

## Big-screen text input system based on gesture recognition

Xing Guo<sup>1,a</sup>, Zhenyu Lu<sup>1,b</sup>, Rongbin Xu<sup>1,c</sup>, Zhengyi Liu<sup>1,d</sup>, Jianguo Wu<sup>1,e</sup>

<sup>1</sup>Computer Science and Technology Department, Anhui University, Hefei City Anhui Province, China 230031

<sup>a</sup>15803771@qq.com; <sup>b</sup> 294833476@qq.com; <sup>c</sup>2867633607@qq.com

<sup>d</sup>315857812@qq.com; <sup>e</sup>215430214@qq.com

**Keywords:** Gesture Recognition; Gesture Input; Chinese Character Input; SIFT;

**Abstract.** Now more and more based on gesture interaction system applications, but there are simple gestures to operate the mouse interaction, no text input to the system function. In this paper, the blind letters gestures as input gestures, using kinect to get the depth image, gestures split, use SIFT feature extraction, then get manual alphabet as Pinyin input method to provide an Chinese character input to the system.

### Introduction

With the rapid development of projection technology, modern display screen is bigger and bigger, and brings a visual impact on the big screen, but interaction of the big screen and the audience is more and more difficult. The literature<sup>[1]</sup> provides a way to control the mouse to operate computers through remote gesture, which effectively solve the problem of interaction of the people with big-screen, but the literature<sup>[1]</sup> does not provide a valid text input method, if the people want to search or query something, when input the text, they will find the literature<sup>[1]</sup> can not meet their need. In this paper, we will realize the text input by using recognize the blind sign language gesture. Gestures are divided into two types: dynamic gesture and static gesture. Dynamic gesture using the blind sign's movements to express semantic as text input, and now, the feasibility is low for the dynamic hand gesture to recognize text input<sup>[2][3]</sup>. Static hand gesture recognition can make a gesture to correspond to a Chinese characters, but because there are ten thousand characters, so it is very difficult to design hand gesture to correspond the character and remember it. So this paper uses sign language alphabet and numbers of gesture as a static hand gesture recognition object to input character through Pinyin input method.

### System design

The System use General Chinese Gesture Language of alphabet and digit as standard input, which shown in Fig. 1. First recognize the input gesture, then get the letters or number indicated by gesture, and the letters and numbers are input to the Pinyin input method, at the last, you can get the characters by the Pinyin input method.

The System get the image through the image collection device, such as camera, then do the gesture segmentation in the input image, extract the gesture edge through the Canny algorithm, and get the edge feature, obtain edge feature vector, and compare the vector with the template library vector using similarity, and find the most similar gesture, then get the letter which corresponds to the gesture. Input this letter into Pinyin input method, do the loop until all pinyin are input into the Pinyin input method and get the correct Chinese character.

