [1]. The human-computer interaction technology has been constantly developed [2]. Gesture recognition is an important part of human-computer interaction [4] and plays an indispensable role in people's daily life, and it has a wide application in computer games, virtual reality, medical care and other areas [5,6]. Kuang et al. [7] used the zed stereo camera to get the gesture depth image, segmented the gesture image by the depth information and color information detection, carried out the fingertip detection, and recognized the five kinds of digital gestures by support vector machine (SVM). The average recognition rate was 94.9%, which indicated the high validity of the method. Huang et al. [8] proposed a Gabor filter and SVM based hand gesture recognition method which eliminated the limitation of illumination conditions and obtained a recognition rate of 96.1% in the experiment. Moreover it was found that the use of Gabor filter improved the recognition accuracy from 72.8% to 93.7%, which suggested the high feasibility of the method. Li et al. [9] designed a gesture recognition system, used previous facial knowledge based adaptive skin region segmentation algorithm to segment gesture, and then used SVM to recognize gesture. The experimental results showed that the gesture recognition method had a recognition rate of 95.88%, indicating that the method had excellent performance in gesture recognition and could be applied in real life. Nagarajan et al. [10] proposed a gesture recognition system based on edge histogram features and multi-class SVM to recognize American Sign

Language (ASL) and found that the recognition rate of this method was 93.75%. In the present study, the YCgCr color space based skin color segmentation method was used to segment gesture image. Then the gesture image was recognized by Hu invariant moments based SVM algorithm to study the recognition effect of the method.

## 2 Gesture recognition technology

Man-human interaction technology realizes the rapid communication between human and machine, which brings huge convenience to people's lives. Gesture is one of the daily communication means. Recognizing gesture can help computer understand the behavior act of humans and bring an intuitive experience to users, which is a natural human-computer communication means [11]. Gesture recognition is based on a data glove or vision. In data glove based gesture recognition, relevant information of hands are obtained through data glove, and then the collected data are recognized using computer. It is high-efficient, but is high-cost and complex; hence it is difficult to be promoted. Vision-based gesture recognition is to collect hand images by camera and then recognize them by image analysis. It is practical and has been widely studied. It has been widely used in many fields, such as sign language recognition, somatosensory games and smart home.

In the gesture recognition, the acquired image is segmented firstly to obtain gesture image, and the features of the gesture image are extracted. Then, the gesture recognition algorithm was used for recognizing the

gesture image. A complete gesture recognition system is shown in Figure 1.

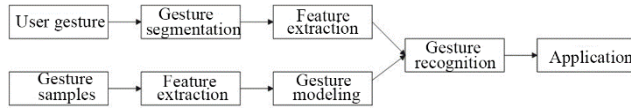


Figure 1: Gesture recognition system.

### 3 YCgCr color space based gesture segmentation method

To accurately recognize gesture, the gesture needs to be separated from the gesture video. The commonly used gesture segmentation methods included skin color segmentation method, background differencing method and pattern matching method. In this study, YCgCr color space based skin color segmentation method was used to segment gesture image.

YCgCr color space has many advantages in gesture segmentation. It is seldom affected by illumination. Y channel can represent the brightness information of the image. Gray image can be extracted directly on Y channel. Cg and Cr components can effectively identify skin color and non-skin color regions.

Fixed threshold value was used to detect skin color. When the pixel value of an image satisfied the condition  $Y \in [35, 230]$ ,  $Cg \in [80, 127]$ ,  $Cr \in [133, 173]$ , it was recognized as a skin color pixel and reset to 255; otherwise it was recognized as a non-skin color pixel and reset to 0. Thus a binary image containing noise was obtained. Then the segmentation process was carried out using this image.

(1) The skin color regions with the area smaller than 400 pixels were eliminated.

(2) The skin color regions with the width and height smaller than 20 pixels were removed.

