Hand gesture recognition based on fingertip detection

Guoqing-Meng
College of Electrical and Control Engineering
Xi'an University of Science and Technology
Xi'an, China

Mei-Wang
College of Electrical and Control Engineering
Xi'an University of Science and Technology
Xi'an, China

***Abstract*—To improve the hand gesture recognition accuracy based on Hu moment features, a new recognition algorithm was developed based on the fingertip structure detection. Firstly, the geometric features, the areas of skin region and the image, were designed to segment the skin region from the background in space of hue, saturation and value of brightness. Secondly, the cam point inspections were carried out and the fingertip was detected after the contour approximation. After that, the 7-dimensional feature vector was built. Finally, the distance marching criterion was utilized to recognize the hand gesture. Compared with the Hu moment feature recognition, the developed recognition algorithm improved the recognition accuracy by 2.7%.**

Keywords - Color segmentation; geometric feature; Contour approximation; Fingertip detection; gesture recognition

I. INTRODUCTION

As a new human-computer interaction method based on computer vision, the hand gesture recognition technology is an eye-catching research and application technology [1].

Early gesture recognition used mainly the sensor glove or special signal marked on the hands. Then the purpose of human-computer interaction was achieved by the signal processing. For instance, researchers recognized the gesture by utilizing the instrument gloves [2]. This approach used a few of input data to get the high recognition accuracy with the good real-time property. But it needed the expensive equipments.

Currently, the relative researches concentrate mainly on the study of bare hand recognitions, which includes the gesture segmentation, gesture modeling and gesture recognition. Among them, the gesture model is important for gesture recognition systems. The model selection depends on the specific application. Hand gesture models include mainly the 3D model and the performance-based model. The former model is suitable for all gestures modeling, and the latter is usually only applicable to the communicative gestures. The 3D model has more parameters and its computation is complex. On the other hand, the performance-based model is simple for its calculation and good for the real-time property. Therefore, the most hand gesture recognition systems use the performance-based models at present [3]. There are 4 types of methods for the performance-based modeling. The first method is to build the gesture model by using the 2D gray image itself [4]. The second method is to utilize the deformable 2D template [5]. The third method is based on the attributes of the image [6]. The fourth method is to calculate the image motion parameters so that the parameters of gesture model are extracted [7-8].

This paper utilizes the image attributes and the deformable 2D template to establish the performance-based model. Meanwhile, the Douglas-Peucker algorithm is used to approximate the contour, and the gesture features are obtained.

This paper is organized as follows. In Section 2 the relevant theories and works of the fingertips detection based on the contour approximation and recognition are described. In Section 3 feature extraction method and the hand gesture algorithm are developed. In Section 4 the experiments of the gesture recognition are implemented, and the comparison and analysis are conducted and the conclusion is drawn in Section 5.

II. RELATED WORK

A. Hand Gesture Segmentation in HSV Color Space

The RGB color space is commonly used for image processing. But it is too sensitive to brightness. The correlations between the three-channel R, G and B are too strong. The hue, brightness and saturation are put together in the RBG space so it is difficult to separate them from each other. To avoid the above problem, the HSV color space is used in this paper because the following reasons. Firstly, HSV color space is not sensitive to light; Secondly, HSV color space is a color space with better digital color processing. Most correlations have been removed from this space. So HSV color space has the better performance for the skin clustering, and thus can be used for skin color segmentation with a constant threshold.

HSV color space was created by A. R. Smith according to the color visual characteristics in 1978. It is represented by using hue (H), saturation (S) and value of brightness (V). It belongs to the non-linear color space. It meets the following conditions:

$$0^\circ < H < 360^\circ, 0 < S < 1, 0 < V < 1.$$

According to reference [9], the conversion formula of the RGB space to the HSV space can be described. The three-channel pixel values of the HSV space are denoted by H, S and V.

978-1-4799-2886-6/13 $31.00 © 2013 IEEE
DOI 10.1109/GCIS.2013.23

$$H = \begin{cases} H_1 / 360, B \le G \\ 360 - H_1, otherwise \end{cases} \quad (1)$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (2)$$

$$V = \frac{\max(R,G,B)}{255} \quad (3)$$

where $H_1 = \cos^{-1}\left\{ \frac{0.5[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right.$, the pixel values of the three channels are denoted by R, G and B respectively.

B. Invariant Moments Theory and Parameter Selection

The definition of the moments is deduced from the probability theory, which is an important digital feature. The moment also has the physical significance intuitively. The ($p + q$) order origin moment of gesture image functions $f(x, y)$, is defined as below.

