Sensors & Transducers, Vol. 169, Issue 4, April 2014, pp. 67-72

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

The Research and Application of SURF Algorithm Based on Feature Point Selection Algorithm

[1] Zhang Fang Hu, [2] Lin Yang, [2]Luan Luo, [2]Yi Zhang, [3]Xiao Chuan Zhou

[1] Key Laboratory of Optical Fiber Communication Technology Chongqing Education Commission, Chongqing University of Post and Telecommunications, Chongqing 400065, P. R. China

[2] Research Center of Intelligent System and Robot, Chongqing University of Post and Telecommunications, Chongqing 400065, P. R. China

[3] China Petroleum Pipeline Telecommunication & Electricity Engineering Corporation, Langfang City of Hebei province, 065000, P. R. China

[1] Tel.: +86-18725875566

[1] E-mail: huzf@cqupt.edu.cn

Received: 26 January 2014 /Accepted: 27 March 2014 /Published: 30 April 2014

Abstract: As the pixel information of depth image is derived from the distance information, when implementing SURF algorithm with KINECT sensor for static sign language recognition, there can be some mismatched pairs in palm area. This paper proposes a feature point selection algorithm, by filtering the SURF feature points step by step based on the number of feature points within adaptive radius r and the distance between the two points, it not only greatly improves the recognition rate, but also ensures the robustness under the environmental factors, such as skin color, illumination intensity, complex background, angle and scale changes. The experiment results show that the improved SURF algorithm can effectively improve the recognition rate, has a good robustness. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Sign language alphabet, Feature point selection algorithm, Adaptive radius, Improved speed up robust features algorithm.

1. Introduction

Sign language recognition based on machine vision is a hot research topic in recent years. Qinghua University's Shin-Han Yu, Chung-Lin Huang identified 40 Taiwan Sign Language symbols by applying parallel Markov (PHMM) method, and the accuracy rate is 94.04 % [1]; Rini Akmeliawatil, Melanie Po-Leen Ooi used the fingers with highlight visual gloves as input, applying skin color segmentation and neural network method for the identification of Malaysian Sign Language alphabet, and the accuracy rate is 95 % [2]. These methods have a good recognition rate, but most of them did not take the robustness of factors like illumination, background, and angle variations into consideration. Researchers from the Engineering Research & Development Center of Information Accessibility of Chongqing University of Posts and Telecommunications achieved the intelligent wheelchair interaction based on static hand gesture by using KINECT sensor with Hu invariant moments algorithm. This method controls the motion of intelligent wheelchair through the identification of predefined gestures, the application of depth information effectively overcomes the interference of environmental factors includes light, complex background and angle change [3, 4]. Furthermore, researchers achieved static sign language alphabet recognition by comparing depth information of the real time image with the template image using KINECT sensor with SURF algorithm [5].

However, as the pixel information of the depth image is derived from the distance information, there can be some mismatched pairs in palm area in the use of SURF algorithm for static sign language recognition based on depth image. Therefore, this paper proposes an improved SURF algorithm based on feature point selection algorithm combined with the SVM method to realize the static sign language alphabet recognition.

2. Extraction of Feature Points Based on SURF Algorithm

SURF (Speed Up Robust Features) [6] is a kind of matching algorithm based on fast robust features. Feature points are extracted by using SURF algorithm, first to generate integral images of the images to be detected, then use [7] gradually enlarge box filter of convolution operation, to get the scale space of real-time image, $\vec{H}\left(x,y,\sigma\right)$. The approximation of the determinant of matrix Happrox:

$$\det\left(\vec{H}approx\right)=D_{xx}D_{yy}-\left(wD_{xy}\right)^{2}, \qquad (1)$$

where D_{xx}, D_{yy}, D_{xy} are the results of convolution, ω is weighted coefficient, can restricted by F (Frobenius) norm.

$$\omega=\frac{\left\|L_{\mathrm{xy}}\left(1.2\right)\right\|_{F}\cdot\left\|D_{\mathrm{xx}}\left(9\right)\right\|_{F}}{\left\|L_{\mathrm{xy}}\left(1.2\right)\right\|_{F}\cdot\left\|D_{\mathrm{xy}}\left(9\right)\right\|_{F}}=0.912\approx0.9, \qquad (2)$$

