2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics

Hand Gesture Recognition Based on HU Moments in Interaction of Virtual Reality

Yun Liu,Yanmin Yin, Shujun Zhang
College of Information Science and Technology
Qingdao University of Science and Technology（QUST）
Qingdao, China
yymhehe@126.com

***Abstract*—Hand gesture recognition is the key technology to realize interaction of virtual reality. The accuracy and stability of the current hand gesture methods are still unsatisfactory due to various reasons. In YC_bC_r color space, the distribution of skin color has a good clustering. The paper uses skin detection in YC_bC_r color space for hand regional image, and proposes a hand gesture recognition method based on *HU* moments which have invariance property of translation, rotation and size. By extracting *HU* moments of the hand contour, this paper applies support vector machine (*SVM*) method to achieve the real-time hand gesture recognition. Experiments prove that the proposed method has very good precision and stability in virtual reality interaction.**

Keywords-Hand gesture recognition; Virtual reality interaction; YC_bC_r color space; HU moments; SVM

I. INTRODUCTION

Computer input devices people daily used usually are mouse and keyboard [1]. With the rapid development of interactive entertainment and augmented reality technology, the interaction becomes more and more extensively applied [2], for example, the application of virtual reality interaction to medical hand rehabilitation training [3]. The traditional interaction way can't meet the increasing requirements of users; therefore development of a convenient and effective way of interaction becomes one of the urgent needs. In recent years, the hand gesture as a natural, intuitive, easy to learn means of human-computer interaction has aroused wide concern among people [4]. The hand gesture itself has diversity and multi-meaning features, and the hand is a complex visual variant. These problems become a difficulty of the hand gesture recognition [5].

II. HAND GESTURE RECOGNITON PROCESS

First, the image is converted to YC_bC_r color space for skin color segmentation [6] after removing the blue background. YC_bC_r color space has a certain resistance on the intensity of the light; hand gesture can be extracted because the skin color has clustering on the C_b, C_r two channel. Then the image noise is removed by morphology. The Laplace operator is used to extract the hand contour. After extracting the *HU* invariant moments of the hand contour, the gesture modeling is trained offline by *SVM*. Then gesture recognition is implemented online using the gesture modeling. The process is illustrated in Fig. 1.

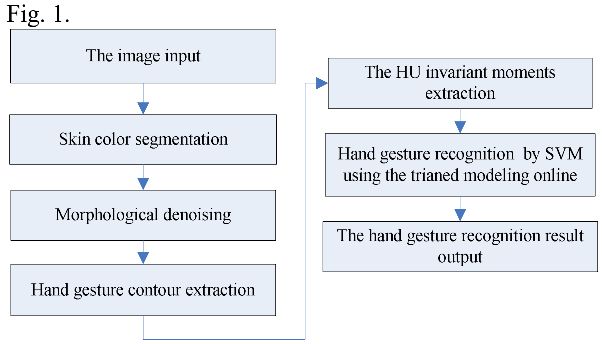


Figure 1. Hand gesture recognition process

III. HAND GESTURE PREPROCESSING

A. *Skin color segmentation*

The paper uses skin detection for hand regional image. According to the research, the effect of skin color segmentation in YC_bC_r color space is better, and the reasons are:

(1) different skin colors in YC_bC_r color space are distributed in a very small range;

(2) the difference in different color in brightness is much higher than the difference in the chroma;

(3) the range in YC_bC_r color space of brightness changes greatly, and the range in YC_bC_r color space of chroma is small;

(4) in YC_bC_r color space, the skin color segmentation results are accurate.

In YC_bC_r color space, the skin color information can be well expressed. The distribution of skin color in YC_bC_r color space showed a good clustering.

The transformation formula between YC_bC_r color space and *RGB* color space is:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cb &= 0.615R - 0.515G - 0.100B \\ Cr &= -0.147R - 0.289G + 0.436B \end{aligned} \tag{1}$$

The C_r, C_b values can eliminate part influence of the light through the analysis of the characteristics of YC_bC_r color space and the experimental verification. The values of C_r, C_b in the

978-0-7695-4721-3/12 $26.00 © 2012 IEEE
DOI 10.1109/IHMSC.2012.42

normal light intensity interval are relatively stable, but in the strong light, the values change. After a lot of experiments, the skin color in the YC_bC_r color space distribution meets $77 \le C_b \le 127, 133 \le C_r \le 170$.