(3) The center of gravity  $(x_z, y_z)$  of the remaining skin color region was calculated.

$$x_z = \frac{m_{10}}{m_{00}}, y_z = \frac{m_{01}}{m_{00}}, \quad (1)$$

where

$$m_{10} = \sum_{x=1}^w \sum_{y=1}^h xf(x, y), m_{01} = \sum_{x=1}^w \sum_{y=1}^h yf(x, y), m_{00} = \sum_{x=1}^w \sum_{y=1}^h f(x, y) \quad (2)$$

(4) The ratio of height (H) to width (W) of the skin color region was calculated and defined as  $\sigma$ ,

$$0.7 \leq \sigma = \frac{H}{W} \leq 3.0.$$

### 4 Hu invariant moment based gesture feature extraction

To improve recognition effect, feature extraction was performed on the segmented gesture binary image, i.e., selecting the features which could represent gesture as the feature vector. Features extracted included feature of

normalized moment of inertia (NMI), the Fourier descriptor and geometrical characteristic [12]. Hu invariant moment was selected to extract feature of the gesture image.

Hu moments are invariant to target translation and rotation. Hu invariant moment theory include seven  $t$  moments defined as:

$$\phi_1 = \eta_{20} + \eta_{02}, \quad (3)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 - 4\eta_{11}^2, \quad (4)$$

$$\phi_3 = (\eta_{30} - 3\eta_{21})^2 + (3\eta_{21} + \eta_{03})^2, \quad (5)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \quad (6)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} - \eta_{12}) + [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (7)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{21} + \eta_{03}), \quad (8)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) + [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (9)$$

Due to calculation complexity of high-order moments,  $\phi_1 - \phi_4$  were selected as features. Moreover the ratio of area to circumference of the gesture image was calculated and also taken as the feature parameter.

Circumference (L) refers to the sum of pixels on the border line:

$$L = \sum \sum h(x, y),$$

where

$$h(x, y) = \begin{cases} 1, & \text{when point}(x, y) \text{ is the contour point of gesture} \\ 0, & \text{when point}(x, y) \text{ is the non-contour point of gesture} \end{cases} \quad (10)$$

Area (S) refers to the sum of pixels in the hand region in the image:

$$S = \sum \sum f(x, y), \text{ where}$$

$$f(x, y) = \begin{cases} 1, & \text{when point}(x, y) \text{ is the gesture area} \\ 0, & \text{when point}(x, y) \text{ is the non-gesture area} \end{cases} \quad (11)$$

The ratio of the area to the circumference of the eighth characteristic parameter is  $A = \frac{S}{L}$ .

### 5 Gesture recognition algorithm

The commonly used gesture recognition algorithms include dynamic warping algorithm [13], Hidden Markov Model (HMM) and neural network. In recent years, SVM has been frequently used in gesture recognition [14]. In this study, SVM was selected as the gesture recognition algorithm.

### 5.1 SVM algorithm

Suppose there was a sample set  $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ , where  $x_i \in X, y_i \in Y = \{1, -1\}$ . If there was a hyperplane  $(w \cdot x) + b = 0$ , then linear discriminant function was  $g(x) = w \cdot x + b$ , where  $w$  stands for weight vector and  $b$  is a constant. Then class interval was:

$$d(w, b) = \min_{\{x_i|y_i=1\}} \frac{w \cdot x_i + b}{|w|} - \max_{\{x_i|y_i=-1\}} \frac{w \cdot x_i + b}{|w|} = \frac{1}{|w|} - \frac{-1}{|w|} = \frac{2}{|w|} \quad (12)$$

If the condition  $y_i[(w \cdot x_i) + b] \geq 0, i = 1, 2, \dots, N$  was satisfied and the class interval was the largest, then this was the optimal hyperplane.