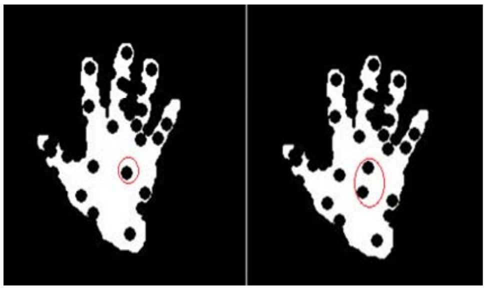
Combine formula 1 and formula 2, when the $\det\left(\vec{H}approx\right)$ is positive, the point is the extreme point; on the other hand, is not the extreme point. Through the comparison of three-dimensional neighborhood of each extreme point with the 26 corresponding inner points, we obtained a local maxima. Then use the Taylor expansions of two three-dimensional quadratic equations for surface fitting [8], so as to realize the non maxima suppression (NMS) of the extremism point location to get the accurate coordinates and scale σ of feature point.

But in order to overcome the interference of complex background and the illumination changes opponent pixel segmentation, we set the depth image as the image to be detected, and pixel information of the depth image is derived from the distance information, through the binarization algorithm the segmentation result of hand was obtained. For the fearful similarity of pixel value information in palm part, it is easy to calculate the incorrect feature points, which can bring interference to the following steps of recognition, so we need to improve the SURF algorithm.

3. Improved SURF Algorithm Based on Feature Selection Algorithm

3.1. Filtering of the Feature Points

Take a common palm as an example, Fig. 1 shows the feature points calculated by using the results of the original SURF. Fig. 1(a) is a real-time image, Fig. 1(b) is a template image. We can see both of the two resulting pictures, have incorrect feature points, which are generally concentrated in the palm area.



(a) real time picture (b) temple picture

Fig. 1. Wrong feature points calculated by SURF.

According to this characteristic, this paper puts forward a kind of improvement measures, through contrasting the number of detected feature points in neighborhood and the distance between feature points to make selection, so as to reduce the impact from the incorrect feature points. The improved algorithm flow chart is shown in Fig. 2.

Assume real-time image and the template image are I and I', the feature point sets are $C=\left\{C_1,C_2,\cdots C_n\right\}$, $C'=\left\{C'_1,C'_2,\cdots C'_n\right\}$ respectively, the corresponding matching feature points are P, P', which $P\in C$, $P'\in C'$, within the radius of R, if the two images have only translation and rotation, then should meet the following conclusions:

1) The total number of the feature points are equal in the same neighborhood. That is, the total number of feature points within radius of R with P as the center should be equal to that when P' is the center.

2) The distance between the corresponding feature points are equal. Assume $C_i, C_j\in C, C'_i, C'_j\in C', C_i$, C_i' is a pair of corresponding feature points, C_j, C_j' is a pair of corresponding feature point too, and the distance

between C_i and C_j should be equal to the distance between $C_i{'}$ and $C_j{'}$, $d(C_i, C_j) = d(C_i', C_j')$.

Based on the above principle, the incorrect feature points can be eliminated according to the following steps:

1) Define an adaptive radius R, if the number of feature points is less than two within a radius of R, then calculate the number of feature points after doubling the value of R; if the number of feature points is greater than two, then do the following two steps.

2) Calculate the number of feature points within a radius of R with P and P' respectively. If the number of feature points are the same, then P is the correct feature point, otherwise, need to be removed.

3) Calculate the distance between the feature point with other neighboring feature points, and use this parameter to further eliminate incorrect feature points. Assume P and P' meet principle (2), Suppose the number of feature points within a radius of R is n, record the distance between P,P' and its neighboring feature points respectively in descending order as $D_1 = \{d_1, d_2, \cdots d_n\}$ and $D_2 = \{d_2, d_2, \cdots d_n\}$. If the deviation is within the allowable range, then P is considered to be the correct feature point. Otherwise, need to be removed. In this way, incorrect feature points in the palm area are eliminated effectively.