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

where, x^p and y^q is variable kernel. (p,q=0,1,2,…).

If the function $f(x, y)$ of the gesture image is piecewise continuous; meanwhile the function $f(x, y)$ has the non-zero value in a limited section in the plane, then each order moments of the function $f(x, y)$ exists. In order to keep moments with shifting invariance, the coordinates of the centroid is used to construct central moments μ_{pq}.

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y) dx dy \quad (5)$$

where (x, y) is the coordinates of the function $f(x, y)$. The gesture centroid coordinate is denoted by $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$.

To offset the impact of the center scale change, the other order central moments are normalized by zero-order central moment. Normalized central moments are

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \quad (6)$$

where r=1+ ($p + q$)/2.

According to Hu moments theory, we can get 7 features by using a variety of the mathematical combinations of different level geometric moments. When the image is shifting, rotation and scale size changes, the moments have properties to maintain its value unchanged. This seven features are called invariant moments. Since higher order moments not only have large computational work, but also are easily affected by noise, the first four Hu moments, M_1, M_2, M_3 and M4, are taken as features in this paper [10].

$$M_1: \quad \phi_1 = \eta_{20} + \eta_{02} \quad (7)$$

$$M_2: \quad \phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (8)$$

$$M_3: \quad \phi_3 = (\eta_{30} - \eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (9)$$

$$M_4: \quad \phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (10)$$

where □ is the normalized central moments.

C. Contour Approximation

In his paper, we uses Douglas-Peucker algorithm as contour approximation algorithm. According to reference [11], it is often used for curve data compression, and it also is a vertical distance limit method. In the digitization process, we often need to sample and simplify the curve. That is, some points on the curve are sampled to turn the curve into polyline. To some extent, that this algorithm maintains the original shape. So it can obtain a polygon after the polygonal approximation of the contour gesture by using the Douglas-Peucker algorithm, which is conducive to detect the fingertip.

Douglas-Peucker algorithm is described as follows.

Step 1: Calculate the threshold ε which equals the product of the circumference of the contour and a factor. We select that the factor is 0.025.

Step 2: Calculate all the distances from all the points of the reference vertex pairs to the reference line.

Step 3: If there is one distance or there are more distances more than ε, we select the farthest point as the *nth* point, and the reference line is removed.

Step 4: The first point and the *nth* point are marked as the reference vertex pairs. The last point and the *nth* point are also marked as the reference vertex pairs.

Step 5: If all the distances from all points to the reference line are less than or equal ε, all the points are deleted. Only the reference vertex pairs are kept as the control vertices of curve.

Step 6: Repeat Step 2 to Step 5 until all the distances are less than or equal ε.

III. FINGERTIP DETECTION AND HAND GESTURE RECOGNITION

Gesture recognition system mainly includes gesture segmentation and gesture detection, gesture image preprocessing, feature extraction and gesture recognition. The recognition process is conducted directly by comparing the sample features with the features from real-time detection. Firstly, the gesture sample templates are read and preprocessed. These preprocesses include filtering, and HSV color space threshold segmentation and contour extraction. Secondly, the contour is approximated to obtain the polygon. The fingertip is detected. Lastly, the number of the fingertips, Hu moments and other parameters are gotten and saved with the chain form. Each video image is processed in the same way. Furthermore, the features are extracted when the gesture is detected and compared with the sample features. Finally, recognition result can be output.

A. Fingertip Detection and Feature Extraction Method Described

Feature extraction is the first step in recognition process, and it directly affects the result of the recognition. So we need to find the features with the larger differences. Usually the more features selected, the high accuracy of the recognition. But it also brings a large amount of data and the real-time problem, and even leads to the dimensional disaster. Therefore, we want to minimize features when features are selected. Meanwhile, we hope the differences between features large enough so that we can recognize the gesture easily. For these problems, we select 10 gestures as the classes. They are shown in Figure 1.

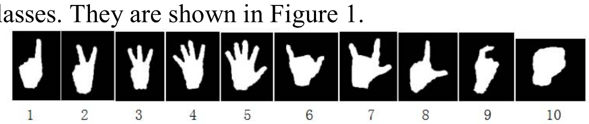

Figure 1. The classes of gestures

In the process of gesture recognition, the fingertip detection is very important. Because gesture transform is mainly achieved by the shape changes of the fingers. If we can accurately detect the fingertips, then it is very conducive to the recognition of the next step.