Linearly separable SVM could be rewritten as an optimization problem.

$$\begin{cases} \min & \frac{\|w\|}{2} \\ \text{st.} & (y_i(w \cdot x_i) + b) - 1 \geq 0, i = 1, 2, \dots, N \end{cases} \quad (13)$$

where  $w$  and  $b \in R, X_n \in R^M$  are eigenvectors and  $y_n \in (-1, 1)$  represents the affiliated category value. Lagrangian multiplier was used for solution. Then the problem could be written as:

$$w(\alpha) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n \alpha_i \quad (14)$$

where  $\alpha_i$  stands for Lagrangian multiplier .  
The final classification function was:

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b) \quad (15)$$

If it was linearly inseparable, slack variable  $\zeta$  needed to be introduced. The objective function was:

$$\begin{cases} \min & \frac{\|w\|^2}{2} + C \sum_{i=1}^N \zeta_i \\ y_i [(w \cdot x_i) + b] \geq 1 - \zeta & i = 1, 2, \dots, N \\ \zeta \geq 0 \end{cases} \quad (16)$$

where  $C$  stands for the penalty factor.  
The following equation could be obtained after the solution based on Lagrangian multiplier:

$$w(\alpha) = \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^n \alpha_i, \quad (17)$$

where  $K(x_i, x_j)$  stands for the kernel function .

The final classification function was:

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i K(x_i \cdot x) + b) \quad (18)$$

### 5.2 Common kernel functions

(1) Linear kernel function is:

$$K(x, x_i) = x \cdot x_i \quad (19)$$

(2) Polynomial kernel function is:

$$K(x, x_i) = [(x \cdot x_i) + 1]^p \quad (20)$$

where  $p$  stands for the polynomial order.

(3) Radial basis kernel function is:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{\sigma^2}) \quad (21)$$

Different kernel functions will affect the classification performance of SVM. It is found that radial basis kernel function has better performance. Therefore radial basis kernel function was used in this study.

## 6 Verification of gesture recognition system

### 6.1 Establishment of sample library

Gestures were collected with the video camera in the laboratory environment. Six gestures were designed, as shown in Figure 2. An experimenter repeated every gesture in front of the camera for 100 times. The first 60 samples of every gesture were taken as the training samples, and the remaining 40 samples were taken as the experimental samples. There were 360 training samples and 240 experimental samples, totally 600 samples.

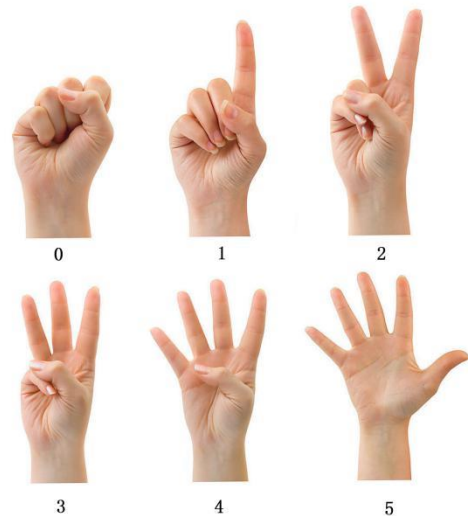


Figure 2: Experimental gestures

### 6.2 Gesture recognition results

Firstly YCgCr color space based skin color segmentation method was used to segment the samples. Images of the gestures obtained after segmentation are shown in Figure 3.

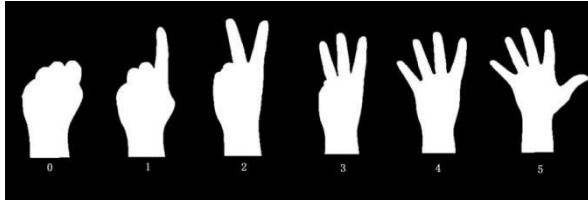


Figure 3: The segmentation results of gestures

It was found that the used gesture segmentation method accurately segmented gestures. According to the obtained segmentation results, features of the training samples were extracted, and five eigenvalues were obtained, including 4 invariant moments and the ratio of area to circumference A, as shown in Table 1.

	$\phi_1 (10^{-3})$	$\phi_2 (10^{-7})$	$\phi_3 (10^{-8})$	$\phi_4 (10^{-9})$	A
0	8.72645	3.32452	1.75426	2.71256	-1.42658
1	1.02154	3.59875	1.02568	2.31456	-2.25454
2	8.62135	2.61245	2.26589	4.501247	-3.26589
3	6.62354	9.28452	4.12523	4.12589	-4.23654
4	5.62147	5.16521	3.28564	3.25489	-5.07541
5	5.01245	1.06241	1.24658	1.20158	-5.92587

Table 1 The eigenvalues of gestures.

