

# An Improved SIFT Algorithm for Image Matching

Zhang Hui<sup>\*</sup>

Faculty of Information and Control Engineering  
Shenyang Jianzhu University  
Shenyang, China  
e-mail: [arnold0110@sina.com](mailto:arnold0110@sina.com)

<sup>\*</sup> Corresponding Author

Ren Dan

College of Engineering and Technology  
Eastern Liaoning University  
Dandong, China

Wang Li

Faculty of Information and Control Engineering  
Shenyang Jianzhu University  
Shenyang, China

Kan Hongliang

School of Information & Control Engineering  
Shenyang Jianzhu University  
Shenyang, China

Wang Bin

Northeastern University at Qinhuangdao  
Qinhuangdao, China

Zhang Fengzhong

Faculty of Information and Control Engineering  
Shenyang Jianzhu University  
Shenyang, China

Wang Xin

School of Information & Control Engineering  
Shenyang Jianzhu University  
Shenyang, China

Lü JiuYi

Faculty of Information and Control Engineering  
Shenyang Jianzhu University  
Shenyang, China

**Abstract**—Aiming at the problems of large calculating scale and high complexity in Scale Invariant Feature Transform (SIFT) feature matching algorithm, this paper presents an improved SIFT feature matching algorithm based on quasi Euclidean distance. The traditional Euclidean distance can only calculate out the variance of the two images to the corresponding pixel, so when a slight shift or distortion occurs in the image, it may produce a large deviation. The quasi Euclidean distance instead of Euclidean distance is as the similarity measure of feature descriptors to improve the SIFT feature matching. It reduces the dimensions of SIFT feature vector to improve the efficiency of feature matching. Experimental results show that under the condition of keeping the image matching rate and algorithm robust, the method can not only improve the matching accuracy but also shorten the matching time. The new algorithm has better performance than traditional algorithm, it is possible and valid, which are useful for the fields of image recognition, image reconstruction, etc.

**Keywords**- *SIFT algorithm; image matching; keypoints; quasi Euclidean; scale space*

## I. INTRODUCTION

Image matching is an important part of digital image processing, also is one of the most active research areas in machine vision and serves as an important step in many applications. Image matching refers to using effective methods to analyze the consistencies or likelihoods in two

or more images of the same object which are acquired in the different vision conditions, and find the most homologous image points among these images [1]. Nowadays, there are mainly two kinds of image matching methods: one is based on image gray scale matching and the other is based on image character matching. The first method is directly using image gray value to discover the best matching point, which algorithm is simple and high accuracy, it is mostly sensitive to illumination, scale, etc., but the amount of calculation is great too, so it is easy to fail in matching. The second method takes the image character as the matching measurement, such as image edge, image texture and image region statistical character. Compare with the first method, this kind of method has good anti-noise performance. it can be immune from illumination, scale and rotation.

For a long time, many domestic and scholars are dedicated to research image matching algorithm. In 2004, David G. Lowe formally proposed SIFT (Scale Invariant Feature Transform) algorithm [2,3] after 5-year perfection and summary, which is a feature-describing method, and it is applied to extract local feature widely; the SIFT features have good robustness and are invariant to image rotations, illumination changes, scale changes and so on. The SIFT features are local and based on the appearance of the object at particular interest points, which are invariant to image scale and rotation. They are also robust to changes in illumination, noise, occlusion and minor changes in

viewpoint. But in some occasions SIFT cost much time, and it is very difficult to achieve real-time control. In this paper, we presented a improved image matching method based on SIFT by using quasi Euclidean and the experimental results demonstrate the usefulness of the method in image matching.

## II. SIFT ALGORITHM

The SIFT algorithm is widely used local feature-based method for object recognition in image matching and is useful for nearly all computer vision tasks [4]. The steps of the SIFT algorithm are as follows:

### A. Detecting the Local Extrema in Scale-space

#### 1) Establish a scale space

The main idea of the scale space theory is using Gaussian kernel to do the scale transform of original figure, the sequences of image multi-scale space are acquired , then extract scale space feature from these sequences. In order to detect the stabilized key-feature-points effectively, 2D Gaussian kernel [5] is defined as Equation (1):

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

For a corresponding image  $I(x, y)$ , the differential Gaussian scale-space (DOG, Difference of Gaussian scale-space) function  $L$  is defined as the convolution of scale variant Gaussian kernel function  $G(x, y, \sigma)$  and image  $I(x, y)$ .

