A Method of Sign Language Gesture Recognition Based on Contour Feature

Jingzhong WANG, Meng LI.

***Abstract*—The paper gives a method for sign language gesture recognition using contour features. First, the sign language gesture images pre-processing such as background image difference is given. Then the method of extracting edge features and contour characteristics is introduced. Finally, the match method of weighted features with the sample image feature values is given. Aiming to sign language gestures, create sample library and validate the method with plenty of experiments. The results show that this method can carry out classification efficiently of 30 sign language gestures, and its recognition rate reaches 93%.**

***Index Terms*—sign language gesture recognition, YCrCb skin color detection, contour detection, feature extraction**

I. Introduction

Sign language is the language used by the deaf, and Chinese finger language is one kind of it. Chinese "phonetic alphabet finger chart" has a total of 30 different finger styles, representing all Pinyin. Sign language recognition research can make to improve the lives of the deaf. With the extensive application of computer technology and matures of image processing and pattern recognition technology, the human-computer interaction has become a need of development. Therefore gesture recognition, an integrated multi-product of multi-disciplinary field, has far-reaching significance and application prospects [1] [2].

The edge information is one of the most important features of the image, so that many scholars have come to research gesture recognition based on the edge detection. Some gesture recognition methods use the image edge information directly as a feature [3][4][5], but when the edge detection caused ineffective by light and other effects, the recognition rate will be affected. Therefore rely on the image edge information as the feature is not high stability. Combined with edge detection and image geometry moment method [6], geometric moment features have stability, but the process will lose the original edge information. Gestures contour and convex defects [7] as a feature, it can be applied gesture is limited and requires a gesture parallel to the screen.

This paper researches 30 static gestures' recognition, using image processing and recognition methods, while applying the image edges and contours information and calculating multi-feature matching gesture recognition.

Manuscript received July 8, 2014; revised July 28, 2014. This work was supported in part by China the National Natural Science Funds under Grant No.61371142 and Beijing innovation team building promotion plan under Grant ID HT20130502.

J.-Z. W. Author was with Inner Mongolia University, Inner Mongolia, China. He is now working in the College of Information Engineering, North China University of Technology, Beijing, 100144China (e-mail: jingzhongwang@163.com).

M. L. Author is now with the College of Information Engineering, North China University of Technology, Beijing, zip code 100144 China. (Corresponding author, phone : 135-110-58776; e-mail: 410068528@ qq.com).

II. Feature selection

A. Edge Detection

According to a Gaussian noise model, the desired effect of edge detection operators need to meet three conditions: the first is a low probability of failure, that is a real edge points lost as little as possible, while avoiding as much as possible to retain the pseudo-edge points; second is the high position accuracy, edge detection as close to real edge; third, each point of edge is a single pixel edge. For this edge detection should follow three guidelines:

SNR Criterion

The larger SNR is, the higher the quality of the extracted edge has. SNR is defined as follow:

$$SNR = \frac{\left|\int_{-w}^{w} G(-x)h(x)dx\right|}{d\sqrt{\int_{-w}^{w} h(x)dx}} \quad (1)$$

In the formula, $G(x)$ represents an edge function; $h(x)$ represents the impulse response of the filter when the width is w ; d represents the variance of the Gaussian noise.

Positioning Accuracy Standards

Edge positioning accuracy of L is defined as follows:

$$L = \frac{\left|\int_{-w}^{w} G'(-x)h'(x)dx\right|}{d\sqrt{\int_{-w}^{w} h'^2(x)dx}} \quad (2)$$

In the formula, $G'(x)$ represents a derivative of $G(x)$; $h'(x)$ represents a derivative of $h(x)$; the larger L is, the higher positioning accuracy is.

Single-Edge Response Criteria

The average distance of detection operator impulse response derivative's zero crossing point $D(f')$ should be met:

$$D(f') = p \left\{ \frac{\int_{-\infty}^{+\infty} h'^2(x)dx}{\int_{-\infty}^{+\infty} h''(x)dx} \right\}^{\frac{1}{2}} \quad (3)$$

In the formula, $h''(x)$ is the second derivative of $h(x)$.

Canny designed an edge detection algorithm. Its concrete steps are as follows:

--First, use convolution of 2D Gaussian filter template to smooth the image.

--Second, calculate gradient's magnitude and direction using the differential operator.

