Robust Gesture Recognition Based on Embedded System

Fei Wang[1], Yanli Dong[2], Hui Yang[3], Feiyan Zhao[4]

Zhejiang University Kunshan Innovation Institute, Kunshan 215300, China

Abstract— *To improve portability of current gesture recognition technology, this paper proposes a method of the design and implementation of gesture recognition based on embedded system by combining embedded platform with Open CV. Firstly, image denoising using the weighted mean filtering method, and the morphological open operation is used to obtain the complete contour, and the edge extraction is carried out by the improved Canny operator; Secondly, image segmentation using two-dimensional maximum interclass variance. Then analysis and processing of specifies the feature information of gestures images using Hu distance. Finally, the template matching method is used to recognize the gesture image. The experimental results show that the gesture recognition based on embedded system has strong portability, higher recognition rate and easy implementation.*

Keywords— Embedded, template matching, human-computer interaction, Hu distance.

I. INTRODUCTION

Gesture recognition is one of the most intuitive forms of human-computer interface. It has gradually become a hot spot in many scientific fields. At present, the research on gesture recognition technology is mainly focused on the two aspects of vision and sensor hardware [1,2,3]. Gesture recognition based on sensor acceleration hardware requires the help of some hardware equipment to complete the operation; such equipment is expensive, not easy to wear. Furthermore, that needs to be used in a specific situation, the limitations are very large [4,5].

For gesture recognition based on visual information, it is only necessary to complete the communication between man and machine through human gesture action without intermediate media. It makes gesture recognition convenient and effective because of no external device. And then it brings users a new interactive experience and freedom. Dynamic gestures are one of the most intuitive and effective approach for human-computer interaction [6,7,8,9]. Therefore, we present gesture recognition based on embedded system in order to better recognize and obtain gesture information. The gestures images are acquired by camera are preprocessed, and then the gesture features are extracted and recognized.

II. IMAGE PREPROCESSING

2.1 Gesture image denoising

The weighted mean filtering algorithm is also called linear filtering, which is neighborhood averaging method that is using the average value of several pixels gray to replace the gray level of each pixel [10]. General definition of weight filtering is defined as follow:

$$W_{mean}\{I(x,y)\} = \sum_{(x,y)\in\Omega i,j} w(x,y) \times I(x,y) \tag{1}$$

Then the weight value in the filter template is the reciprocal of the Euclidean distance, which is defined as follow:

$$w_0(x,y) = 1/\sqrt{\sum_{(x,y)\in\Omega i,j}[(x-x_0)^2 + (y-y_0)^2]} \tag{2}$$

Where (x_0, y_0) is the position coordinate of the central pixel point. Then the Weight normalized is defined as follow:

$$w(x,y) = w_0(x,y)/\sum_{x,y\in\Omega i,j} w_0(x,y) \tag{3}$$

2.2 Morphological image processing

The morphological opening operation is used to extract the components of the gesture image, which effectively preserves the details of the original image and the basic shape features. The opening operation is defined as follow:

$$A \circ B = (A\Theta B) \oplus B \tag{4}$$

A is defined target image. B is defined structural element. Operation can separate some adhesion targets, which is very significant for removing salt and pepper noise.

2.3 Image threshold segmentation

The maximum between-Cluster variance is derived based on the principle of decision analysis or least square method, which is proposed by the great progress of Japan, also known as the Otsu threshold [11,12]. The maximum between-Cluster variance is defined as follow:

$$\sigma_T^2 = w_A w_B (u_A - u_B)^2 \tag{5}$$

Changing the t value until σ_T^2 is the maximum, then T is the optimal segmentation threshold.

2.4 Image edge detection

For a good edge detection algorithm, it should have three basic conditions: high positioning accuracy, low error rate and suppression of false edges. Candy operator proposed three optimization criteria for these three conditions [13]: maximum noise ratio criterion, optimal zero-crossing positioning criterion and multi-peak response criterion. The implementation steps are as follows:

- Image denoising by Gaussian filter;
- Using the first order partial derivative finite difference to find the direction and amplitude of the gradient; The function is defined as follow:

$$F(x,y) = \sqrt{F_1^2(x,y) + F_2^2(x,y)} \tag{6}$$

$$G_F = arctan\frac{F_2(x,y)}{F_1(x,y)} \tag{7}$$

Where F is amplitude of the image gradient. G_F is Direction of the image.

