Design of control system based on hand gesture recognition

Shining Song
Electricity and Information Institute
Jinan University
Zhuhai, China
240937255@qq.com

Dongsong Yan
School of Electronics and Information Engineering, Jinan University
Rail Transit Research Institute of Jinan University
Zhuhai, China

Yongjun Xie
School of Electronics and Information Engineering, Jinan University
Rail Transit Research Institute of Jinan University
Zhuhai, China
xieyongjun919@jnu.edu.cn

***Abstract*—Vision-based dynamic gesture recognition is an important means of human-computer interaction.** This paper designs a dynamic gesture recognition control system based on USB single camera. After obtaining the image through the camera, the original image is binarized, eroded, smoothed and filled with holes, then the circumscribed rectangles and the centroid points of the gesture image are tracked, then the movement direction of the gesture is determined according to the motion track. Finally, through the gesture control movement of the mouse pointer to move and simulate the left-click operation. The test results show that the system recognition rate is high, therefore the research has some practical values.

Keywords—gesture image; image processing; gesture recognition; motion control

I. INTRODUCTION

The communications of information between people and machines, man-machine interaction are in the continuous development, but the human-computer interaction is still limited by the need of using certain physical equipment to trigger the command, and gradually cannot meet the complexity of many occasions. Therefore, to fulfil the intelligent requirements in interactive process, gesture recognition has become a hot research topic in the field of computer applications.

Gestures occupy an important position in the daily communication of people. They have strong expressive force and impact, which can express information vividly, and contain a large amount of interactive information according to different gestures. Therefore, gestures are very suitable to be the media of human-computer interaction.

At present, the gesture recognition system mainly includes two methods, the system based on a data glove and the system based on computer vision. To operate the former system, the operator needs to wear specialized data glove and position tracker, which brings great inconvenience; while the vision-based gesture recognition system only needs an ordinary camera to obtain the hand gesture picture. Human-computer interaction based on computer vision is still in the theoretical and experimental stages, the number of gestures can be identified is relatively small, which means it is a hot research direction.

II. SYSTEM DESIGN

The research topic of this paper is a control system based on gesture recognition. By calling the camera, four kinds of dynamic gestures can be identified as "up", "down", "left" and "right", and at the same time, the system can realize to control the movement of the mouse pointer and simulate left-click operation.

The system includes four parts: gesture image acquisition, gesture image preprocessing, dynamic gesture recognition and control system. In the gesture image acquisition phase, use OpenCV to call the USB camera to obtain the original image. In the phase of gesture image preprocessing, the color image sequences of the captured gesture images are segmented, and Otsu adaptive threshold segmentation is performed in the Cr channel of YCbCr color space to separate the gesture image from the background. Then the image will be enhanced by smooth and hole filling. In the phase of dynamic gesture recognition, after finding the centroid point of the gesture image, the motion direction is judged according to the motion trajectory of the centroid point. And in the control phase, on the basis of judging the movement direction of the gesture, the system can control the movement of mouse pointer and simulate left-click operation.

System implementation of the algorithm is shown in Fig 1.

ACKNOWLEDGMENT
This research was supported by the Natural Science Foundation of Guangdong Province (NO：2017A030310184), the Guangxi Key Laboratory Open Project Foundation under grant No.16-380-12-012k, the Coordinated Operation Innovation Center Project of Zhuhai Rail Transit.

978-1-5386-5053-0/18/$31.00 ©2018 IEEE

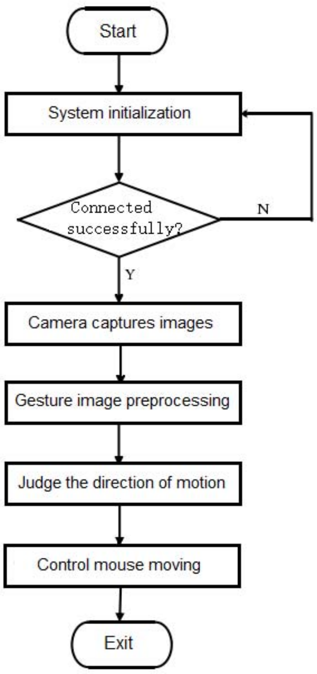

Fig. 1. System flow chart

III. ACQUISITION AND PREPROCESSING OF GESTURE IMAGES

OpenCV library HighGUI module provides multiple functions like reading or writing image files. Using cvCreateCameraCapture () function in HighGUI module to complete the image acquisition with capturing the gesture image from the USB camera, then process the image.