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (2)$$

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3)$$

Where  $I(x, y)$  is a function of the 2D original image,  $\otimes$  is the convolution operation in  $x$  and  $y$ ,  $L(x, y, \sigma)$  is a Gaussian image,  $(x, y)$  represents the pixels position , also named scale-space factor,  $\sigma$  is scale coordinate.

#### 2) Detection extremum points in scale space

The SIFT key-feature-points are made up of the local extrema of  $D(x, y, \sigma)$ ; For detecting the local maxima and minima points successfully, each point is compared with the pixels of all its 26 neighbors, which include 8 adjacent points surrounding it in the same scale and 18 points which are corresponding position of the upper and lower adjacent scale in pyramid . If this value is the maxima or minima comparing with its 26 neighbors in 3x3 regions at the current and adjacent scales, then this point is an extremum point. As the Fig.1 shown, maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles).

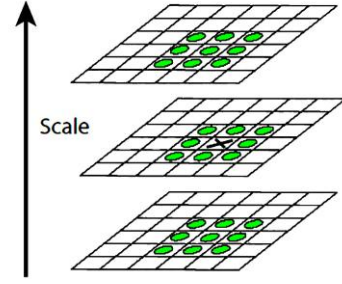


Figure 1. Extreme point detection of DOG scale space

### B. Locating the Extremum Accurately

The SIFT algorithm determines the location and dimension of the extreme value point through the second-order Taylor expansion of the DOG [6,7] function. Because the DOG operator is sensitive to edge responses and noise, we need use curve fitting to improve the stability of the keypoints, when the entire keypoint candidates have been

The 3D quadratic function [8] is fit to locate the keypoints in SIFT algorithm, it is shown in Equation (4):

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (4)$$

Where  $D$  and its derivatives are evaluated at the sample point , and  $x = (x, y, \sigma)^T$  is the offset from this point. The  $\hat{x}$  is determined by taking the derivative of this function with respect to  $x$  and setting it to zero, we can get :

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X} \quad (5)$$

For the experiments, all extrema with a value of  $|D(\hat{x})|$  less than 0.03 were discarded, so if  $|D(\hat{x})| < 0.03$ , the extremum was rejected.

A incorrectness peak in the difference-of-Gaussian function will have a large principal curvature across the edge, but have smaller principal curvature in the vertical direction. The principal curvatures can be computed from a 2x2 Hessian matrix,  $H$  is defined as the following equation (6). The derivatives are estimated by taking differences of neighboring sample points.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (6)$$

The eigenvalues of  $H$  and the principal curvatures of  $D$  are proportional relation. Then, we can compute the  $\text{Tr}(H)$  and  $\text{Det}(H)$  respectively,  $\text{Tr}(H)$  is the sum of the eigenvalues and  $\text{Det}(H)$  is their product from the determinant .

$$Tr(H) = D_{xx} + D_{yy}$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 \quad (7)$$

In order to check whether the principal curvature is in a certain range of  $\gamma$  ( $\gamma$  is the ratio of the maximum eigenvalue and the minimum eigenvalue), we only need to check:

$$\frac{Tr(H)^2}{Det(H)} \left\langle \frac{(\gamma+1)^2}{\gamma} \right\rangle \quad (8)$$

Which eliminates the low contrast and edge response points, and the SIFT algorithm is improved the ability to resist noise.

### C. Keypoint Orientation Invariance

The scale of the keypoint is used to select the Gaussian smoothed image, to ensure that the descriptor possesses orientation invariance, we specify the direction parameters by using the gradient direction distribution characteristic of keypoint neighborhood pixels. The gradient magnitude and orientation of the pixel  $(x, y)$  are given as equation (9).

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

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (9)$$

In the above Equation (9),  $L(x, y)$  is the  $(x, y)$  gradient magnitude at this scale,  $m(x, y)$  is  $(x, y)$  orientation. The gradient magnitude  $m(x, y)$  and orientation  $\theta(x, y)$  of each image are precomputed by using pixel differences. The SIFT feature points with location, scale and orientation are generated.