- Finding the maximum of the image gradient;
- Using double threshold algorithm to discriminate and connect for the edges.

A large amount of noise can be effectively eliminated and the edge contrast of the image can be enhanced by improving the original operator.

III. GESTURE FEATURE EXTRACTION AND RECOGNITION

3.1 Hu invariant moment

The image is preprocessed by the above algorithm to extract the features of the image. The Hu invariant moments are used to extract the features of gesture images since the Hu invariant moments have the invariance of translation, rotation and scaling [14]. Calculating the similarity of the contour between the two images by Hu invariant moment, then the value with high degree of acquaintance is chosen as the recognition criterion. Feature extraction using Hu distance can ensure that some feature quantities of the image remain unchanged in size, position and direction, and then improve the accuracy of gesture recognition [15,16]. The seven moment invariants are used to describe the important features in the image that want to extract the target region. Then the functions are defined as follow:

$$\begin{aligned}
\omega_1 &= \upsilon_{20} + \upsilon_{02} \\
\omega_2 &= (\upsilon_{20} - \upsilon_{02})^2 + 4\upsilon_{11}^2 \\
\omega_3 &= (\upsilon_{30} - 3\upsilon_{12})^2 + (3\upsilon_{21} - \upsilon_{03})^2 \\
\omega_4 &= (\upsilon_{30} + \upsilon_{12})^2 + (\upsilon_{21} + \upsilon_{03})^2 \\
\omega_5 &= (\upsilon_{30} - 3\upsilon_{12})(\upsilon_{30} + \upsilon_{21})[(\upsilon_{30} + \upsilon_{12})^2 - 3(\upsilon_{21} - \upsilon_{03})^2] \\
&+(3\upsilon_{21} - \upsilon_{03})(\upsilon_{21} + \upsilon_{03})[3(\upsilon_{30} + \upsilon_{12})^2 - (\upsilon_{21} + \upsilon_{03})^2] \\
\omega_6 &= (\upsilon_{20} - \upsilon_{02})[(\upsilon_{30} + \upsilon_{21})^2 - (\upsilon_{21} + \upsilon_{03})^2] + 4\upsilon_{11}(\upsilon_{30} + \upsilon_{12})(\upsilon_{21} + \upsilon_{03}) \\
\omega_7 &= (3\upsilon_{21} - \upsilon_{03})(\upsilon_{30} + \upsilon_{12})[(\upsilon_{30} + \upsilon_{12})^2 - 3(\upsilon_{21} + \upsilon_{03})^2] \\
&-(\upsilon_{30} - 3\upsilon_{12})(\upsilon_{21} + \upsilon_{03})[3(\upsilon_{30} + \upsilon_{12})^2 - (\upsilon_{21} + \upsilon_{03})^2]
\end{aligned} \tag{8}$$

3.2 BP neural network

BP neural network [17] is a kind of multi-layer feed forward neural network. Then the threshold and weights are trained by error propagation algorithm to reduce the error between layers and realized the given mapping relationships before and after. The structures consist of an input layer, one or more hidden layers, and an output layer. The network structure is shown in fig 1. The learning process consists of a forward and backward propagation process. Forward propagation is that the input information is processed layer by layer from the input layer to one or more hidden layers, and then to the output layer. Backpropagation is that the error of the output is returned according to the original path when the expected output value is not reached, then the weight between the layers is calculated layer by layer to minimize the error. Gradient descent algorithm is used to train the network. The Sigmoid function is made as the activation function in order to ensure the differentiability of the function.