--Third, take non-maxima suppression of Gradient amplitude. That is to say, take traversal images and combine with the gradient direction and gradient strength to remove the pseudo edge points around the real edges.

--Fourth, use dual-threshold algorithm to detect and connecting edges. By setting the dual-threshold makes edge coherent [8].

References [9] proposed American Sign Language (ASL) use of calculating the image edge gradient identification algorithm, which compared five kind of edge detection methods including the canny operator, Prewitt operator, Sobel operator, Robert operator and the Laplace operator, proved canny operator is best suited for edge detection and gesture recognition.

B. *Features of Contour*

Geometric characteristics of the image are principal basis for image analysis and recognition. Recognition method based on geometric features can reduce the complexity of the algorithm, save storage space needed and time consuming. Due to ease of access to, storing and processing of binary images, a convenient extraction of the image geometric features usually requires image binary process [10]. Obtain the contour information from binary images, in turn, can calculate the centroid, perimeter, area, and other geometric characteristics, eventually combined with a plurality of features to effectively improve the recognition rate.

Center of Mass[11]

For a continuous image of a 2D, $f(x,y) \geq 0$, p+q-order moments m_{pq} is defined as follows:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(\mathrm{x,y})\,\mathrm{dxdy} \quad (4)$$

In the formula, p and q are non-negative integers, for discretization digital image, above changes into follows:

$$m_{pq} = \sum_{j=1}^{N} \sum_{i=1}^{N} i^p j^q f(\mathrm{i,j}) \quad (5)$$

When the order $\mathrm{p+q} \leq 3$, m_{pq} is a real physical meaning: 0-order moments (m_{00}), that is the mass of the object; 1-order moments (m_{10}, m_{01}) represents the centroid of the object; 2-order moments (m_{20}, m_{02}, m_{11}) represents the radius of rotation; 3-order moments ($m_{30}, m_{03}, m_{12}, m_{21}$) describes the object orientation and slope. These moments have the invariant features in translation, rotation and scaling, so it has a wide range of applications on the image feature extraction and recognition. Obtain the centroid coordinate of images by 0-order moments and 1-order moments:

$$i_c = m_{10} / m_{00},\ j_c = m_{01} / m_{00} \quad (6)$$

Area

Area, a measure of the image total size, is a basic feature to describe the size of the image area, and only has relationship with the edge region, no relationship with the internal gradation value. For binary image, the easiest statistical method of area is measure of the number of pixels in the target profile. Assuming that target area $f(x,y)$'s size is $M \times N$, and a logical 1 indicates target portion, logic 0 indicates the background portion, the area is calculated as follows:

$$Area = \sum_{x=1}^{M} \sum_{y=1}^{N} f(x,y) \quad (7)$$

III. GESTURE RECOGNITION SYSTEM

The process framework of gesture recognition system implementation is showed in Figure 1:

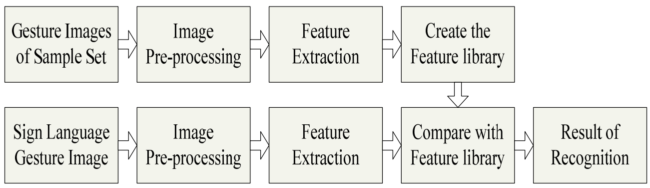

Fig. 1. Gesture recognition system flowchart

As shown in Figure 1, the design of sign language recognition system is divided into two parts.

For the first part: create the gesture samples, images preprocess and feature extraction to the sample set, and finally construct a gesture feature library.

The second part is built on the basis of the first part, for a single gesture image. Gesture image through the same image preprocessing and feature extraction, and then make the gesture feature compare with gesture feature library to obtain the recognition result.

A. *Images pre-processing*

The purpose of gesture image preprocessing is to detect hand position of the image from camera, thus can obtain gesture area of the image as a target area. Its flowchart has been shown in Figure 2:

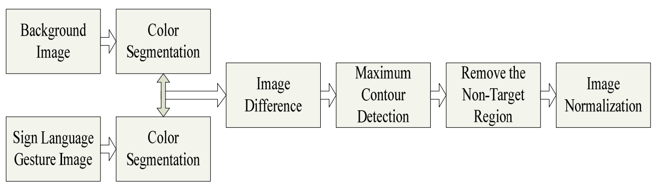

Fig. 2. Gesture recognition system framework flowchart.

