

Improved Characters Feature Extraction and Matching Algorithm Based on SIFT

Yueqiu Jiang*, Yiguang Cheng, Hongwei Gao

Shenyang Ligong University, Shenyang, China

*Corresponding author, e-mail: missjiangyueqiu@sina.com

Abstract

According to SIFT algorithm does not have the property of affine invariance, and the high complexity of time and space, it is difficult to apply to real-time image processing for batch image sequence, so an improved SIFT feature extraction algorithm was proposed in this paper. Firstly, the MSER algorithm detected the maximally stable extremely regions instead of the DOG operator detected extreme point, increasing the stability of the characteristics, and reducing the number of the feature descriptor; Secondly, the circular feature region is divided into eight fan-shaped sub-region instead of 16 square sub-region of the traditional SIFT, and using Gaussian function weighted gradient information field to construct the new SIFT features descriptor. Compared with traditional SIFT algorithm, The experimental results showed that the algorithm not only has translational invariance, scale invariance and rotational invariance, but also has affine invariance and faster speed that meet the requirements of real-time image processing applications.

Keywords: MSER algorithm, Feature Extraction, Character Recognition, SIFT algorithm

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Scale Invariant Feature Transform (SIFT) [1] is a well-known computer vision algorithm. It is an algorithm of feature point detection and matching which has translation, rotation and scale invariance and has a certain degree of robustness at the same time. SIFT algorithm can be divided into two parts: the feature points location detection and feature vector extraction and matching [2].

Early proposed feature detection operator (Detector) mainly includes: Morava corner detector, Harris detector, Harris-Laplace detector, DOG detector, etc. [3], but these detectors are not effective against affine transformation. In order to solve this problem, some characteristic detection operators with affine invariance are successively proposed mainly include Harris/Hessian-Affine, maximum stable extremely (MSER) detection operator and EBR, etc. [4-6]. Mikolajczyk carried on a contrast performance test for numerous detection operators from the Angle of transformation, scale zoom to image compression and so on in 2005 and the results showed that MSER and Hessian-Affine detection operators perform optimally [7]. In addition, the Descriptors proposed in recent years mainly includes: the shape context information [8], multiplex filtering, moment invariant [9], SIFT based on DOG detector and Zernike moment [10, 11]. Mikolajczyk also carried on a series of experiments and performance evaluation to SIFT, moment invariant, Steerable filter and other 10 descriptors, the results show that the correlated method based on SIFT operator is the most stable and has the best performance when the degree of illumination, affine, fuzzy transforms relatively great. SIFT operator has been widely applied owing to operator's significance and robustness [12].

However, the SIFT operator also has its flaws, which limit its application in the modern image processing. Firstly, SIFT is based on DOG detection which extracts circular regions for the feature points location, as a result, it just has scale invariance and can not meet the requirement of the affine invariance. Secondly, the SIFT descriptor is represented by 128 element feature vectors, it will show the disadvantages of time consuming and huge storage space cost on the case that there are many feature points in the image when matching images. To overcome these two problems, a new method of MSER-SITF is proposed, and making the following improvements for the Detector and Descriptor: one is to use MSER detector substitute DOG detector, make the extracted ellipse image region meets the affine invariance; the other is

to calculate the main direction of the region after normalizing the extracted affine invariant elliptical area into a circular area, use SIFT to produce a 128 element descriptor vector and reduce the dimension of 128 element feature vector by the use of PCA (Principal Component Analysis) [13], in order to improve operation efficiency. Numerical experiments verify the effectiveness of the new method

2. Structure of SIFT Descriptor

The structure of SIFT operator mainly includes four stages [10]: DOG scale-space extreme detection, accurate key point location, Orientation assignment, the establishment of feature descriptor.

2.1. Extreme Detection of DOG Scale-Space

Gaussian kernel function is used to analyze the image scale transformation, and adjacent Gaussian images are subtracted to produce the difference-of-Gaussian (DOG) images that build DOG pyramid. Then the local maxima and minima of the difference-of-Gaussian images which will be determined as candidate feature points are detected by comparing a pixel to its 26 neighbors in 3*3 regions at the current and adjacent scales. The cost of this check is reasonably low due to the fact that most sample points will be eliminated following the first few checks.