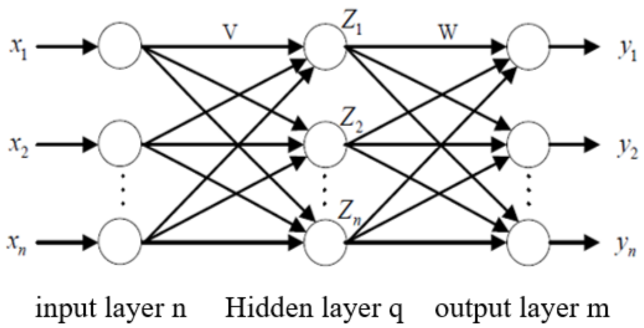

FIGURE 1: BP neural network

IV. SYSTEM DESIGN AND REALIZATION

4.1 Introduction to Hardware and Software

4.1.1 System software environment

1) PC operating system: Fedora14, Windows XP.
2) Virtual machines: VMware-6.5.1.
3) Graphic interface development tools: Qtopia2.2.0.
4) Open source computer visual library: OpenCV2.4.3.

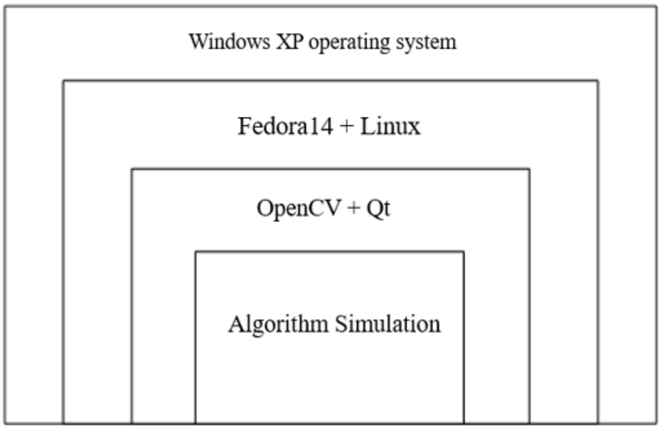

FIGURE 2. The system software diagram

4.1.2 System hardware environment

1) PC hardware configuration: core i5CPU, 2G running memory, 3.19 GHz main frequency.
2) ARM development board configuration: CPU processor Samsung s3c2440A, master frequency 400 MHz, 64M SDRAM memory, 3.5 inch touch screen.
3) Camera configuration: USB2.0 interface.
4) FLASH storage :256 M/1GB Nand Flash, 2M Nor Flash.

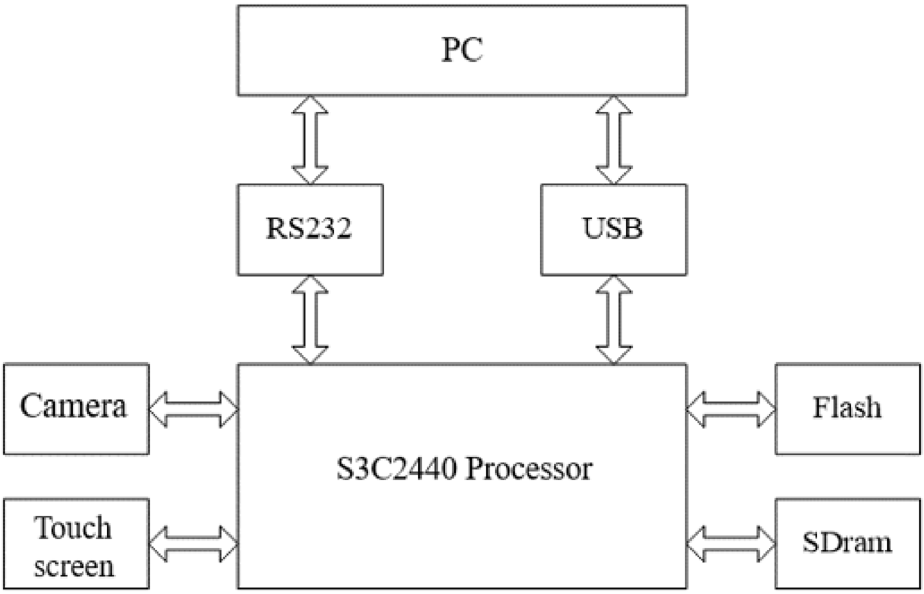

FIGURE 3. The system hardware diagram

4.2 Gesture recognition experiment and analysis

In this study, The Hu invariant moments of the gesture template are compared with the Hu invariant moments of the current acquisition gesture from the camera, then the similar Hu invariant moments value is used as the criterion of gesture recognition. Fig 4 is a template for ten kinds of gesture images. Fig 5 is the contour image of ten kinds of gesture templates.