The developed fingertip feature extraction algorithm is described below.

Step1: Obtain video images by using the camera and detect the hand gesture region in HSV space.

Step2: Extract the gesture contour; calculate gesture gravity and approximate contour polygons.

Step3: Detect convexity defect of contour approximated as the fingertip points of the candidate.

Step4: Find fingertip according to a fingertip filtering rules, and calculate the fingertip number and the finger position.

In addition, the geometry of the gesture has the good abilities to recognize gesture. The geometric features include mainly the relationship between perimeter and area. In this paper, the following two features are used.

(1) Convexity n_1: The outer contour area Oca is divided by the contour convex hull area Ccha.

$$n_1 = Oca \,/\, Ccha \tag{11}$$

(2) Compactness n_2: The squared contour perimeter Scp is divided by the outer contour area Oca.

$$n_2 = Scp^2 / \, Oca \tag{12}$$

These two features have nothing to do with the gesture size, rotation and scaling. So they can be used as the ideal features for gesture recognition.

B. Gesture Recognition Algorithm

The difference D between the compared targets A and B is used to recognize the hand gesture. The difference is shown below.

$$D(A, B) = \sum_{i=1}^{n} \left| \frac{1}{m^{A_i}} - \frac{1}{m^{B_i}} \right| \tag{13}$$

where $m^{A_i} = sign(h^{A_i}) \cdot \log(h^{A_i})$, $m^{B_i} = sign(h^{B_i}) \cdot \log(h^{B_i})$. n is the dimensions of feature vector h^{A_i} and h^{B_i}.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental environment is as follows:

Computer: AMD G4 -3300M;

Processor APU; 1.90GHz;

Windows 7 operating system;

Vc ++ 6.0 and Opencv 1. 0;

The camera resolution is set with 320 × 240, and frame rate is 24fps.

Figure 2 is the different curve approximations of gesture contours by using the Douglas-Peucker algorithm. Figure 3 is the corresponding fingertips detection results.

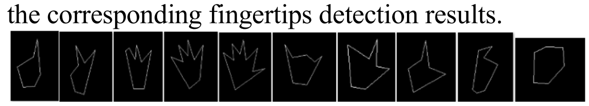
Figure 2. Curve approximations of gesture by using the D-P algorithm

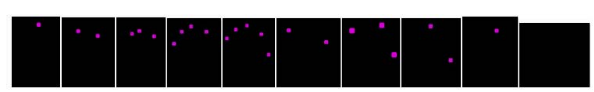
Figure 3. Fingertips detection results

Figure 4 shows the use of the camera capture the results of the gesture recognition, the figure shows real-time gesture bounding rectangle, polygon approximate external, gesture contour, fingertips and other information.

Gesture recognition rate of the gestures is shown in TABLE 1.

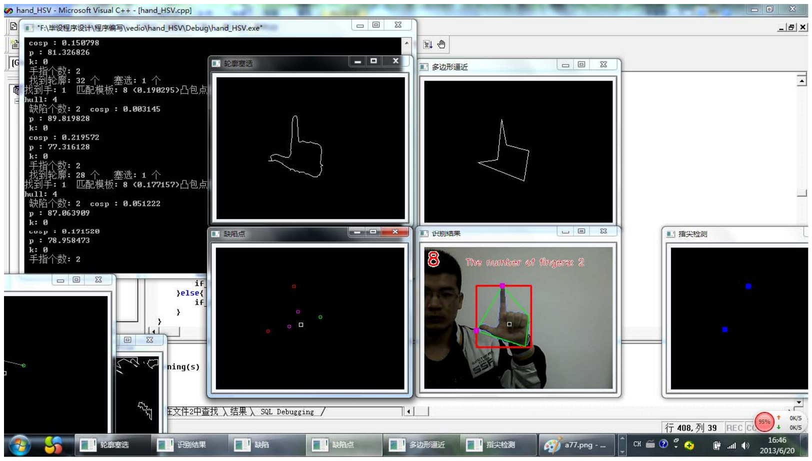

Figure 4. Recognition result of the gesture

TABLE I. GESTURE RECOGNITION RESULTS

Numbers of Gestures and Samples		Recognition Based on Designed Features			Recognition Based on Hu Moment Features		
Gesture No.	***Test No.***	***Number of corrects***	***Number of errors***	***Recognition rate %***	***Number of corrects***	***Number of errors***	***Recognition rate %***
1	100	89	11	89	88	12	88
2	100	91	9	91	89	11	89
3	100	95	5	95	95	5	95
4	100	96	4	96	90	10	90
5	100	95	5	95	93	7	93
6	100	90	10	90	88	12	88
7	100	92	8	92	89	11	89
8	100	93	7	93	90	10	90
9	100	88	11	89	85	15	85
10	100	98	1	99	97	3	97

The recognition accuracy can be calculated according to TABLE 1.