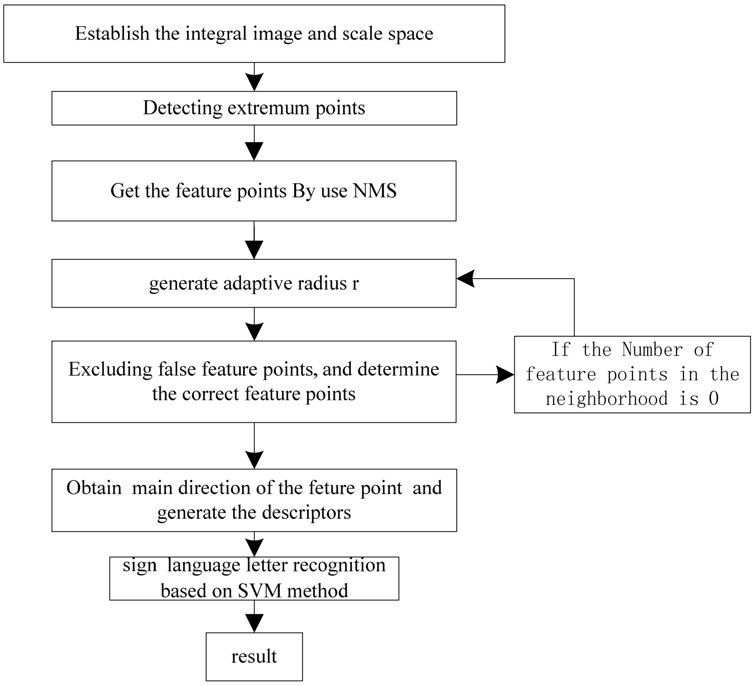


Fig. 2. The improved algorithm flow chart.

3.2. Description of Feature Points

Descriptions of feature points are divided into two parts, first to determine the principle direction of the feature points, and then generate the corresponding feature point descriptors.

Calculate the Harr wavelet responses of size 4σ (σ is scale) within a circle of radius of 6σ and center at feature point, represented by d_x and d_y. Through the Gaussian weighted, d_x, d_y are denoted as denoted as $\vec{W}_{dx}, \vec{W}_{dy}$, in all the 72 sector regions of 60 degree, the gradient of that region, the main direction is the direction that has the maximum gradient gradient. According to the arc tangent of $\vec{W}_{dx}, \vec{W}_{dy}$, the main direction can be expressed in degrees.

After determining the main direction, then place feature points as the core, divide the square window of size 20σ into 16 small squares. Calculate Harr wavelet responses for each region d_x, d_y, and then weighted by Gauss function. Sum those Harr wavelet responses d_x, d_y greater than zero and less than zero respectively, the descriptors obtained as follows:

$$\bar{D}\mathrm{esc}_{squre} = V\left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right), \quad (3)$$

As a four dimensional vector, 4×4 small areas constitute 64 dimensional feature point descriptor vector. The next step comes to sign language alphabet recognition.

4. Sign Language Letter Recognition Based on SVM Method

The SVM method has many advantages in solving small, nonlinear and high dimensional pattern recognition [9]. Moreover, the sign language letter recognition is a multi classification problem, at the same time, SURF feature vector is a high dimensional vector, this paper uses SVM method to carry out the sign language letter recognition.

The SVM multi-classification algorithm creates a series of two categories classifier and put them together to realize multi classification. Mainly includes the following three methods: 1) "one against rest" classification. 2) "one to one" classification. 3) "hierarchy" method. [10]

In order to reduce the training time and ensure a high recognition rate, we apply the "one to one" classification method to realize recognition of sign language letter. This method will train $m(m-1)/2$ SVM models for the m categories, then the samples to be classified are input into each SVM model for classification. If the sample x belongs to class i, then this class.scores a point. Finally, the category with the highest score is assigned to this unknown sample. To achieve the classification of 27 sign language letters, 351 SVM trained models are needed. In the training process, first we need to apply formula (4) to each kind of sign language letter feature extracted by SURF for feature size normalization ([0 1]), in order to avoid the flood of big data and overfitting problem. In formula (4), x'_n is the i-dimensional feature vector for the nth characteristics. Next, input the feature vectors to those 351 models for classification. The category which wins the most votes will be recognized as the recognition result.

$$x'_n = \frac{x'_n - \min x'_n}{\max x'_n - \min x'_n}, \tag{4}$$

5. Analysis of Experimental Results

5.1. The Experimental Environment

Microsoft Kinect sensor and Windows XP laptop are the main hardware for the experiments. Software programming environment is VC++2008, includes precompiled library Opencv and OpenNI.

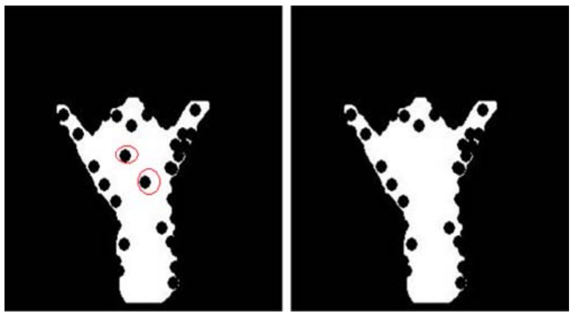
Among them, OpenCV is a publishing cross platform computer vision library based on license BSD, which consists of a series of C and C++ library for the basic image processing. OpenNI[11] is a must-install API if the Kinect sensor is running in Windows environment. Fig. 3 shows the sign language letter template used in our experiments.