ISBN: 978-988-19252-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

--First, skin color detection for both the background image which was saved before and the collected gesture image by YCrCb color space image segmentation. For color detection, a common approach is changing the image from RGB color space to HSV color space [12] or YUV color space [13]. Due to the HSV color space segmentation is very sensitive to light, it is better to choose YUV color space to detection segmentation.

--Second, after skin segmentation, process the images by image difference and binarization, thereby removing the influence of background. As a result, obtain the binary image of skin color region except for the background.

--Third, morphological dilation. That is to eliminate noise and smooth image.

--Fourth, maximum contour detection. According to the outline information from obtained binary image, compare with the contained area of each contour shape and elect the outline with the largest area, then record its boundary information.

--Fifth, obtain the target area. According to the boundary conditions on the step of recording, remove the area outside the boundary and obtain the hand area. Both original and binary gesture images need to area division. According to the direction of gesture, remove the arm region by using improved method of projection. According to the position of inflection points, remove the upper arm. Then find the wrist position using the horizontal projection or vertical projection to cut the lower arm.

--Sixth, image reduction normalized. Due to the size of each hand region is not same, each of the resulting images should be normalized to a bitmap which size is 64 * 64. That is can reduce the amount of image data and redundancy; enhance its smoothness and sharpness for further processing.

--Seventh, gray processing. After the image pre-processing, obtain the gray-level image which size is 64 * 64, as shown in Figure 3:

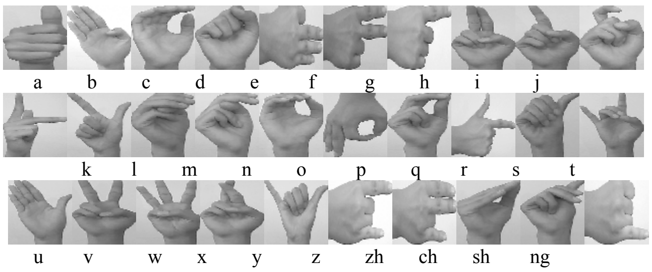

Fig. 3. Gesture images after image preprocessing

After the process of image preprocessing, obtain a gray-level image and a binary hand-type region image which size is 64 * 64. The gray-level images is used to edge detection of feature extraction, contour extraction of binary image is used to calculate the geometric characteristics.

B. Feature Extraction

Features of this paper researches are based on the contour and the edge information. Calculate the eigenvalues of sample images to generate the corresponding feature library that edge feature library (denoted as SL1) and geometric features of the library (denoted as SL2), calculated as follows:

Based on Edge Information

Process the gray-level image after image preprocessing by canny edge detection and 3 * 3 area template dilation twice, making the boundary point spread to the neighborhood. The expanded images as an edge template, which is an edge feature library SL1 of gesture feature library. As shown in Figure 4 is a partial edge feature database:

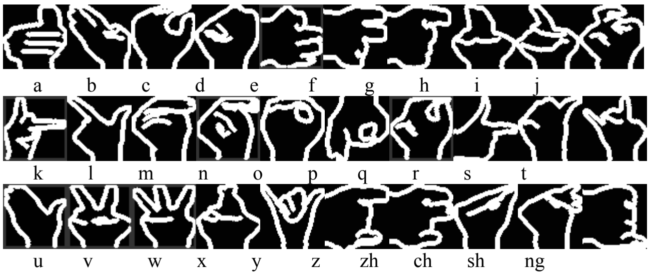

Fig. 4. Partial gesture images of edge feature library

Have a logical AND operation between images after edge detection in test set (denoted as $f_{\text{test}}(x,y)$) and edge template images in feature library (denoted as $f_{sl1}\left(x,y\right)$):

$$g(x,y) = f_{test}(x,y) \wedge f_{sl1}\left(x,y\right) \quad (8)$$

In the formula, $g(x,y)$ represents two-dimensional function of the image result of the operation.

The intersection of two sub-images can be obtained by the logical AND operation, calculate the coincide proportion between image of test set and the edge of the sample set image as a feature (denoted as f_1), the formula is defined as follows:

$$f_1 = \frac{\sum_{y=0}^{64}\sum_{x=0}^{64} g(x,y)}{\sum_{y=0}^{64}\sum_{x=0}^{64} f_{\text{test}}(x,y)} \quad (9)$$

Based on contour information

--First, obtain the contour of binary image after image preprocessing.