The scale space of an image is defined as a function, $L(x, y, \sigma)$, that is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (1)$$

Where σ is the scale of an image, \otimes is the convolution operation in x and y , and $G(x, y, \sigma)$ is a two dimensional Gaussian function, and

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

2.2. Accurate Key Point Localization

Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures.

In the SIFT algorithm, in order to improve the stability of the key points, we need a curve fitting for DOG function of scale space. Using the Taylor expansion of the scale-space DOG function:

$$f(\mathbf{x}) = f(\mathbf{0}) + \frac{\partial f^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 f}{\partial \mathbf{x}^2} \mathbf{x} \quad (3)$$

Where $f(\mathbf{x})$ and its derivatives are evaluated at the sample point and $\mathbf{x}(x, y, \sigma)$ is the offset from this point. The location of the extremum $\hat{\mathbf{x}}$ is determined by taking the derivative of this function with respect to \mathbf{x} and setting it to zero, giving

$$\hat{\mathbf{x}} = \frac{\partial^2 f^{-1}}{\partial \mathbf{x}} \frac{\partial f}{\partial \mathbf{x}} \quad (4)$$

This function value at the extremum, $f(\hat{\mathbf{x}})$, is useful for rejecting unstable extrema with low contrast. This can be obtained by substituting equation (4) into (3), giving

$$f(\hat{x}) = f(0) + \frac{1}{2} \frac{\partial f^T}{\partial x} \hat{x} \quad (5)$$

Through a new method for fitting a 3D quadratic function [14] to locate key points at the location and scale of the central sample point, at the same time, the feature points that have low contrast can be removed. In addition, the difference-of-Gaussian function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise. And for stability and good ability to resist noise, it is not sufficient to reject key points with low contrast. In order to overcome these problems and to improve the stability, the Hessian matrix [15] was involved in, and considered its nature, the edge response points of the DOG operator extreme will be removed.

2.3. Orientation Assignment

To achieve invariance to image rotation, the key point descriptor can be represented relative to the orientation by assigning a consistent orientation to each key point based on local image properties. SIFT assigns a local orientation by calculating the gradient magnitude of each extreme point [16]. The scale of the key point is used to select the Gaussian smoothed image, L , with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample $L(x, y)$ at this scale, the gradient magnitude $m(x, y)$ and orientation $\theta(x, y)$ is recomputed using pixel differences:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (6)$$

$$\theta(x, y) = \tan^{-1} \left[\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right] \quad (7)$$

An orientation histogram is formed from the gradient orientations of sample points within a region around the feature point, Peaks in the orientation histogram correspond to dominant directions of local gradients. The highest peak in the histogram is detected, and then any other local peak that is within 80% of the highest peak is used to also create a keypoint with that orientation [10]. Therefore, for locations with multiple peaks of similar magnitude, there will be multiple keypoints created at the same location and scale but different orientations.

2.4. The Local Image Key Point Descriptor

Rotate the Gaussian image according to the current sample point's main orientation, select the neighborhood area of the feature point after rotation as the object, and divide it into 4*4 sub areas, then calculate gradient histogram with 8 orientation bins in each for every sub area, and as a result, the SIFT descriptor is formed from a 4*4*8=128 element feature vector containing the values of all the orientation histogram entries [16]. By far, the SIFT feature vector had removed the influence of geometrical deformation factor such as rotating and scale change, etc. Finally, the feature vector is modified to reduce the effects of illumination change with normalizing the vector to unit length.

3. MSER Elliptical Region Extraction and Orientation Calculation

3.1. MSER Elliptical Region Extraction

MSER algorithm is put forward by Mates. A MSER is carried out to obtain the final area after selecting the appropriate threshold to an image to get connected components, and testing the stability of the connected components.

For an image $I(x), x \in \Lambda$, Λ is a finite set of real functions, τ is a topology parameter, and the element in the Λ represents a pixel, in a word, Λ is defined as. $\Lambda = [1, 2, \dots, n]^n$

τ domains 4 neighborhood and 8 neighborhood, but $n=2$ is not limited.