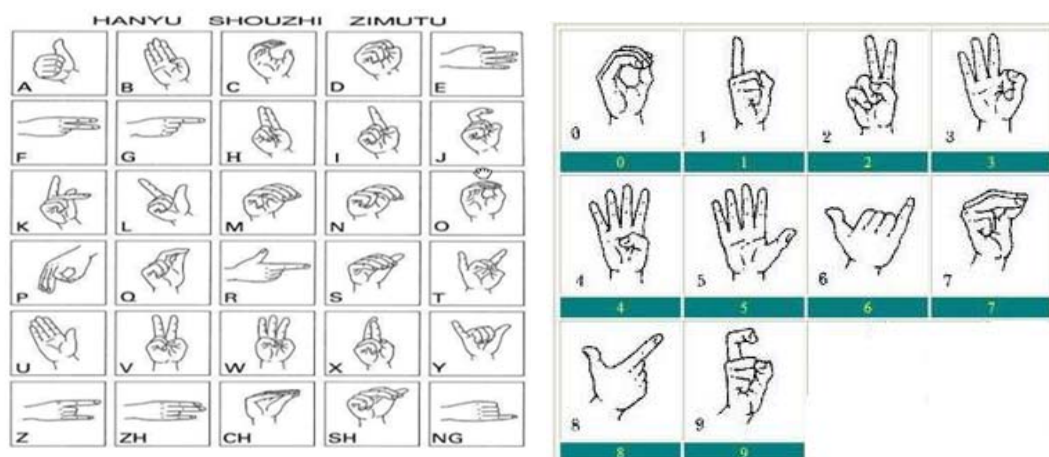


Fig.1 Character、Digit Hand Sign

## Gesture segmentation and edge extraction

At present, the gesture segmentation method<sup>[4]</sup> mostly use skin color model which base on the YCbCr or HSV or RGB color space, and then get the whole hand area using color information method. But because of the instability of the light, the skin color can change greatly in different lighting conditions ,which will produced in part areas of the skin can not be detected or non skin region is considered to be an area of skin, result in the hand detecting failure.

According to the above problems, this paper uses a equipment named Kinect. This device can provide depth image. And the depth image includes the depth information from the objects to the device. By the depth of image, we can extract hands without the affection by the light interference.

Fig.3 shows the color image and the depth image in different light environment. Contrast these two images, the depth image obviously do not disturb by the light. In our hand segmentation test, all base on a hypothesis: people's hands is the first objects, that is to say there have no other objects between the device and the hands.

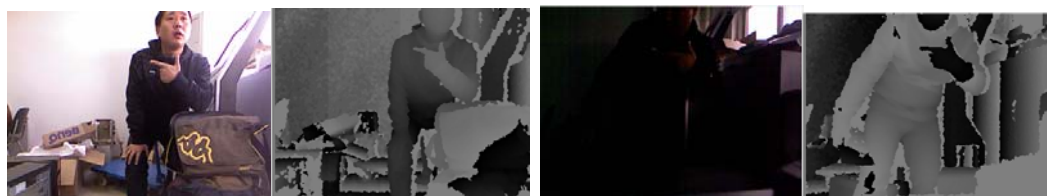


Fig .3 RGB and Depth Image in different Light Environment

On the basis of the hypothesis, we can statistic the pixel counts every 1cm from near and far. Once we detect the number reaches the hand size in one block in some one distance, then this distance is the threshold. In this paper, the experimental distance threshold is 75~85cm, the gesture segmentation is shown in figure 4. Then detect this gesture edge through the Canny detection algorithm, as shown in figure 4 too.

## Feature extraction

In the present, Most extraction of the gesture recognition feature is by geometric moments<sup>[5-7]</sup> or Fourier description operator<sup>[8]</sup>. Geometric moment can effectively recognize the gesture which is rotated, scaled, moved, but only the gesture classification is few, recognition accuracy was improved obviously. Fourier description operator have better description ability for the feature, but if the series selected low, the gestures feature detail will lost, if the series of selected was high, the computational complexity will increase, and it will very sensitive for the feature details. This paper choose the SIFT algorithm to describe extraction feature, the algorithm is a local feature description, can keep the invariant of the rotation, scaling similarly.

SIFT algorithm<sup>[9]</sup> was published by David Lowe in 1999, then completed summary in 2004. The

algorithm found out the contours of the key points. The key point is having the direction information of local information value. SIFT algorithm to find the points which are change big in the DOG space, that are the local extreme points.

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (1)$$

After determining the key point, using the key point of the field pixel gradient direction to distribute the characteristics and let the descriptors be invariant to rotation. The key points of the gradient magnitude  $m(x, y)$  and gradient  $\theta(x, y)$  direction formula is as follows:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (2)$$

$$\theta(x, y) = \tan^{-1} \left[ \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right] \quad (3)$$

Then, for any key point, with the key point as the center of the  $8 \times 8$  field as the sampling window. In the region, calculated for each pixel in the 8 direction of the gradient direction histogram in the  $2 \times 2$  region, rendering each gradient direction of the accumulated value, to form a seed point. Every seed point is described by 8 dimensional histogram operator array descriptions. Then histogram statistics the next  $2 \times 2$  region, form a seed point. Such key point need to 16 seed points to describe it, eventually forming a 128 dimensional feature vector. The vector is the SIFT descriptors, namely SIFT feature vector. The Figure 4 is the key point which is to find.