A. Color segmentation

YCbCr is a color coding scheme commonly used in consumer video products such as DVD, camcorder and digital TV. Y is the luminance component, Cb is the blue chrominance component, and Cr is the red chrominance component. YCbCr is a color model obtained by a linear transformation of RGB, which is computationally simple as shown in Equation 1. YCbCr separates luminance information independently, Cb and Cr channels are barely affected by light, and these three channels are separated from each other, therefore it is suitable to achieve color clustering.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1440 \\ -0.1687 & -0.3313 & 0.5000 \\ 0.5000 & -0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Equation 1

B. Gesture image binarization

The color images collected from the camera need to be binarized first to highlight the area of interest, in order to ensure the accuracy of subsequent operations. It is hard to have the best segmentation result in different background because the threshold is set in advance. Therefore, the system in paper adopts the algorithm of Otsu adaptive threshold segmentation in Cr channel.

The main idea of Otsu method is clustering. This algorithm assumes that an image consists of foreground colors and background colors. A parameter is chosen statistically so that this parameter can separate foreground colors from background colors as much as possible. In another words, this method is to divide the gray level of the image into two parts according to the gray level so that the difference of the gray value between the two parts is the largest and the difference of the gray level between each part is the smallest. A suitable gray degree level to divide would be found through the calculation of variance.

The process of Otsu algorithm achieves as following: set t as the segmentation parameter of foreground and background, W0 is the number of foreground points accounted for the image ratio, the average gray value U0; W1 is the background points accounted for the ratio of the image, the average gray value U1. The average gray value of the image is: u = W0 * U0 + W1 * U1. From the minimum grayscale value to the maximum grayscale value over t, when t makes the value g = W0 * (U0-u) 2 + W1 * (U1-u) 2 maximum, t is the optimal parameter for segmentation.

C. Gesture image enhancement

Since other objects in the background will interfere with the image segmentation, some irregular small holes or small protrusions in the edge will exist in the binarized image. Therefore, the result needs to be processed in a mathematical morphology, mainly using open operation to eliminate small bumps and close operation to fill small holes.

Eroding first then expansion is called open operation and it is used to eliminate small objects and separate objects at slender points. It can also remove tiny structures in the image and smooth the boundaries of larger objects without significantly changing its area. Expansion first then eroding is called close operation, which is used to fill small holes in the object and connect neighboring objects and smooth the boundary without significantly changing its area.

In order to improve the quality of the real-time video image acquisition, the image features should be extracted and processed stably, and the noise needs to be removed in order to smooth image. This system adopts a linear Gaussian filter smoothing operation, using the Gaussian filter and each pixel of the input image to complete the convolution calculation, to obtain the final pixel output value .

After Gaussian filtering, some small residual holes are still existing in the binary gesture images, so these gesture images need hole filling process by filling small areas in order to smooth entire gesture image. Use function cvFindContours () to find all the outer and inner contours in the image. When the inner contour area is less than the threshold, filling all the areas enclosed by the inner contour with the function cvDrawContours ().

IV. Dynamic Gesture Recognition and Control

A. Getting the bounding rectangle and centroid point of the gesture image

First get the outline of the gesture image from the gesture image, then capture the circumscribed rectangle of the gesture image, and finally get the center of the gesture image. The algorithm is shown in Fig 2. Set parameter as 6000.