Sampling in the feature points within a region around the keypoint, then an gradient orientation histogram is formed. Every 10 degree of the gradient histogram is as a column, the orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and choose main peak value of the histogram as principal orientation of the feature points, the local value more than 80% of the peak value is selected as auxiliary orientation.

### D. Generation of Feature Descriptor

In order to achieve orientation invariance, the coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation. To create the keypoint descriptor, firstly the gradient magnitude and orientation at each image sample point are computed in a region around the keypoint location. These samples are then accumulated in orientation histograms that summarize the contents over  $4 \times 4$  subregions, the length of each arrow corresponds to the sum of the gradient

magnitudes near that direction within the region. Finally, a  $4 \times 4 \times 8 = 128$  element feature vector is generated for each keypoint. At present, feature vector has removed the influence of geometrical deformation factors, such as scale change, rotation, etc. then the vector is normalized to unit length. so the feature vector is reduced the effects of illumination change. The description of the keypoints is shown in Figure 2.

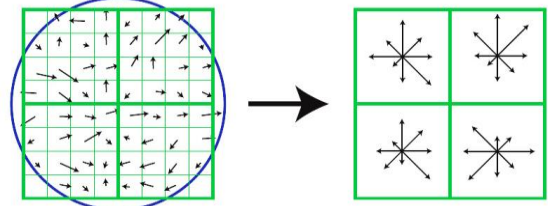


Figure 2. Creating the keypoint descriptor

### E. Feature Point Matching

For image matching, we need extract SIFT features from a set of training images in advance and then stored in a database. By comparing each feature individually between the previous database and the new image, A new image will be matched. The best candidate match for each keypoint is acquired by identifying its nearest neighbor in the database of keypoints from training images. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector. Euclidean distance refers to the linear distance between two pixels, in 2D space, The Euclidean distance[9] is given by Equation (10):

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (10)$$

The correct and incorrect matches probability is determined in terms of the ratio of distance from the closest neighbor to the distance of the second closest neighbor of each keypoint. To realize object recognition, we reject all matches if its distance ratio is greater than 0.8, which eliminates 90% of the false matches while discarding less than 5% of the correct matches.

## III. IMPROVED IMAGE MATCHING METHOD

Because the feature vector is as high as 128 dimensional, the large amount of calculation will reduce the matching speed. When applied the SIFT algorithm matching image, we must calculate the distance between the feature points in an image to another image all the feature points, and each distance has 128 dimensional data, the complexity of the calculation was palpable.

Euclidean distance, the most common distance measurement standard, has been used as a matching rule in many image matching algorithms based on SIFT. In this paper, we present an improved SIFT image matching method based on quasi Euclidean distance [10,11]. with the quasi Euclidean distance replaced the Euclidean distance, and through the limit of geometric constraints to eliminate many mistakes on match point, the matching time is shortened, and the matching efficiency of algorithm is improved.

The quasi Euclidean distance is Euclidean matrix according to the level, vertical, and object collection

segmented estimate all the Euclidean distance. It can be described as:

$$d_0 = |x_1 - x_2| + (\sqrt{2} - 1)|y_1 - y_2| \quad (11)$$

Use Obviously, calculation  $d_0$  is simple than calculation  $d$ , but the obtained the calculated values are small, we use appropriate linear combination  $\alpha d$  instead of  $d_0$ .  $\alpha$  is uncertain, after many matching test, the best values of  $\alpha$  is 0.600-1.00, it can make the calculation simple and the calculation of deviation.

In a calculation of  $d$ , the SIFT algorithm need 128 times multiplication and a square root, the improved algorithm only need a multiplication. If the image is generated  $N$  keypoints of 128 dimensional, the reduction of multiplication computation is  $127N$  respectively. Computing the distance of two points with the shortest eigenvector is the matched points of matching image .So in this way can obviously shorten the operation time and improve the efficiency of the algorithm.

#### IV. EXPERIMENTAL RESULTS

Experimental conditions is Windows7 operating system, U5 processor, 1.7GHz frequency, 2GB memory, and programming environment is MATALB software platform.