SVM was trained using the selected samples. Then it was used to recognize the experimental samples. To verify the recognition effect of the SVM, HMM and SVM recognition performance were compared. The results are shown in Table 2.

Table 2 shows that the SVM has a very favourable performance. Only 2 out of 240 samples were misjudged. The recognition rate of 4 gestures out of 6 gestures reached 100%, and the overall recognition rate reached 99.2%. The number of errors of HMM algorithm was relatively large, totally 24, and the average recognition rate was 90%. This showed that the effect of gesture recognition based on SVM was better than the one based on HMM. The recognition based on SVM could accurately recognize the gesture samples after gesture segmentation and feature extraction, with few number of wrongly recognized samples and high recognition rate; therefore the method was reliable.

Recognition results of SVM						
Gesture	0	1	2	3	4	5
Number of correctly recognized samples	40	40	39	40	39	40

Number of wrongly recognized samples	0	0	1	0	1	0
Recognition rate	100%	100%	97.5 %	100%	97.5 %	100%
Average recognition rate	99.2%					
Recognition results of HMM						
Gesture	0	1	2	3	4	5
Number of correctly recognized samples	38	36	36	35	34	37
Number of wrongly recognized samples	2	4	4	5	6	3
Recognition rate	95%	90%	90%	87.5 %	85%	92.5 %
Average recognition rate	90%					

Table 2 Comparison of the recognition results of SVM and HMM.

### 7 Discussion

In the aspect of the gesture segmentation, the skin color segmentation method based on YCgCr color space was used in this study [15]. It is less limited by light, and the calculation involved is simpler. It can segment the gesture image more accurately to facilitate feature extraction and recognition.

The segmented gesture image contains a large amount of data. In order to recognize the image effectively, the feature vectors are needed. The extraction of feature vectors has a great influence to the recognition accuracy. In this study, Hu invariant moments were selected to extract the features of the gesture image. In order to reduce the computation, only the first four invariant moments were used together with the ratio of the area to the circumference of the gesture. In total, five feature parameters were used for gesture recognition.

The SVM algorithm was selected due to its unique advantages in classification of small sample sets and non-linearity. It was also successfully used in several other pattern recognition applications, data mining and other aspects. In this paper, the radial basis function is used as the kernel function. In the experiment, it was shown that the support vector machine algorithm has a good recognition performance. In the recognition of six gestures, the recognition rate is as high as 99.2%, while the recognition rate was 90% when HMM was used. The SVM clearly outperformed the HMM.

The gesture recognition approach described in this paper has a good performance and can be used in the actual human-computer interaction. For example, in the

application of the intelligent remote control, gesture 0 can be used for indicating shutdown, gesture 1~5 can be used to indicate channel 1~5. It can also be used in intelligent audio, gesture 0~5 corresponding to shutdown, last song, next song, volume plus and volume down. With further training and recognizing additional gestures this could be applied even more extensively.

## 8 Conclusion

Gesture recognition can demonstrate a simple human-computer interaction and can have a great applied value in people's daily life. In this study, firstly, the YCgCr color space based skin color segmentation method was used for segmenting the collected hand gesture images. Then, features were extracted using the first four features of the Hu invariant moment theory together with the ratio of area to circumference of the hand gesture images. Next SVM was used for gesture recognition in images. In experimental setup the recognition rate of SVM was 99.2%, which was 9.2% higher than the HMM approach, indicating the high reliability of the SVM algorithm. This work provides a theoretical support for the application of the SVM in human-computer interaction.

## Acknowledgement

This study was supported by Research on the Application of SVM in Human-Computer Interaction under grant number 1351MSYYB002 and Research on Some Key Technologies of the Internet of Things Architecture and Intelligent Information Processing Theory under grant number YB2018005.