Fig. 4 75~85cm Extract Gesture and Canny Edge Detection and Gesture Key Point

## Gesture input

Using SIFT algorithm, we build feature vector template library for the 30 Chinese Pinyin and 10 digital sign language gestures. Using similarity computation of the detected gesture feature vector and template library, we select the max value of similarity as the impute gesture letter. if the template feature vector is  $R_i = (r_{i1}, r_{i1}, \dots, r_{i128})$ , and the detected gesture features vector is  $S_i = (s_{i1}, s_{i1}, \dots, s_{i128})$ . Then there similarity is

$$d(R_i, S_i) = \sqrt{\sum_{j=1}^{128} (r_{ij} - s_{ij})^2} \quad (4)$$

## Experimental analysis

This system is programming with VS2010 program in win7. First experiment invites experimenter to face the camera and do the gestures of 30 letters and 10 numbers as a sample template. Based on the template, experiment has two parts. The first part of the experiment was compare Fourier feature extraction and SIFT feature extraction. The second part of experiment compare on the text input time and error rate while the template space is or not be updated according to the context.

Experiment 1: because there are large number of gesture, experiment only selected the A, J, M, N, D, O six gestures to be compared. And do this experiment 20 times. The experimental results as shown in table 1:

Table 1. Fourier and SIFT compare

| Mode         | A  | J  | M  | N  | D  | O  |
|--------------|----|----|----|----|----|----|
| Fourier      | 20 | 20 | 13 | 14 | 17 | 15 |
| SIFT feature | 20 | 20 | 18 | 17 | 19 | 20 |

From the table 1, we can see the edge is more relatively clear, gestures letters which have no

similar gesture with recognize better, A and J can reach 100% recognition rate, but on the M,N and D,O two group of letters which easily confuse each other, Fourier can't solve the details of the problem effectively, lead to recognition rate was lower than SIFT, the Fourier can reach about 70%, while the Sift can reach 90%. So SIFT can divide similar gestures in image details.

## Summary

This article using the static hand gesture recognition propose a way of text input to the system for large screen interaction, the method uses the depth camera, to avoid light interference, improves the hand gesture segmentation accuracy greatly. Using SIFT algorithm to extract the gesture feature, the algorithm extract the feature can maintain detail feature effectively, at the same time do not lose rotation, scaling invariance, it can improve recognition results significantly. The system preliminary realized the function of character input, but there still exist some problems in the current system such as switch gesture, words backspace, control gesture design, such as if system recognize the wrong letter, due to the absence of design control gesture, now we use the keyboard to do fallback operation. Because exist these problems, lead the whole system text input efficiency is still very low, speed is very slow. These are the next step work focus.

## References

- [1] Guo Xing, Liu Zhengyi, LI Wei, et al. Implementation method of human-computer interaction system with large screen[J]. Computer Engineering and Applications, 2012, 48 (1): 176-179.
- [2] Wang Zhaoli, Wang Li. A Simple and Efficient Method of Dynamic Hand Gestures Recognition [J]. Computer Engineering and Applications,,. 2002, 19(8): 73-75.
- [3] Ren Haibing,Zhu Yuanxin,Xu Guangyou. Spatio-Temporal Appearance Modeling and Recognition of Continuous Dynamic Hand Gestures [J]. Chinese Journal Of Computers. 2000, 23(8): 824-828.
- [4] Chen Duansheng, Liu Zhengkai. A Survey of Skin Color Detection [J ]. Chinese Journal Of Computers, 2006,24(2) :65- 69.
- [5] A.Jain and A.Vailaya, Shape based retrieval:a case study with tradernark image databases, pattenRecognition,vol.31, No.9, pp.1369 — 1390, 1998.
- [6] LU Kai ,LI Xiaojian ,Hand Signal Recognition Based on Skin Color and Edge Outline Examination[J]. J. NORTHCHINAUNIV. OFTECH.,2006,28(3):14-16.
- [7] Yangjiang Wang ,Baozong Yuan. A Novel Approach for Human Face Detection from Color Images under Complex Background. Pattern Recognition, 2001 ,34 (10) : 1983-1992.
- [8] LIU Yin,TENG Xiao-long,LIU Chong-qing. Hand Gesture Recognition Based on Fourier Descriptors with Complex Backgrounds [J ]. Computer Simulation, 2005,22(12) :127- 129.
- [9] Lowe D G. Distinctive image features from scale-invariantkeypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.