--Second, calculate the gesture images' geometrical feature set (denoted as S), including: region area contained by contour(denoted as area); the centroid coordinate of the image(denoted as pc(x, y)); the longest distance from the centroid to point on the contour as radius (denoted as R）; the distance from centroid of the contour to the horizontal secant(denoted as d1); the distance from centroid point to the left contour point at the same height of the centroid (denoted as d2); the longitudinal distance from the centroid to the point at above section contour (denoted as d3). Above steps to sign O as an example, shown in Figure 5:

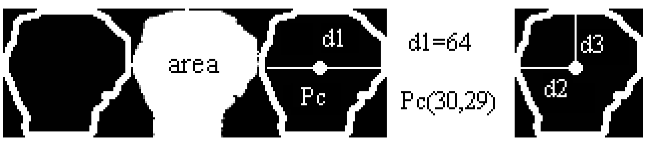

Fig. 5. The process of geometric feature extraction

Following the steps above for all the samples, write the eigenvalues of S in a XML file as the geometric feature database (denoted as SL2). According to the test set, calculate the geometric eigenvalues using the same way and obtain a test feature set (denoted as Stest). Then calculate the eigenvalue distance from the test image to each image of SL2 one by one, finally obtain the overall spatial distance as the feature2 (denoted as f_2), the equation is defined as follows:

$$f_2=\sqrt{f_{2\_1}^{\ 2}+f_{2\_2}^{\ 2}+f_{2\_3}^{\ 2}+f_{2\_4}^{\ 2}+f_{2\_5}^{\ 2}+f_{2\_6}^{\ 2}} \quad (10)$$

Among the equation,

$$f_{2\_1}=\frac{d(p_{ctest}-p_{csl2})}{64\sqrt{2}}=\frac{\sqrt{(x_{test}-x_{sl2})^2+(y_{test}-y_{sl2})^2}}{64\sqrt{2}} \quad (11)$$

$$f_{2\_2}=\frac{|area_{test}-area_{sl2}|}{64*64} \quad (12)$$

$$f_{2\_3}=\frac{|r_{test}-r_{sl2}|}{64\sqrt{2}} \quad (13)$$

$$f_{2\_4}=\frac{|d1_{test}-d1_{sl2}|}{64} \quad (14)$$

$$f_{2\_5}=\left|\frac{d2_{test}}{d1_{test}}-\frac{d2_{sl2}}{d1_{sl2}}\right| \quad (15)$$

$$f_{2\_6}=\frac{|d3_{test}-d3_{sl2}|}{64} \quad (16)$$

Each eigenvalue has been normalized to a value within the range 0 to 1 in the process of calculation, so that reduce the impact rate.

C. Gesture recognition match

In gesture recognition system, test images calculate the eigenvalues with the corresponding value of feature library by the method introduced before. Then, obtain N (the number of sample in the feature library) matching values (denoted as match). The formula is defined as follows:

$$\text{match}=\max(w_1 f_1+w_2 f_2) \quad (17)$$

In the formula above, w_1 and w_2 are the weight values of the edge feature and geometrical feature respectively, and need to meet $|w_1|+|w_2|=1$. Due to the bigger edge coincidence rate (f_1) is and the smaller the spatial distance (f_2) to the sample is, the closer to the sample the test image is. The parameters should be met $w_1>0, w_2<0$.

In N matching values obtained, the maximum value of the corresponding sample number is result of gesture recognition.

IV. THE EXPERIMENTAL RESULTS

According to 30 alphabet gestures for the gesture recognition system, create three sets of gesture library (each set of images come from different people's hand respectively). There are 450 gesture images totally in three set of gestures library. Each gesture has recorded five letters in a set which size is 640*480 as a bitmap. Among them, choose one set as a sample set and two sets as test set.

According to this algorithm, take a lot of experiments to test their impact on the recognition rate for 30 Pinyin gestures.