Fig. 1 **Fig. 2** **Fig. 3** **Fig. 4** **Fig. 5**

Fig. 6 **Fig. 7** **Fig. 8** **Fig. 9** **Fig. 10**

FIGURE 4: Ten kinds of gesture templates

Fig. 1 **Fig. 2** **Fig. 3** **Fig. 4** **Fig. 5**

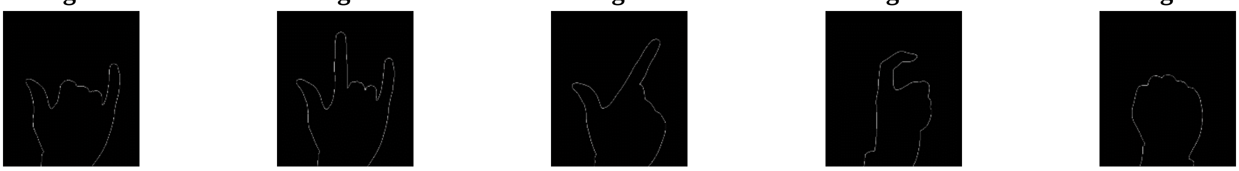

Fig. 6 **Fig. 7** **Fig. 8** **Fig. 9** **Fig. 10**

FIGURE 5: Ten kinds of gesture contours

In both simple and complex background, the gesture video is detected and recognized by the embedded development board, and then the number of gestures is given. This paper lists three kinds of gesture recognition effects: three, five, eight as shown in the following fig 6,7:

International Journal of Engineering Research & Science (IJOER) ISSN: [2395-6992] [Vol-6, Issue-5, May- 2020]

FIGURE 6: Under the simple background gestures rendering

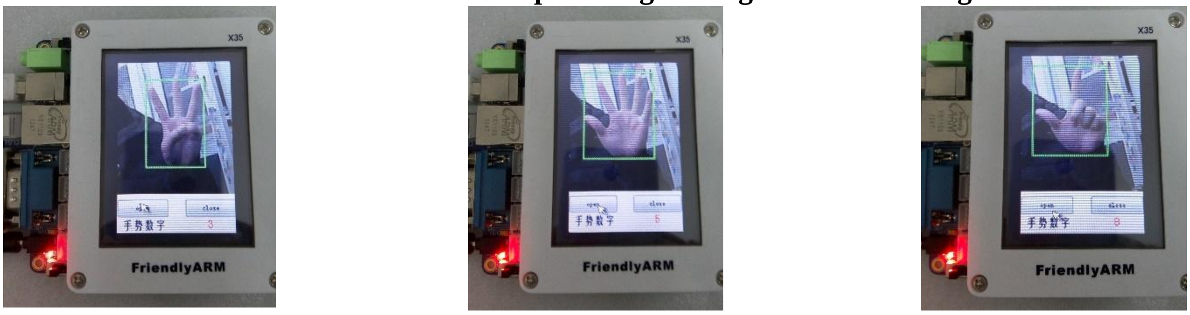

FIGURE 7: Under complicated background rendering

One hundred images were randomly collected for each gesture to identify in order to verify the effectiveness and practicability of the gesture recognition system. Then the recognition rate of each gesture before and after the improvement of the BP neural network algorithm was obtained.

TABLE 1
TEN KINDS OF GESTURE RECOGNITION RATE UNDER DIFFERENT BACKGROUND

Data set	Original algorithm		Improved algorithm	
	In a simple background	In a complex background	In a simple background	In a complex background
1	83%	75%	96%	92%
2	79%	70%	93%	90%
3	81%	73%	90%	87%
4	76%	69%	93%	88%
5	79%	75%	96%	91%
6	80%	70%	92%	89%
7	78%	68%	91%	86%
8	85%	74%	94%	90%
9	73%	65%	90%	85%
10	86%	80%	97%	93%

From Table1, it can be seen that the gesture recognition rate in different backgrounds is increased by 16% and 23% respectively compared with the previous one by improving the algorithm. The gesture recognition rates are also different in different background. The recognition rate is higher in simple background. The recognition rate is reduced in complex background. The different light intensity and the similarity of each gesture cause the final effect of gesture recognition. When extracting gesture features in a complex background, it will affect the Hu moments value of the gesture, resulting in a large difference between the Hu moments value and the template, which will affect the recognition rate of the whole system. The characteristic value and quality of various gesture images are so stable that the recognition rate of gesture system is higher in simple background.