The images after skin color segmentation in YC_bC_r color space still have noise for the bright light. Mathematically morphology can eliminate the noise in the images. The skin color segmentation result and morphological denoising result are as shown in Fig. 2.

(a) (b)

(a) Skin color segmentation result (b) Morphological denoising result

Figure 2. Morphological denoising result

B. *Hand gesture contour extraction*

Laplace Gaussian operator is a kind of second order derivative operator .The Laplace Gaussian operator tests the edge points by looking for the zero-crossing edge of the second order differential in the grey image values. The principle is: through differential operator, the edge formed by gray slow varying form a unimodal function, and the peak value corresponding to the position of edge points; the value at the peak of differential on the unimodal is 0, and on the both sides of the peak the values have opposite sign; the original extreme value point correspond to the zero crossing point of the second differential. The image edges can be extracted by detecting the zero crossing point. The Laplace Gaussian operator of the function is realized by template. The template used in this paper is shown in Fig. 3.

0	-1	0
-1	4	-1
0	-1	0

Figure 3. The Laplace Gaussian operator template

C. *The HU invariant moments*

A group of numbers or symbols (descriptors) are used to represent some certain features of the described object in the image, which is image description. The common image descriptions are geometric descriptor [7], Fourier descriptor, chain code, Hough transform, invariant moment, etc. The invariant moments of the continuous image have been widely used, which can't be impacted by the size, position, direction of the identified target. The feature of the regional information has strong anti-interference ability.

Let $f(x, y)$ is a digital image; the $(p+q)$ order moment is defined as

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \qquad (p,q) = 0,1,2,... \tag{2}$$

The $(p+q)$ order central moment is defined as

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \quad (p,q) = 0,1,2\ldots \tag{3}$$

In which

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \tag{4}$$

$(p+q)$ order normalized central moments defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \quad r = \frac{p+q}{2} \quad p+q = 2,3,4\ldots \tag{5}$$

Constituted by the linear combination of the second and third order central moments, the specific expression of the seven invariant moments is as follows.

$$\begin{aligned}
&\Phi_1 = \eta_{20} + \eta_{02} \\
&\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2 \\
&\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
&\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
&\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} - \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&+(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&\Phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\
&4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
&\Phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&+(3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{6}$$

The *HU* moments have the following physical meaning: the low-order moments describe the overall characteristics of the image, the zero-order moment reflects the target area, the first order moment reflects the target center of mass, second moment reflects the length of principal, auxiliary axis and the orientation angle of principal axis, higher moments describe the details of the image, e.g. the target of the distortion and kurtosis distribution.

The seven invariant moments are fit for description the overall shape of the target, so they have wide applications in edge extraction, image matching and object recognition. The calculation of Hu's seven invariant moments is different and the included information content is not the same. Useful information in the image is generally concentrated on a relatively small amount of computation low order moments. However, high order moments have high computing capacity, and contain some of the details which are susceptible to noise interference. The differences between the high orders moments are not easy to be distinguished for each target.

IV. SUPPORT VECTOR MACHINE(*SVM*)

SVM is based on the theory of statistical learning theory. The method of traditional statistical pattern recognition when is in machine learning emphasizes empirical risk minimization. Simple empirical risk minimization will have "overfitting" problem, and its promotion ability is bad. The promotion ability is correctly predicting the future output though the machine learning (i.e. predicting function, learning function, or learning model).

Using different function as the kernel function *K (x, y)* of the support vector machine can construct to realize the learning machine of the input space of different types of non-linear decision. At present, the commonly used and ideal effect kernel functions are polynomial kernel function, RBF kernel function, Sigmoid kernel function.

