

A Algorithm for Static Gesture recognition Using combination of object features

Duan Hong*

Software School of Xiamen University
Xiamen city,China

*corresponding author: hduan@xmu.edu.cn

Luo Yang

Software School of Xiamen University
Xiamen city,China

e-mail:779446731@qq.com

Abstract—Proposed a static gesture recognition method for identifying characteristics of the object in combination. With the feature vector composed of five features, such as the number of fingers, gesture outline convex defect characteristics, the length and area of contour and Hu matrix, we adopted the template-matching method to conduct the matching of featured parameters. Experiments show that the method successfully recognized static gestures under complex background and could reduce the impact of environmental change simultaneously.

Keywords-Gesture Recognition, combined features, static gesture, template-based match

I. INTRODUCTION

Traditional input devices (such as keyboard, mouse, etc.) in the human-computer interaction activities cannot act flexibly and naturally. With the development of computer technology, the human-computer interaction techniques that meet interpersonal communication habits become hyperactive. These studies include face recognition, gesture recognition and body potential recognition, etc[1] which has a natural nude monocular vision-based hand gesture recognition that is natural, direct and concise compared to other human-computer interaction. The input of gesture recognition technology commonly uses image data. Current identification methods mainly include shape-base recognition technology [2], skin-based identification technology [3], as well as the use of certain characteristics, such as Haar features, SIFT / SURF characterized[4], generalized hough transform on the image data classification training technology. The difficulty of research in the field is that the degree of precision and accuracy in the complex context, as well as the real-time nature of the algorithm.

In this paper, we proposed a method for recognizing static features basing on a combination of characteristics. With the feature vector composed of five features, such as the number of fingers, gesture outline convex defect characteristics, the length and area of contour and Hu matrix, we adopted the template-matching method to conduct the matching of featured parameters. Experiments show that the method successfully recognized static gestures under complex background and could reduce the impact of environmental change simultaneously which can basically meet the practical requirements of real-time.

II. EXTRACTION OF THE CHARACTERISTIC PARAMETERS

In recognition of images, the selection of the characteristics of targeted objects have a great influence on the identification results. Single characteristic is often affected by the environment and produce relatively large interference. From the literature [5],we know that feature extraction is to measure various aspects of the object to obtain a set of characteristics, known as a set of feature vectors of the object $X = (X_1, X_2, \dots, X_n)$. After the first few chapters of the method of processing, we get the outline of the hand. After repeated experiments for the binary image and gesture contour of the hand region, we extracted five characteristic parameters, which were used to form the characteristics vector.

A. Gesture outline convex defect characteristics

We first explain the concept of convex defect. In Figure 1, the picture of hand is used to illustrate the concept. The black line depicting the hand around the convex hull is shown, the marked areas of A to H are the convex hull of each convex defects. As seen in Figure 1, the convexity defects provide a method to present characteristics of hand and hand state [7].

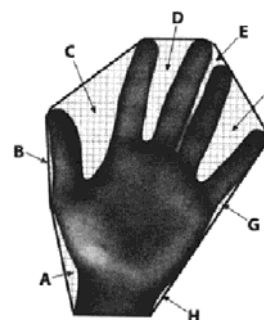


Figure 1 Convex hull of the contour of the hand and convex defect

Our algorithm first find the contour convexity defects, store them in an array. The data structure of each of the convex defects are three components, start is the starting point of the defects on the convex hull, end is the termination point of the convex hull defects, depth point is a defect in the distance of the convex hull side edge (with the convex hull of the defects related) furthest point, the last parameter depth is the distance of the farthest point and the package edge (eDGE).

The process of calculation is as follows:

- 1, initialize the count value $n = 0$, loop through each convex defect
- 2, on each of the convex defects, satisfy equation 1
Distance (start, depth point) $> c$ (Equation 1)
ORDER count value n is incremented by 1, where c is a constant threshold, so that $c = \text{box height} / 5$ can obtain a better effect, after repeated experiments, box minimum bounding rectangle for the hand region.
- 3, n is seeking convex defects characteristic value

B. Characters of the length of contour.

The length of the contour is number of points that the gesture contour contains. In order to reduce the distance from the lens, we use the relative contour length to show this characteristic. We took the length of gesture outline as $L1$, and the perimeter of the external matrix of the gesture outline as $L2$.