The recognition accuracy of the proposed method is

a_1 = (0.89+0.91+0.95+0.96+0.95+0.90+0.92 +0.93+0.89+0.99)/10
=92.9%

The recognition accuracy of the traditional method is

a_2 = (0.88+0.89+0.95+0.90+0.93+0.88+0.89+0.90 +0.85+0.97)/10
=90.4%

The recognition accuracy is increased by

a_3 = (a1-a2)/a2=(0.929-0.904) /0.904
=2.7%

V. CONCLUSION

Gesture recognition system of detection fingertips is based on contour approximation, not only effectively detect fingertips position and finger numbers, and but also can achieve real-time gesture recognition for 10 classes of gestures. This method is suitable for gesture recognition of monocular vision. Currently, we only select the Hu moments, finger numbers, compactness and convexity as feature parameters, which have robustness for shifting, plane rotation and scaling. The method is simple, and the detection for the fingertip numbers is accuracy. Average recognition rate increased 2.7% by the method compared with that of the direct use of the Hu moments features for recognition.

ACKNOWLEDGMENT

This paper is supported by Scientific Research Foundation for Returned Scholars, Ministry of Education of China ([2011]508) and Natural Science Foundation of Shanxi Province (2011JM8005).

REFERENCES

[1] Junyeong Choi, Jungsik Park,Hanhoon Park and Jong-II Park, “iHand: an interactive bare-hand-based augmented reality interface on commercial mobile phones,” POpt. Eng, 2rd ed., vol. 52, February 2013.

[2] Pradhan G,Prabhakaran B and Chuanjun Li, “Hand-Gesture Computing for the Hearing and Speech Impaired,” MultiMedia, IEEE, 2rd ed., vol. 15, June 2008, pp.20-27.

[3] Ben Abdallah. Mane,Kallel. Mohamed and Bouhlel. Mohamed Salim, “An overview of gesture recognition,” Conference on SETIT 2012,IEEE Computer Society, March 2012, pp. 20-24.

[4] XIE Fang-fang,YANG Wen-fei,CHEN Jing,LI Fang and YU Yue, "A comparison of two methods for model matching algorithm," Information Technology, China, 4rd ed., vol. 10, 2012, pp. 92-95.

[5] Varkonyi-Koczy and Tusor, "Human–Computer Interaction for Smart Environment Applications Using Fuzzy Hand Posture and Gesture Models," IEEE Transactions on Instrumentation and Measurement, 5rd ed., vol. 60, May 2011, pp. 505 -1514.

[6] Chia-Feng Juang,Chia-Ming Chang, Jiu-Rou Wu and Lee D, "Computer Vision-Based Human Body Segmentation and Posture Estimation," IEEE Transactions on Systems, Man and Cybernetics,Part A: Systems and Humans, 1rd ed., vol. 39, January 2009, pp. 119 - 133.

[7] Alon, Jonathan,Athitsos, V., Quan Yuan and Sclaroff, S, "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 9rd ed., vol. 31, 2009, pp. 1685-1699.

[8] Daeho Lee,Kyung Hee Univ, "Vision-based remote control system by motion detection and open finger counting," IEEE Transactions on Consumer Electronics, 4rd ed., vol. 55, 2009, pp. 2308 - 2313.

[9] Lu-Kai, Li Xiaojian and Zhou Jinxiang, "Hand Signal Recognition Based on Skin Color and Edge Outline Examination," Journal of North China University of Technology Beijing China, 3rd ed., vol. 18, 2006, pp. 12-15.

[10] ng Lifeng, Ruan Jun, Ma Qiushi,Wang Lei., "The gesture identification based on invariant moments and SVM," Microcomputer & Its Applications, 6rd ed., vol. 31, 2012, pp. 32- 35.

[11] Zhong Xie,Huimin Wang and Liang Wu, "The improved Douglas-Peucker algorithm based on the contour character," 2011 19th International Conference on Geoinformatics, 2011, pp. 1-5.