$S(x)$ is a level set of image $I(x)$ and the grayscale is not more than that in $I(x)$, $x \in \Lambda$.

$$S(x) = \{y \in \Lambda : I(y) \leq I(x)\} \quad (8)$$

The sequence (x_1, x_2, \dots, x_n) is a connectivity sequence of pixels, such as x_i and x_{i+1} are a four neighborhood or a 8 neighborhood, and $i = 1, \dots, n - 1$. The connected component C is a subset of Λ , $C \subset \Lambda$, a couple of pixels $(x_1, x_2) \in C^2$ can be connected with a path in C . If any connected component C contains C is equal to C , we called C the maximum connected component. Extremal region R is defined as the maximum connected components of the level set $S(x)$. The collection of all extrema areas of the image I is represented by $R(I)$.

Among $R(I)$ extremal regions, we are only interested in the special region which can meet a certain steady standard as described below. The standard assumptions a extremum zone R and $I(R)$ is the maximum value of the image can be obtained in R :

$$I(R) = \sup_{x \in R} I(x) \quad (9)$$

Set $\Delta > 0$ set $R_{+\Delta}$ to contain the minimum extremely regions of R and its strength Δ larger than R at least,

$$R_{+\Delta} = \arg \min \{ |Q| : Q \in R(I), Q \supset R, I(Q) \geq I(R) + \Delta \} \quad (10)$$

As the same, set $R_{-\Delta}$ to contain the minimum extremely regions of R and its strength Δ smaller than R at least,

$$R_{-\Delta} = \arg \min \{ |Q| : Q \in R(I), Q \subset R, I(Q) \leq I(R) - \Delta \} \quad (11)$$

Area transformation is defined:

$$\rho(R; \Delta) = \frac{|R_{+\Delta}| - |R_{-\Delta}|}{|R|} \quad (12)$$

If the regional R is the regional minimum transform R is the most stable area. In the following understanding: Whether any extremely region Q contains R or R contains Q , $\rho(R; \Delta)$ is smaller than $\rho(Q; \Delta)$. R and Q are two extreme regions, if $R \supset Q$ and only if the another extremism zone R meets $R \supset R \supset Q$, then $R = R$, saying R contains Q , and the definition works only when Λ is a finite set.

3.2. MSER Regional Fitting

After the completion of the image MSER area detection, it is essential to fit the rule area to ellipse in order to facilitate the normalization and extracting a feature description. The important information of a region and shape is its location, size and orientation, and the oval can be more effective to reflect three types of information. The center of the ellipse as the center of gravity of the MSER, the two axes of the ellipse passed through the center of gravity respectively, Corresponding to the two axes of the second-order central moments in the major axis minor axis direction respectively, the maximum and minimum (Hu proposed Image Moments and Moment-based invariance systematically in 1962).

For an area ξ of the image $I(x, y)$, its $(p+q)$ order two-dimensional geometric moments defined as:

$$m_{pq} = \iint_{\xi} x^p y^q I(x, y) dx dy \quad p, q = 0, 1, 2, 3 \dots \quad (13)$$

Geometric first moment $m_{00} = \sum_{\xi} I(x, y)$ represents the area of a region (MSER), equals to the number of element of which density value is 1.

Geometric first moment m_{01} and m_{10} :

$$m_{01} = \sum_{\xi} xI(x, y), \quad m_{10} = \sum_{\xi} yI(x, y) \quad (14)$$

The position of the center of gravity of the region can be got through standardized calculation.

$$x_c = \frac{m_{10}}{m_{00}}, \quad y_c = \frac{m_{01}}{m_{00}} \quad (15)$$

Center Second Moment is $U_2 = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$. We are more concerned about its center matrix after calculating the center of gravity of the region, and we can get the so-called center matrix with moving the origin point to the center of gravity and calculating, as shown below:

$$\begin{aligned} \mu_{20} &= \sum_{\xi} (x - x_c)^2 I(x, y) \\ \mu_{11} &= \sum_{\xi} (x - x_c)(y - y_c) I(x, y) \\ \mu_{02} &= \sum_{\xi} (y - y_c)^2 I(x, y) \end{aligned} \quad (16)$$

As mentioned above, the long axial direction in the elliptical fitting region θ represented the direction of the region, semimajor w and semiminor l represent the shape of the region, as shown in Figure 1. These three parameters can be got by calculating the center second matrix of the image U .

$$\begin{aligned} w &= \sqrt{\frac{\lambda_1}{m_{00}}} \\ l &= \sqrt{\frac{\lambda_2}{m_{00}}} \\ \theta &= \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \end{aligned} \quad (17)$$