The experiments examine the algorithm in the aspect of improvement, robustness, real-time performance, and the recognition rate.

Fig. 3. Sign language letter template.

5.2. Algorithm Improvement Effect Test

In order to prove the effectiveness of the improved algorithm in eliminating the incorrect feature points, take letter "Y" as an example, compare results shown in Fig. 4(a) with the result of the improved algorithm shown in Fig. 4(b), we can easily observe the improvement of the algorithm in eliminating the false feature points in the palm area.

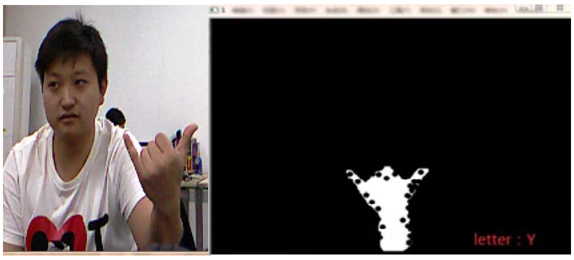


(a) original (b) improved

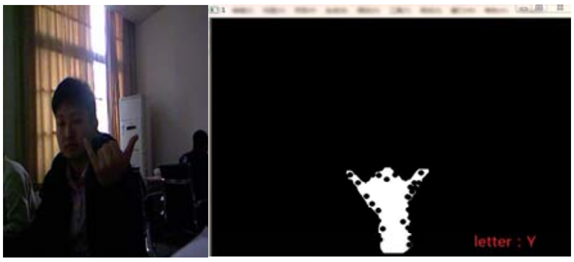
Fig. 4. The Feature point calculation of Y.

5.3. Robustness Test under Different Lighting Conditions and Complex Background

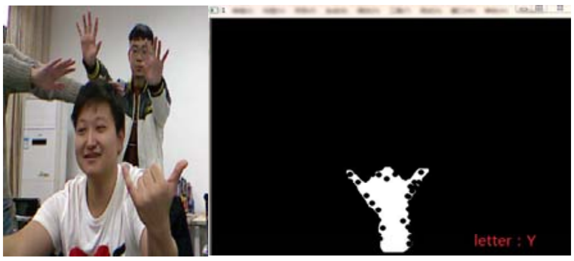
In order to verify the robustness of the algorithm under varied lighting conditions, complex background, conduct the experiments under strong and weak lighting conditions, complex background respectively, Take letter "Y" as an example, in Fig. 5(a) – Fig. 5(c), color image collected by KINECT sensor is on the left, and the result of recognition is on the right.



(a) Matching results under strong light condition.



(b) Matching results under weak light condition.



(c) Matching results under complex background.

Fig. 5. Matching results for letter Y under different conditions.

Experiments have shown that, were correctly identified in three environments. Also, from Fig. 5(a) – Fig. 5(c), due to the use of the hand pixel segmentation based on image depth, exclude the interference of complex background, illumination change in the segmentation, the recognition results of the three cases are almost the same, which demonstrates the robustness of this method.

5.4. Real Time Performance Test

Real time mainly depends on the running time of the algorithm. Table 1 gives the average running time of sign language letter recognition required by the original algorithm and the improved one. We can see the use of improved algorithm requires about 93 ms in image processing, that is to say, it can handle 10-11 pictures per second, which just take 8 more milliseconds than the original algorithm. Therefore, although the improved algorithm is complex than the original one, the handling process is quite fast, which proves that the improved algorithm makes no difference in terms of real-time performance.

Table 1. Algorithm running time.

Algorithm	Running time
Original	85 ms
improved	93 ms

5.5. Recognition Rate Test

In addition, the recognition method in this experiment were used with the method only using SURF algorithm and improved SURF algorithm are proposed for 150 trials on each letter, The corresponding recognition rates are shown in Table 2.

Table 2. The manual alphabet recognition rate table.