As shown the formula:

$$\text{lengthRatio} = L1/L2 \quad (\text{Equation 2})$$

As shown in figure 2:

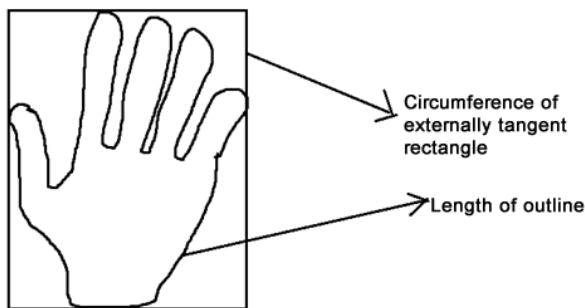


Figure 2 the length of the outline

C. Contour area features

We divided the external matrix of the gesture outline into 2 parts, we took the area surrounded by the upper portion of region contour as $S1$, and the area surrounded by the lower portion of region contour as $S2$, and we use the ratio of them as the area's characteristic. As shown the formula:

$$\text{areaRatio} = S1 / S2 \quad [7]. \quad (\text{Equation 3})$$

The area of the outline is the total number of the pixels in the contour area. We can get the result by scanning contour diagram of the binary rectangular area, calculating the total number of the black dot(hand area in binary diagram). As Shown in figure 3:

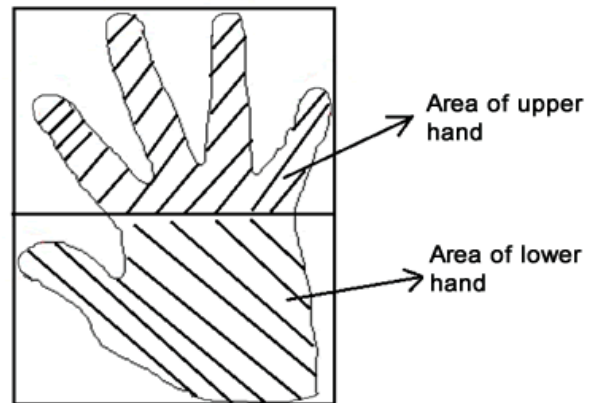


Figure 3 the area of the outline

D. Hu moment character

If we took the binary image or a grayscale images as the two-dimensional density distribution function, then we can bring up the concept of moment descriptors. In 1962, Hu proposed the concept of the moment invariants at the first place, he used the nonlinear combination of moment draw a set of moment invariants which has rotational invariance, translational invariance and the proportion of invariance.

First of all, let's introduce some definition about moment: In the continuous case, the image functions is $f(x,y)$, then the $p+q$ order geometric moments of image(standard moments) has a definition like this:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (\text{Equation 4})$$

The definition of $p+q$ order central moments:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (\text{Equation 5})$$

\bar{x} and \bar{y} represent the center of gravity of the image and are calculated as:

$$\bar{x} = m_{10} / m_{00}$$

$$\bar{y} = m_{01} / m_{00} \quad (\text{Equation 6})$$

For discrete digital image, we use sum NO. instead of integration:

$$m_{pq} = \sum_{y=1}^N \sum_{x=1}^M x^p y^q f(x, y) \quad p, q = 0, 1, 2, \dots \quad (\text{Equation 7})$$

$$\mu_{pq} = \sum_{y=1}^N \sum_{x=1}^M (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad p, q = 0, 1, 2, \dots \quad (\text{Equation 8})$$

Among that, N and M are respectively the height and width Of the image.

Normalized central moments is defined as:

$$\eta_{pq} = \mu_{pq} / (\mu_{00}^\rho) \quad (\text{Equation 9})$$

Among that, $\rho = (p + q) / 2 + 1$

In 1962, Hu use algebraic invariant theory for scale standardization moments to construct seven famous moment invariants. Seven moment invariants is constructed of a linear combination of the second-order central moments and third-order central moment, the specific expression is as follows:

$$\begin{aligned} M1 &= \eta_{20} + \eta_{02} \\ M2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ M3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ M4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ M5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \\ &+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \\ M6 &= (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \\ &+ 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ M7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \\ &- (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \end{aligned} \quad (\text{Equation 10})$$

The amounts of calculation of the seven Hu moments are not equal and so it is with the information content. In general, useful information in the image is centralized in the low-order moments of which the amount of calculation is relatively small. The higher moments are computationally intensive, and contain some details that are more susceptible to noise interference, therefore it is not easy to distinguish between the various target the difference between the higher moments. Consider the needs of the system, we select the first two Hu moment as the feature, namely M1, M2.

E. characters of the number of fingers.

This system uses K vector method to calculate the number of fingers. More details were described in [6].