Polynomial kernel function:

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d \tag{7}$$

RBF kernel function:

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d \tag{8}$$

Sigmoid kernel function:

$$K(x_i, x_j) = \tanh(scale \times (x_i, x_j) - offset) \tag{9}$$

In this paper, the kernel function is RBF kernel function, the parameter γ=0.5.

V. EXPERIMENTSALS RESULTS AND ANALYSIS

We developed a virtual-reality system using the proposed method in Visual C++6.0 environment with OpenCV library.

In virtual-reality interaction, we put forward the driving engine technology of the mixed reality environment, which integrates the acquisition, modeling and interaction in a set of engine structure, in order to drive different kinds of virtual environment. We have built up a multi-camera acquisition platform, and the engine system is built in Client/Server architecture. The system supports multi-Client real-time acquisition image of the camera angle, which is transmitted over the LAN to the Server. The Server terminal integrates the functions of modeling, recognition, interaction and graphic rendering.

We have built up real-time DreamWorld-QD collecting device without marking objects which is based on multiple cameras, with gathering space for 1 m×1 m ×1 m, as is shown in Fig. 4. On the platform, we choose one of the cameras for hand gesture recognition.

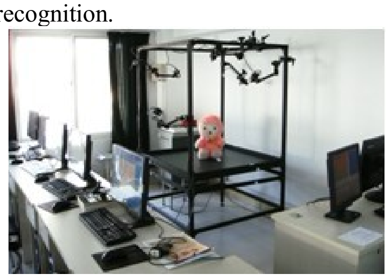
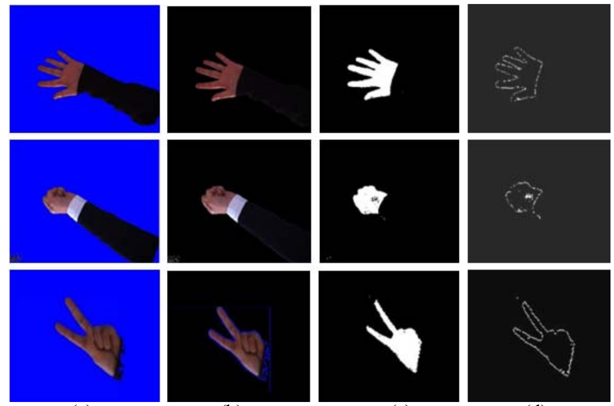


Figure 4. Multi-camera acquisition platform DreamWorld-QD

The results of the image from the platform after skin color segmentation, hand contour extraction are shown in Fig. 5.



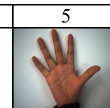
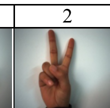
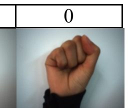
(a) (b) (c) (d)

(a) The original image (b) Removal of the blue background (c) Skin color segmentation (d) Contour extraction

Figure 5. The hand gesture extraction

We define three patterns of hand gesture shown in Table1.To ensure the rate of the hand gesture recognition, we take 1000 images from the platform of different distances and different light intensity as the sample library.

TABLE I. HAND GESTURE DEFINITION

The hand gesture definition	5	2	0
Image			

To gather statistics of the experimental results, more than 1000 images from the platform are used to realize the online hand gesture recognition by *SVM*. The experimental results are shown in Table2. From the results, we know that the hand gesture recognition rate is high. The results are satisfactory.

TABLE II. EXPERIMENTAL RESULTS

The type of hand gesture	5	2	0
Recognition rate	99.42%	98.37%	98.93%

Based on the recognition method, we develop virtual-reality interaction system as shown in Table3. In order to enrich the experimental results, we develop many scenes, such as the Tumbler, Drum, Ball and the book. The interactive scene and hand gesture recognition results are shown in Fig. 6.

VI. CONCLUSIONS

The paper presents a hand gesture recognition method in interaction of virtual reality. Proved by experiments, the hand gesture segmentation in YC_bC_r color space is very accurate. The HU moments of the hand gesture can well characterize the hand geometric features. In the interaction of virtual reality, the results are satisfactory.