A. Gesture Detection

In order to complete the detection of the hand-type region, using a series of image processing algorithm mentioned above. Interception hand-type region accurately is the basis for feature extraction; especially those have complex background or arm in the gesture images. Thus, for the gesture recognition based on vision, accurate detection of a gesture needs to be solved for a long-term. The mainly result images of pre-processing as shown in Figure 6:

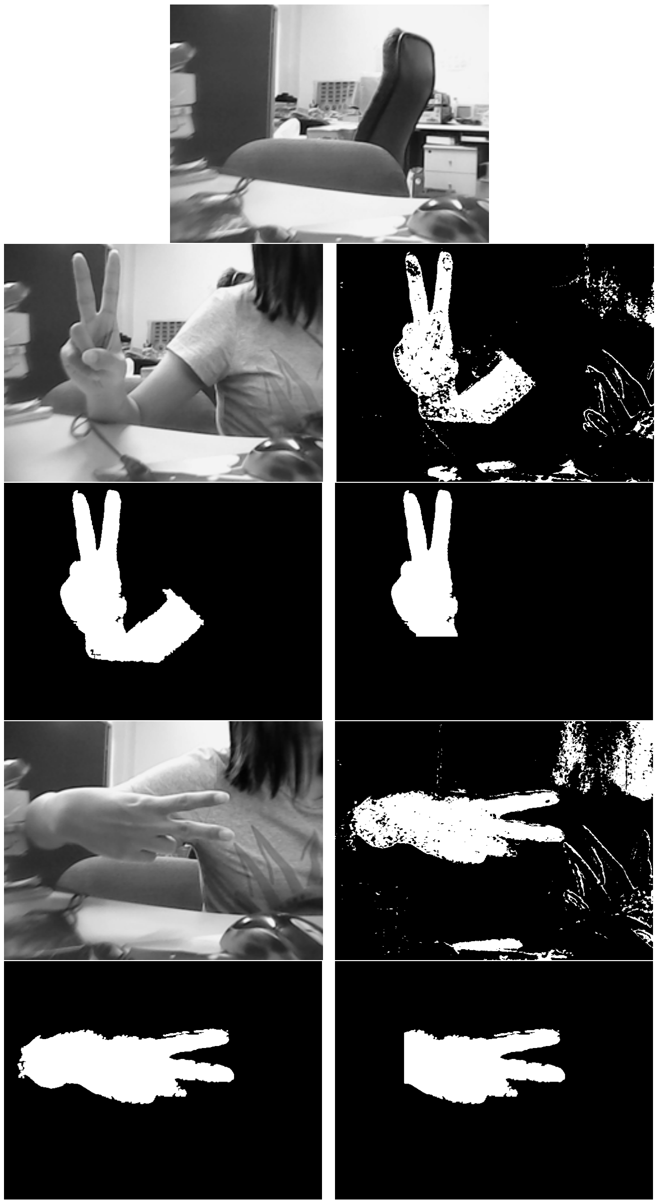

Fig.6. The process of detection of the hand-type region

B. *Logic operation of the edge*

As for using the image edge as a part of features, either to curve fitting or to calculate the Euclidean distance for the edge, have lost its original information. This paper gives a method to have a logical AND operation between gesture edge image and edge image of sample, and use the edge information directly without any loss. Then get the intersect sub-image of two edge images, can intuitive reflect the coincidence of two gestures.

Using the edge image after feature extraction of the test image as the original image and the image of feature library SL1 as the background image ,then have the AND operation between original image and background image respectively. Using gestures U as an example, the following calculation results shown in Figure 7:

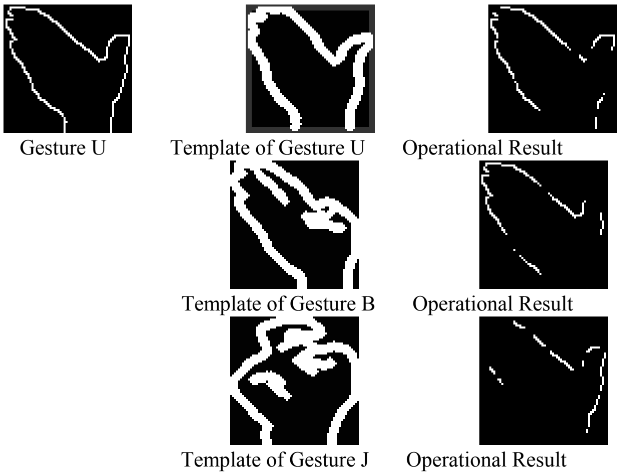

Fig.7. The calculation process of edge coincidence rate

After calculation, the edge's coincidence rate of test image (gesture U) and the template edge corresponding gesture G in feature library of sample is 70.2%. As to the gesture B which is similar to the gesture U, its edge's coincidence rate is 60.8%. As to another gesture such as gesture J, it likes far from gesture U, and its edge's coincidence rate is just 26%. Thus, the operator can do its work for the gesture recognition. For the images which edges are similar to each other, may appear result in similar coincidence rate, so using this method to recognize while calculating the geometric features of images to aid identification.