## References

- [1] Hasan H. S., Kareem S. A. (2015). Human Computer Interaction for Vision Based Hand Gesture Recognition: A Survey. *Artificial Intelligence Review*, pp. 1-54. <https://doi.org/10.1007/s10462-012-9356-9>
- [2] Grudin J., Carroll J. M. (2017). From Tool to Partner: The Evolution of Human-Computer Interaction. *Extended Abstracts of the Chi Conference*. Morgan & Claypool, Morgan & Claypool, pp. 183. <https://doi.org/10.2200/S00745ED1V01Y201612HCI035>
- [3] Biswas K. K., Basu S. K. (2012). Gesture recognition using Microsoft Kinect®. *International Conference on Automation, Robotics and Applications*. IEEE, Wellington, New Zealand, pp. 100-103. <https://doi.org/10.1109/ICARA.2011.6144864>
- [4] Panwar M. (2012). Hand gesture recognition based on shape parameters. *International Conference on Computing, Communication and Applications*. IEEE, Dindigul, Tamilnadu, India, pp. 1-6. <https://doi.org/10.1109/ICCCA.2012.6179213>
- [5] Ren Z., Meng J., Yuan J. (2011). Depth camera based hand gesture recognition and its applications in Human-Computer-Interaction. *Communications and Signal Processing*. IEEE, Singapore, pp. 1-5. <https://doi.org/10.1109/ICICS.2011.6173545>
- [6] Córdova-Palomera A., Fatjó-Vilas M., Kebir O., et al. (2011). Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey. *International Journal of Computer Science & Engineering Survey*, pp. 122-133. <https://doi.org/10.5121/ijcses.2011.2109>
- [7] Kuang D., Yang C., Wang M., Peng G. (2018). An improved approach for gesture recognition. *Chinese Automation Congress*. IEEE, Jinan, China, pp. 4856-4861. <https://doi.org/10.1109/CAC.2017.8243638>
- [8] Huang D. Y., Hu W. C., Chang S. H. (2011). Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination. *Expert Systems with Applications*, pp. 6031-6042. <https://doi.org/10.1016/j.eswa.2010.11.016>
- [9] Li J., Zheng L., Chen Y., et al. (2013). A Real Time Hand Gesture Recognition System Based on the Prior Facial Knowledge and SVM. *Journal of Convergence Information Technology*, pp. 185-193.
- [10] Nagarajan S. S., Subashini T. (2013). Static Hand Gesture Recognition for Sign Language Alphabets using Edge Oriented Histogram and Multi Class SVM. *International Journal of Computer Applications*, pp. 28-35. <https://doi.org/10.5120/14106-2145>
- [11] Itkarkar R. R., Nandi A. V. (2017). A survey of 2D and 3D imaging used in hand gesture recognition for human-computer interaction (HCI). *IEEE International Conference on Electrical and Computer Engineering*. IEEE, Pune, India, pp. 188-193. <https://doi.org/10.1109/WIECON-ECE.2016.8009115>
- [12] Zhao S., Zhang Y., Zhou B., Ma D. (2014). Research on gesture recognition of augmented reality maintenance guiding system based on improved SVM. *International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test and Measurement Technology and Equipment*. International Society for Optics and Photonics, pp. 61-81. <https://doi.org/10.1117/12.2067852>
- [13] Hartmann B., Link N. (2010). Gesture recognition with inertial sensors and optimized DTW prototypes. *IEEE International Conference on Systems Man and Cybernetics*, Istanbul, Turkey, pp. 2102-2109. <https://doi.org/10.1109/ICSMC.2010.5641703>
- [14] Aguilar W. G., Cobeña B., Rodriguez G., et al. (2018). SVM and RGB-D Sensor Based Gesture Recognition for UAV Control. *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, Cham, pp. 713-719. [https://doi.org/10.1007/978-3-319-95282-6\\_50](https://doi.org/10.1007/978-3-319-95282-6_50)
- [15] Altairi ZH, Rahmat RW, Saripan MI, Sulaiman PS. Skin segmentation using YUV and RGB color spaces. *Journal of Information Processing Systems*, 2014, 10(2):283-299.