The images used in the experiment are shown in Fig. 3 , the reference image is taken from real environment. The matching result of SIFT algorithm and improved algorithm are shown in Fig. 4 and Fig. 5 separately. The improved algorithm can effectively increase the number of feature points, and matching result is accurate.



Figure.3 Stereo images

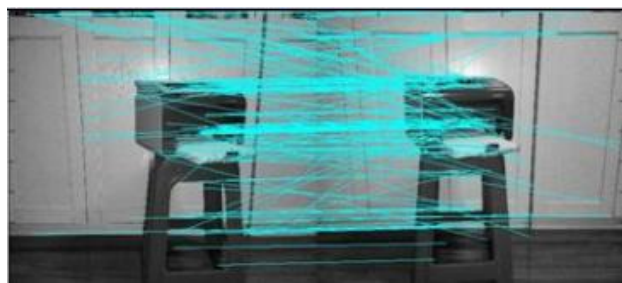


Figure 4. SIFT matching result

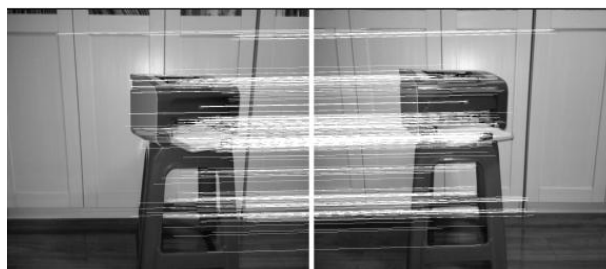


Figure.5 Improved matching result

#### V. CONCLUSIONS

This paper describes a matching algorithm in image processing. Experimental results show that the improved SIFT matching algorithm based on quasi Euclidean distance is effective and feasible. This method can achieve real-time and accurate stereo matching, the complexity of calculation has been reduced, and the amount of calculation has been reduced greatly.

#### REFERENCES

- [1] G. Fu, W. Fu, H. Wu, "A Fast Parallel Image Matching Algorithm Based on MPP Computing Model, " 2012 IEEE International Conference on Computer Science and Automation Engineering, Zhangjiajie, China, vol. 1, pp.765-769, May 2012.
- [2] Wang Xin, Huang Kuan, Gao Zhijun. The Primary Users' Signals Recognition Algorithm in Cognitive Radio Networks via KPCA and Random Forest[J].ICIC Express Letters,2015, 9(4):1083-1088.
- [3] Lowe David G., "Object Recognition from Local Scale-invariant Feature," International Conference on Computer Vision, Kerkyra, Greece, vol. 1, pp. 1150-1157, September 1999 .
- [4] L. Zhang, D.Wang, and A. Vincent, "Adaptive Reconstruction of Intermediate Views from Stereoscopic Images," IEEE Trans. Circuits Syst. Video Technol., vol. 16(1), pp. 102-113, 2006.
- [5] P.J. Besl and N.D. McKay, "Method for registration of 3-D shapes," International Society for Optics and Photonics , vol.6, pp. 586-606, 1992.
- [6] Wang Xin,Guo Lili,Ma Lina. Cooperative Spectrum Sensing Algorithm Based on Second User Selection and Random Forest Classification[J].ICIC Express Letters, Part B: Applications,2015, 6(3):845-850.
- [7] Rafael C.Gonzalez and Richard E.Woods , "Digital Image Processing," Beijing : Electronic Industry Press, 2010.
- [8] LU Guanming, LIU Ruizhen,"3D face recognition based on SIFT feature descriptor in circular neighborhood," Journal of Nanjing University of Posts and telecommunications , vol.35, pp.11-16, Jun. 2015.
- [9] LI A, SHAH S, GAO W, "Coupled bias-variance tradeoff for cross-pose face recognition," Image Processing, IEEE Transactions on , vol.21(1), pp. 305-315, 2012.
- [10] D Z CHENG, Y J LI, and R X YU, "Image Matching Method Based on Improved SIFT Algorithm," Computer Simulation, vol. 28(7), pp. 285-288, 2011.
- [11] Z X Le: "Digital image information processing," Beijing : National Defence Industry Press, 2003.