Where λ_1 and λ_2 are two characteristic values of the Second Moment $U = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$, and their specific value are as below:

$$\lambda_1 = \frac{(\mu_{20} + \mu_{02}) + [(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2]^{\frac{1}{2}}}{2},$$

$$\lambda_2 = \frac{(\mu_{20} + \mu_{02}) - [(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2]^{\frac{1}{2}}}{2}$$

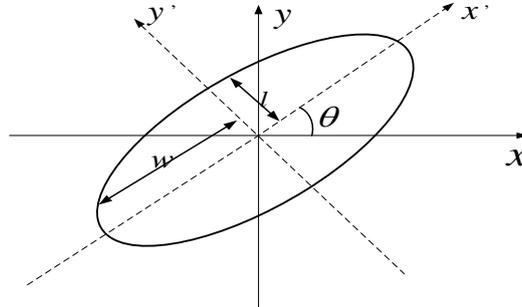


Figure 1. Elliptic rotating schematic diagram

3.3. Orientation Assignment

Similar to the SIFT method of calculating the main direction, this paper also uses gradient direction around the feature point to determine the main direction. Firstly we extract ellipse area though section 3.1, and then normalize the elliptic area into a circular area (32*32 pixels), the normalized affine transformation relation is:

$$x = sA\hat{x} + m, \quad A = 2RD^{\frac{1}{2}} \quad (18)$$

Among them, x is the coordinate of the measurement area, \hat{x} is the coordinate of normalized area, D is the similarity transformation matrix of the covariance matrix generated from ellipse fitting (real symmetric matrix).

Normalization into a circular region aims at making each pixel in elliptical region map to the correct division unit when calculating the gradient distribution. And when calculating the main direction of the normalized circular area, each pixel's gradient and phase of the circular area should be calculated at first, and then weight the amplitude and the Gaussian function of each pixel's phase, and overlay them onto the histogram according to the gradient direction. At last, take the maximum value of phase histogram as the main direction of the current feature point. When other directions are close to the direction of peak value, preserve it and identify it as the second main direction. By assigning a stable main direction to each feature point, the descriptor which generated from the main direction has invariance to the rotation of image.

It is noteworthy that, the significance of the normalized circular area and the meaning of the SIFT circular region that determined by the scaling is not the same. Because the former is obtained by conversion to the matrix of the shape which has a nature of affine invariant, and the corresponding image region may not change at all after transformation, while in the SIFT the corresponding two circular regions are determined by the scaling, there will be information redundancy or insufficient information occasionally. Therefore, the main direction based on normalized circular area that calculated in this paper will be more robust in affine transformation, compared with original SIFT.

3.4. Key Point Descriptor

As section 2.4 shows, SIFT feature descriptor is a 128-element vector, this descriptor describes the size of the 8 directions of the 16 sub-regions. Take it into account that the farther the distance from the key point [13], the smaller the impact to the gradient information of feature points. In this paper, the feature points of the area is divided into eight fan-sectors and using Gaussian function to weight the gradient information field to construct a new SIFT feature descriptor. Specific procedures are as follows:

Taking feature point as the center, a circular region the radius of which is r is divided into eight equiangular fan-shaped area, as shown in Figure 2. Lowe noted a 16×16 neighborhood contains sufficient information without causing a large amount of calculation, and thus the approximate size of the feature point neighborhood is used here to construct characteristics descriptor, taking the radius of the circular area as 8.