C. *Gesture Recognition*

From the experiments, in the process of the calculating features, f_1 is larger than f_2 almost an order of magnitude. In order to balance the impact of the value gap between two characteristic values, the corresponding weight in the equation (17) need to meet $|w_1| < |w_2|$. Different weight values has great impact on the recognition rate, therefore, need to determine the optimal weights determined by experiment. Under the different weights, test set gesture recognition rate as shown in Table 1:

TABLE I
RECOGNITION RESULTS OF TEST SETS

ω_1	ω_2	recognition rate of test set 1（%）	recognition rate of test set 2（%）	recognition rate of all test set（%）
0.2	-0.8	90	87.33	88.67
0.25	-0.75	92.67	90.67	91.67
0.3	-0.7	92.67	94	93.33
0.35	-0.65	91.33	92.67	92
0.4	-0.6	91.33	90.67	91

As can be seen from the Table 1, there is the best recognition rate of integral test sets reaches 93.33% when ω_1 = 0.3 and ω_2 = -0.7. For the test set 2, the recognition rate is 94.0%.

Using geometric moments and edge histogram as features to recognize 30 letters gesture, its recognition rate is 90% from the references 6 mentioned. Compared with it, the paper's method improves the gesture recognition rate effectively.

V. CONCLUSION

This paper researches the 30 finger gestures recognition o f Chinese phonetic alphabet, and gives an algorithm based on multi-feature matching, which through a series of processing methods of digital image processing for image feature extraction and classification. Experiments show that this method makes full use of the information of the edge and outline of image, and thus provides an easy and efficient way to the gesture recognition, and it also has stability of the panning and zooming image. The light and excessive rotation gestures has a certain influence on the recognition, just because of this the follow-up research will focus on the adaptive gesture recognition system.

REFERENCES

[1] W. Gao, J. Guo, B.-K. Zeng. The direction and status of the sign language research [J]. Application of Electronic technique, 2002, 11: 000.
[2] X. Wu, Q. Zhang, Y.-X. Xu. An Overview of Hand Gestures Recognition [J]. Electronic Science and Technology, 2013, 26(6): 171-174.
[3] L.-G. Zhang, J.-Q. Wu, W. Gao, et al. Hand Gesture Recognition Based on Hausdorff Distance [J]. Journal of Image and Graphics (A), 2002, 7(11).
[4] Q. Yang, M. Wang. Hand Gesture Recognition Algorithm Based on Euclidean Distance [J]. Microcomputer Information, 2007, 9: 265-266.
[5] L.-J. Sun, L.-C. Zhang. Static Sign Language Recognition Based on Edge Gradient Direction Histogram [J]. MICROEL ECTRONICS & COMPUTER, 2010, 27(3).
[6] Y.-Q. He, Y. Ge, L.-Q. Wang. Gesture Recognition Algorithm Based on Invariant Moment and Edge Detection [J]. Computer Engineering, 2005, 31(15): 165-166.
[7] H.-L. Weng, Y.-W. Zhan. Vision-Based Hand Gesture Recognition with Multiple Cues [J]. Computer Engineering & Science, 2012, 34(2): 123-127.
[8] Canny J. A computational approach to edge detection [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1986 (6): 679-698.
[9] Thepade S D, Kulkarni G, Narkhede A, et al. Sign Language Recognition using Color Means of Gradient Slope Magnitude Edge Images [J]. Intelligent Systems and Signal Processing (ISSP), 2013, 216 – 220.

[10] W.-X. Feng, M. Tang, B. He. Visual C++ digital image pattern recognition technology [M]. China Machine Press, 2010.
[11] WANG Bing, Zhi Qinchuan, Zhang Zhongxuan, et al. Computation of Center of Mass for Gray Level Image Based on Differential Moments Factor[J]. JOURNAL OF COMPUTER-AIDED DESIGN & COMPUTER GRAPHICS, 2004, 16(10): 1360-1365.
[12] Zhang X N, Jiang J, Liang Z H, et al. Skin color enhancement based on favorite skin color in HSV color space [J]. Consumer Electronics, IEEE Transactions on, 2010, 56(3): 1789-1793.
[13] D. Qiu. Research about skin color detection based on HSV and YCrCb color space [J]. Computer Programming Skills & Maintenance, 2012 (10): 74-75.