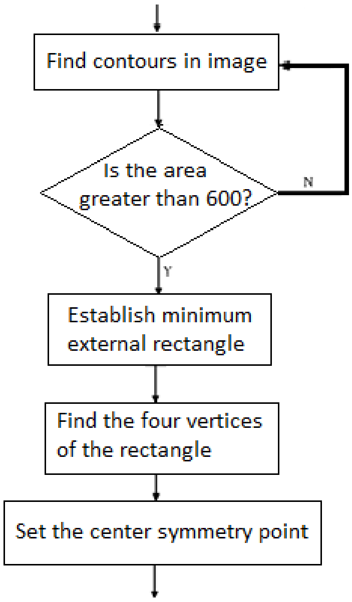

Fig. 2. Algorithm flow chart

B. Dynamic gesture recognition of the direction of movement

After obtaining the centroid of the gesture, analyzing the trajectory of the centroid point can judge the direction of the gesture. The algorithm is to connect the centroid point of each frame of gesture image and use function cvLine () to draw the motion trajectory of centroid point. When the centroid of the gesture moves more than 50 pixels, system starts to judge the direction. First, it is determined whether the direction of motion is horizontal or vertical, then determine the direction is left or right.

Use the judgement of 'right' as an example to illustrate the idea. X is the horizontal displacement of the center of mass point shifted by 50 pixels and y is the vertical displacement of the center of mass point displaced 50 pixels. If abs (x)> abs (y), the movement direction of mass center point is on the horizontal direction. If x> 120, the direction of the gesture is considered as 'right'.

C. Direction and motion control

After recognizing the direction of the gesture movement, the control commands can be issued according to the control directions. To visualize the interface, we draw the motion trajectory of the centroid of the gesture on the screen. In order to adapt to the coordinates of the entire computer screen, the coordinates of the centroid of the gesture (p1.x, p1.y) are scaled up and assigned to the mouse pointer coordinates (mouseInput.mi.dx, mouseInput.mi.dy). Use function SendInput to control the movement of mouse, and output the real-time coordinates of the mouse pointer.

In addition to the control of the movement direction, as well as the action of control, here simply to achieve the left mouse button click on the action control, for example, according to the minimum external rectangle gesture image vertex coordinates, we can get the rectangle long 'imageLength' and wide 'imageWidth'. When imageLength / imageWidth≤1.1 while the gesture is the making a fist, the output mouse action is 'left-click'. Under other condition, the output of mouse action is 'no action'.

V. System Implementation

The system development environment uses Visual Studio 2012 MFC programming interface design, with OpenCV 2.4.9 library. Camera resolution is 640 * 480 and frame rate is 30fps.

While the system is running, click on the 'open camera' button, the system will be ready to operate. Click the 'Start / Pause' button to start gesture recognition and mouse virtual control. The recognized movement direction is indicated by an arrow in the upper left corner of the 'video window', and the text 'up', 'down', 'left' and 'right' would be output in the recognition result to indicate the movement direction. In the same time, the mouse pointer coordinate action would be output as 'left-click' or 'no action'. The system interfaces are shown in Fig.3 and Fig.4.