III. MATCHING OF THE CHARACTERISTIC PARAMETERS

The recognition methods for current static gesture recognition are mainly template matching, neural network and SVM classifier method. According to gestures feature vector characteristics in this paper, We chose the template matching method for the identification of gestures and the improved distance of European distance, European distance computation formula is as follows:

$$D = \sqrt{\sum_{i=1}^n (\theta_i - T_i)^2} \quad (\text{Equation 11})$$

D is the total distance, θ is to identify gestures eigenvector of a characteristic quantity, T1 is the corresponding template gestures, n is the number of characteristic. In order to

improve the calculation efficiency, we improved the calculation method. A new calculation method is shown as follows type:

$$D = \sum_{i=1}^n \alpha_i |\theta_i - T_i| \quad (\text{Equation 12})$$

α_i is a factor to adjust the size, used to balance different data size to a on the orders of magnitude. Because in feature extraction part of our common extraction five features six characteristic value, so the formula 4-12 of the value of n is 6. After experimental comparison,

Factor set to 8.33, 8.33, 1, 1, 54.95, 135.14 can achieve good effect. Calculation for recognition gestures feature vector and template library every gesture form between the total distance after, find the minimum value, minimum value corresponding template is the identification results.

IV. EXPERIMENTAL RESULTS AND CONCLUSIONS

We defined 8 kinds of gestures during experiment, as shown in figure 4:

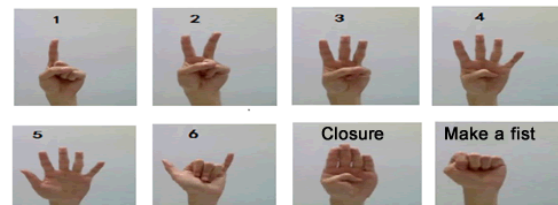


Figure 4 8 kinds of gestures

First of all, we conducted four sets of experiments to test the influence which the increases of the kind of gestures make on the correct rate of recognition. The 2,4,6,8 are the gesture to be identified while experiment, we tested 100 times for each group. We do the same number of tests for every gesture in each group. We recorded the total errors and calculated the correct rate of recognition during these experiments, the experimental result shown as table 1:

Table 1 Different correct rate under different gesture

Gesture kinds	Total test number	Total errors	Correct rate
2	100	1	98%
4	100	2	96%
6	100	4	95%
8	100	5	93%

Through the result of the experiment, the correct rate will decline with the increase of the kinds of gestures to be recognized, but still maintained at a high level. So this identification method is feasible.

At the same time, in order to test the change of correct rate in the different environmental contexts, we implemented eight different gesture experiments under the simple background and the complex background. 100 experiments for each gesture. Figure 5 is the simple background and the complex background of environment.



Figure 5 Background of experiment

Table 4-2 and Table 4-3 are the results of the experiment:
Table 2 Gesture recognition results under simple background

Gesture	The number of experiments	Correct	Correct rate
1	100	100	100%
2	100	100	100%
3	100	91	91%
4	100	90	90%
5	100	95	95%
6	100	92	92%
Closure	100	100	100%
Make a Fist	100	99	99%

Table 3 Gesture recognition results under complex background

Gesture	The number of experiments	Correct	Correct rate
1	100	99	99%
2	100	100	100%
3	100	97	97%
4	100	89	89%

5	100	94	94%
6	100	15	85%
Closure	100	100	100%
Make a Fist	100	98	98%

Through the result of the experiment, as for the eighth gesture, the correct rate of recognition is always beyond 80 percent in different environments. Although correct rate will decline in a complex environment, But in general, the algorithm is still practical.

REFERENCES

- [1] Ren haibing, Zhu yuanxin, Xu guangyou, Lin xueyin, Zhang xiaoping Research based on the visual gesture recognition[J]. ACTA Electronica Sinica, 2000, 28(2) : 119-121
- [2] Erdem Yörük, Ender Konukoğlu, Bülent Sankur, Senior Member, IEEE, and Jérôme Darbon. Shape-Based Hand Recognition[J]. IEEE Transaction on image processing, 2006, 15(7):1803-1815
- [3] JongShill Lee, YoungLoo Lee, EungHyuk Lee, SeungHong Hong, Hand region extraction and Gesture recognition form video stream with complex background through entropy analysis[J]. 26th Annual International Conference of the IEEE 2004, 2:1513-1516
- [4] Luo Juan, Oubong Gwun, A Comparison of SIFT, PCA-SIFT, and SURF[J]. International Journal of Image Processing, 2009, 3(4):143-152
- [5] D. M. Gavrila, L.S.Davis. Towards 3D model-based tracking and recognition of Human movement:a multi-view approach[C]. In: Workshop on Automatic Face and Gesture Recognition, Switzerland, 1995:272~277.
- [6] Cai kunyu. Fingertips Recognition and PowerPoint Display Application Based on Computer Vision[D]. Xiamen University Bachelor thesis. 2012.
- [7] Yu shiqi, Liu ruizhen.Study opencv(Chinese version) [M] .TsingHua University Press. Beijing : , 2009.
- [8] Hu. Visual Pattern Recognition by Moment Invariant[J].IRE Trans Information Theory. 1962(8):179-187.