Case	Rate1	Rate2	Case	Rate1	Rate2
A	88.4 %	99.3 %	N	86.2 %	95.3 %
B	86.2 %	98.6 %	O	87.9 %	96.0 %
C	87.7 %	98.0 %	P	88.4 %	98.0 %
D	80.2 %	98.0 %	Q	85.7 %	95.3 %
E	90.5 %	99.3 %	R	87.5 %	98.6 %
F	89.3 %	99.3 %	S	88.4 %	97.3 %
G	88.4 %	99.3 %	T	86.2 %	96.0 %
H	82.0 %	98.6 %	U	87.8 %	98.6 %
I	87.9 %	98.6 %	V	85.7 %	98.6 %
J	85.7 %	98.6 %	W	88.4 %	98.6 %
K	85.7 %	97.3 %	X	85.7 %	95.3 %
L	85.7 %	97.3 %	Y	85.3 %	98.6 %
M	80.6 %	97.3 %	Z	87.8 %	98.6 %

Recognition rate 1 is the result after simply using SURF algorithm, and recognition rate 2 is the result of the method proposed in this paper. Comparing the two results, the latter get better recognition rate, which with an average recognition rate of 97.7 %. Among this, the recognition rate of letters "A", "E", "F", and "G" reach 99.3 %.

6. Conclusions

This paper studies a method of sign language alphabet recognition based on machine vision. Feature extraction is performed through the depth information obtained by Kinect sensor, then the application of improved SURF algorithm which is based on feature points filter algorithm for feature information extraction, and finally the usage of SVM method to train the sign language letter template for sign language letter recognition. A lot of experiments show that this method can obtain good recognition rate and robustness, and operation time can also meet the real-time requirements. But as the letter shapes are complex and with stereo, the next step will begin to use KINECT sensor to obtain 3D hand model, and the realization of sign language letter recognition in 3D.

Acknowledgements

This paper is supported by the Science and Technology Research Project of Chongqing Education Commission (KJ110518) and Natural Science Foundation of Chongqing University of Post and Telecommunications(A2009-49).

References

[1]. Shin-Han Yu, Chung-Lin Huang, Shih-Chung Hsu, et a1, Vision-based continuous sign language recognition using product HMM, in *Proceedings of the 1st Asian Conference on Pattern Recognition (ACPR)*, Beijing, 2011, pp. 510-514.

[2]. Rini Akmeliawati, Melanie Po-Leen Ooi, Ye Chow Kuang, Real-time Malaysian sign language translation using colour segmentation and neural network, in *Proceedings of the IEEE Instrumentation and Measurement Technical Conference on Synergy of Science and Technology in Instrumentation and Measurement (IMTC'2007)*, Warsaw, 2007, pp. 1-6.

[3]. Luo Yuan, Xie Yu, The design and implementation of a gesture-Driven system for intelligent wheelchairs based on the orientation histogram method, in *Proceedings of the International Conference on Key Engineering Materials and Computer Science (KEMCS'11)*, Shenzhen, China, 2012, pp. 109-113.

[4]. Luo Yuan, Xie Yu, Zhang Yi, The design and implementation of a gesture-driven system for intelligent wheelchairs based on Kinect sensor, *Robot*, Vol. 34, Issue 1, 2012, pp. 110-114.

[5]. Hu Zhang-Fang, Yang Lin, Luo Yuan, Zhang Yi, A novel static sign language letters recognition method based on improved SURF algorithm, *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, Vol. 25, Issue 4, 2013, pp. 544-548.

[6]. S. Somayajula, A. A. Joshi, R. M. Leahy, Mutual information based non-rigid mouse registration using a scale-space approach, in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, Paris, France, May 2008, pp. 1147-1150.

[7]. B. R. Pires, K. Singh, J. M. F. Moura, Approximating image filters with box filters, in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*, Brussels, 2011, pp. 85-88.

[8]. Hui Ni, Zhong Li, Hongxing Song, Moving least square curve and surface fitting with interpolation conditions, in *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM'2010)*, Taiyuan, Vol. 13, 2010, pp. 300-304.

[9]. Zhu Haizhou, Jia Yinshan, Remote sensing image classification based on support vector machine, *Science Technology and Engineering*, Vol. 10, Issue 5, 2010, pp. 3659-3662.

[10]. Li Jianwu, Lu Yao, A fast multi-class support vector machine, *Pattern Recognition & Artificial Intelligence*, Vo. 20, Issue 3, 2007, pp. 301-307.

[11]. Villaroman Norman, Rowe Dale, Swan Bret, Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor, in *Proceedings of the ACM Special Interest Group for Information Technology Education Conference (SIGITE'11)*, Ouro Preto, 2011, pp. 227-231.