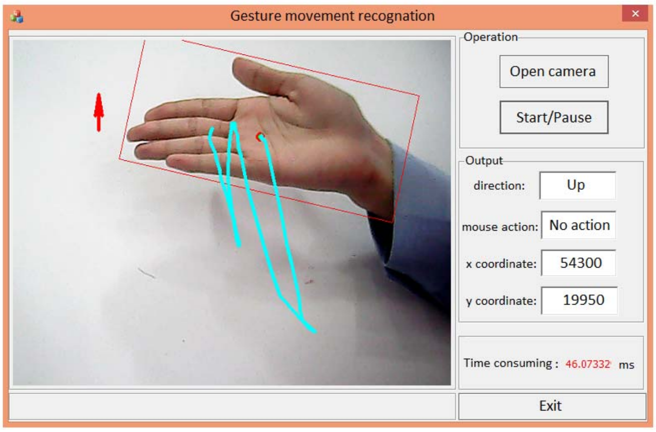

Fig. 3 . Control result is: up, no action

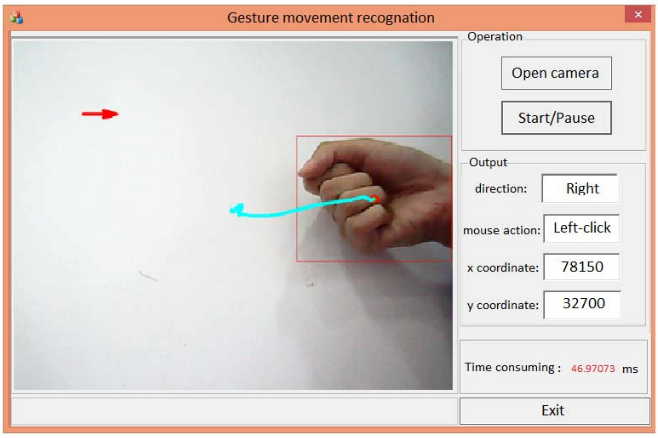

Fig. 4. Control result is: right, left click

In order to test the recognition effect, three testers were selected to do the experiment. Each gesture has done 30 times per person, and the experimental results were obtained under single background and complex background respectively.

TABLE I. EXPERIMENTAL RESULTS UNDER SINGLE BACKGROUND

Action	Up	Down	Left	Right	No action	Left-click
Correct number	86	87	88	88	83	86
Recognition rate	95.6%	96.7%	97.8%	97.8%	92.2%	95.6%
Average recognition rate	95.95%					

TABLE II. EXPERIMENTAL RESULTS UNDER COMPLEX BACKGROUND

Action	Up	Down	Left	Right	No action	Left-click
Correct number	73	75	80	82	70	74
Recognition rate	81.1%	83.3%	88.9%	91.1%	77.8%	82.2%
Average recognition rate	84.07%					

The single background in the experiment is a white wall, which has a strong difference compared with hand. The complex background in the experiment is in a library, there are many students moving around. As we can see from the data in the two tables above, the overall recognition rate in single background is higher than that in complex background; the recognition rate in the horizontal direction is slightly higher than that in the vertical direction, because the experimenters are accustomed to use right hand, so they have better control while moving. The recognition rate of the operation 'left click' is higher than 'no action'.

The experimental results show that in one hand, the recognition rate of the four moving directions of the dynamic gesture reaches more than 80%, and the recognition speed is fast, and the virtual control of the mouse can be well achieved. In the other hand, the recognition adaptability in the complicated background needs to be further improved. Therefore, the gesture movement recognition system has accomplished the purpose of this application research.

VI. SUMMARY

This gesture movement recognition system uses common USB camera and the method of skin color segmentation to get the image, and the input image is converted from RGB space to YCbCr space with Otsu algorithm for threshold segmentation on the Cr channel of skin color clustering to get a better segmentation result. In the process of image preprocessing, filling small area algorithm is used to improve segmentation result by filling small outline inside the gesture image with white, so as to solve the problem of the hole in the gesture image and improve the recognition accuracy of the system, which help realize the gesture recognition and mouse virtual control system. This system achieves using gestures instead of the mouse clicking, pointer movement and other virtual operations. Compared with the data glove and multi-function camera, the required equipment is simple and low cost, which is in favor of gesture recognition application and promotion.

The development of human-computer interaction has been developed increasingly. The combination of gesture recognition and mouse control will make people's lives more convenient and intelligent, therefore this gesture movement recognition system has huge practicality and space for further development.

REFERENCES

[1] Jibin Fu, Xin Bai, Baode Ju. A video target detection algorithm based on pixels texture correlation background model[C]. IEEE, Beijing China: Intelligent · Visual Surveillance (IVS), 2011 Third Chinese Conference on, 2011: 61-64.

[2] Lu Yishu, Research of Vision Based Gesture Recognition, [D]. Tianjin Industry University, 2016.

[3] Cao Liang, Image Segmentation Based on Otsu Algorithm,[D].Wuhan University of Technology, 2008.

[4] Qiu Di, Study of skin color detection based on HSV and YCrCb color space[J]. Computer programming skills and maintenance, 2012(10): 74-75..

[5] Zhang Ruoyu, Research and Implementation of Key Algorithms of Gesture Recognition in Human - Computer Interaction.[D]. Guangdong University of Technology, 2014.