To sum up, through a large number of experiments, it can be seen that the whole gesture recognition system has a good recognition effect, which can meet the basic application requirements and also confirms the reliability and practicability of the gesture recognition system.

V. CONCLUSION

At present, many gesture recognition techniques have been applied to PC machines, resulting in great limitations, poor portability and practicability. Therefore, this paper realizes the embedded gesture recognition system by using the combination of embedded system and image processing technology based on the existing gesture recognition technology. Video stream is collected by USB camera, then different gesture commands are obtained to achieve human-computer interaction through gesture processing in the video stream. Finally, the whole gesture recognition system is verified to ensure that the whole system is real-time and effective.

REFERENCES

[1] V. Pavlovic, R. Sharma, T. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction[J]. IEEE Transition on Pattern Analysis and Machine Intelligence. 1997, 9(7): 677-695.

[2] Y. Wu, T. Huan. Vision-based Gesture Recognition[A]. France: Proceedings of the International Gesture Recognition Workshop. 2009, 4(9): 73-79.

[3] Hardenberg CV, Berard F, Bare-Hand human-computer interaction[C]. In: Proc. of the ACM workshop on Perceptual User Interface, 2001: 15-17.

[4] RenqiangXie, Juncheng Cao. Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching[J]. IEEE Sensors Journal. 2016, 4537-4545.

[5] S. Hazra, A. Santra. Robust Gesture Recognition Using Millimetric-Wave Radar System[J]. IEEE Sensors Letters. 2018,1-4.

[6] W. Fang, Y. Ding, F. Zhang and J. Sheng. Gesture Recognition Based on CNN and DCGAN for Calculation and Text Output[J]. IEEE Access. 2019,28230-28237.

[7] G. Plouffe and A. Cretu. Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping[J]. IEEE Transactions on Instrumentation and Measurement. 2016, 305-316.

[8] R. Xie, X. Sun, X. Xia and J. Cao. Similarity Matching-Based Extensible Hand Gesture Recognition[J]. IEEE Sensors Journal. 2015, 3475-3483.

[9] G. Li, H. Wu, G. Jiang, S. Xu and H. Liu. Dynamic Gesture Recognition in the Internet of Things[J]. IEEE Access.2019, 23713-23724.

[10] NeelaChithirala, Natasha. B, Rubini. N. Weighted Mean Filter for Removal of High Density Salt and Pepper Noise[J]. International Conference on Advanced Computing and Communication Systems. 2016,22-23.

[11] Wesley E. Snyder and Hairong Qi, Machine vision[M]. China Machine Press. 2005.

[12] A. Thayananthan, B Stenger, PHS. Torr, R Cipolla. Shape Context and Chamfer Matching in Cluttered Scenes[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recogniyion.2003: 1-8.

[13] A P Tzannes, D H Brooks. Detection of point targets in image sequences by hypothesis testing a temporal test first approach[J]. IEEE. Proc. 2000, 39(39): 2270-2278.

[14] SouvikHazra, AvikSantra. Robust Gesture Recognition Using Millimetric-Wave Radar System[J]. IEEE. 2017.

[15] S Mckenna, S Jabri, Z Duric, A Rosenfeld, H Wechsler. Tracking groups of people activities[J]. IEEE Trans. On Pattern Analysis and Machine Intelligence. 2002, 22(8): 809-830.

[16] S Mckenna, S Jabri, Z Duric, A Rosenfeld, H Wechsler. Tracking groups of people activities[J]. IEEE Trans. On Pattern Analysis and Machine Intelligence. 2002, 22(8): 809-830.

[17] Y. Cao and L. Zhang. Data fusion of heterogeneous network based on BP neural network and improved SEP[J]. International Conference on Advanced Infocomm Technology (ICAIT). 2017, 138-142.