VII. ACKNOWLEDGEMENTS

This paper is supported by the National Natural Science Foundation (No. 61040047 & No.60903064), Natural Science

Foundation of Shandong Province (ZR2011FQ003) and the Open Foundation of the State Key Laboratory of Virtual Reality Technology and Systems (BUAA-VR-11KF).

REFERENCES

[1] PAVLOVIC V, SHARMA R, HUANG T S. Visual interpretation of hand gestures for human-computer interaction : a review[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1997, 19(7):677–695,.

[2] Zhang Shujun,Wu Wei.Voxel visual hull parallel optimization modeling method[J].China Joural of image and graphics,2011,16(4):686-69

[3] FANG Yikai, CHENG Jian, WANG Kongqiao, et al. A hand gesture recognition method with fast scale-space feature detection[J]. Journal of Image and Graphics, 2009, 14(2): 214–220.

[4] GU Lizhong. Appearance-based hand gesture recognition and study on human-robot interaction [D]. *Doctoral Dissertation*. Shanghai: Shanghai Jiaotong University, 2008.

[5] Deng-Yuan Huang a,Wu-Chih Hub, Sung-Hsiang Chang . Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination.Expert Systems with Applications 38 (2011) 6031–6042.

[6] Jochen Triesch, Christoph vonder Malsburg , A System for Person-Independent Hand Posture Recognition against Complex Backgrounds , IEEE Transaction on Pattern Analysis and Machine Intelligence , Dec. 2001,Vol . 23, No. 12, pp. 1449–1453.

[7] Hu MK. Visual Pattern Recognition by Moment Invariants[J]. IRE Trans. Information theory, 1962, IT(8) : 179–187.

TABLE III. THE SCHEME OF VIRTUAL ENVIRONMENT AND GESTURES INTERACTION

Virtual object	Gesture recognition	Operation
Tumbler	5	With the palm gesture the tumbler can be pushed down
Drum	0	With the fist gesture the drum can sound
Ball	5	With the palm gesture the ball can be held up
	0	With the fist gesture the ball can stop rising
Book	5	With the palm gesture the book can be opened
	2	With the Scissors gesture the book can be closed

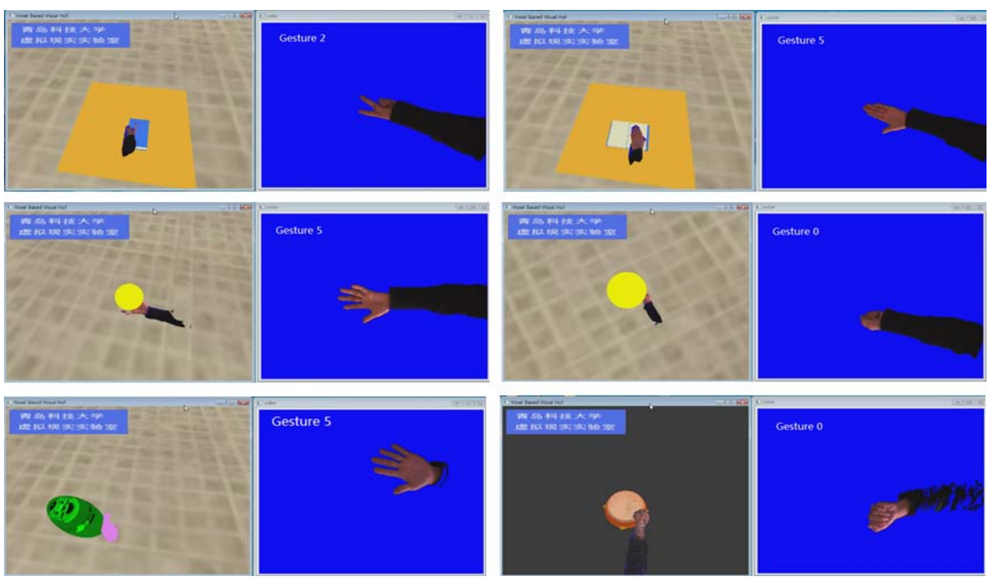


Figure 6. Results of virtual reality interaction