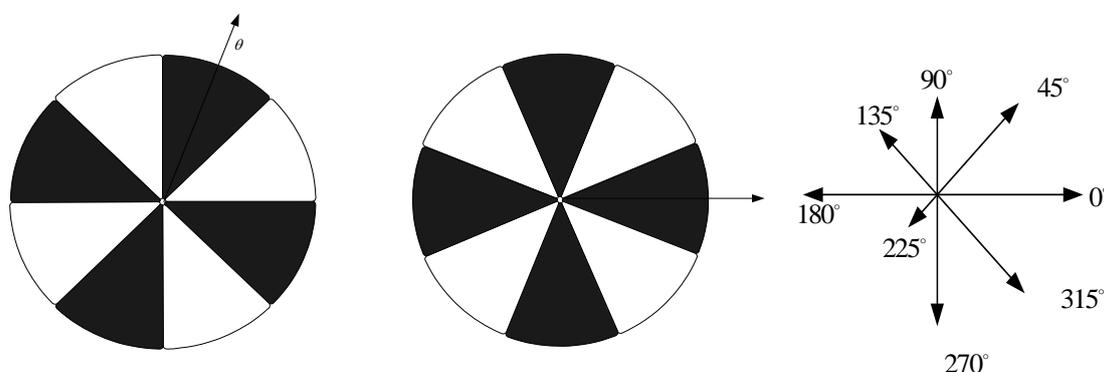


Figure 2. Improved SIFT feature descriptor

Rotate the feature region by the main direction, as shown in Figure 4, after the rotation, calculating eight direction gradient accumulated value of the sector region by a Gaussian function to achieve the descriptor. First, calculate the size and direction of the gradient for each pixel then stats gradient accumulated value of each fan-shaped area in eight directions. In order to reduce the influence of the gradient of pixel away from the feature point to gradient information of the feature point, using a Gaussian function to weight the gradient accumulated value of the feature point. Then, mark the fan-shaped region in a clockwise direction with 1~8, in the 1st region, 8 gradient accumulated value sort as the first to eight elements, in the 2nd region, 8 gradient accumulated value sort as 9 to 16 elements, and so on. 8 sectors for $8 * 8$ elements, the $1 * 64$ vector is defined as a new characteristic descriptor of the feature point. Finally, do a standard normalization processing to this vector to reduce the impact of illumination change to feature descriptor.

New descriptor dimension is from 128 down to 64 dimensions compared with the original characterization descriptor, further reducing the complexity of the algorithm and matching time.

4. Experimental Results And Analysis

In order to verify the run rate of the intra-difference method and the improved SIFT feature extraction algorithm that proposed in this paper, and the validity of the detection of moving targets in complex environments (different lighting conditions, changes in the background and particle noise interference), this chapter made a comparison between the traditional SIFT features extraction [17] and the improved SIFT feature extraction through an image sequence whose maximum size is $768 * 576$ captured by a DH_CG400 capture card and an analog camera. Three random collected pictures in strong, normal and weak light conditions are showed in Figure 3.

Experimental results compared between the traditional SIFT feature extraction and improved SIFT feature extraction such as Table 1.

As shown in Figure 4~7, the traditional SIFT can obtain not only characteristics from the strong to the weak light case, but also a large number of other non-target feature vectors, which will make the next step feature match longer time-consuming; Owing to using MSER algorithm to generate the feature vector on the basis of the maximum stable region, improved SIFT can not only get the target's eigenvectors under different lighting conditions, but also greatly reduce the number of descriptors and thus greatly improved the speed of matching.

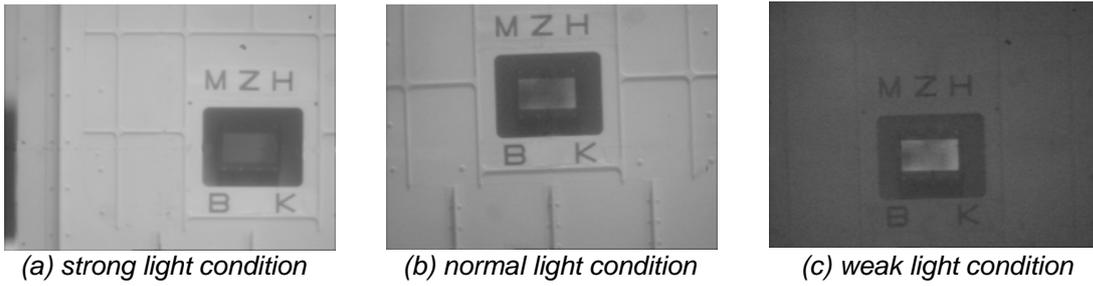


Figure 3. An image in different times under different light conditions

Table 1. The Experimental Data of the Different Images

Group No.	1		2		3		4	
Algorithm Style	traditional	Improved	traditional	Improved	traditional	Improved	traditional	Improved
Extrema Number	554	233	508	296	262	128	365	162
Matching Number	176	161	138	129	46	34	54	44
Running Time(ms)	568	443	711	466	278	177	558	343
Matching Rate(%)	31.77	69.1	27.17	43.58	17.56	26.56	14.79	27.16

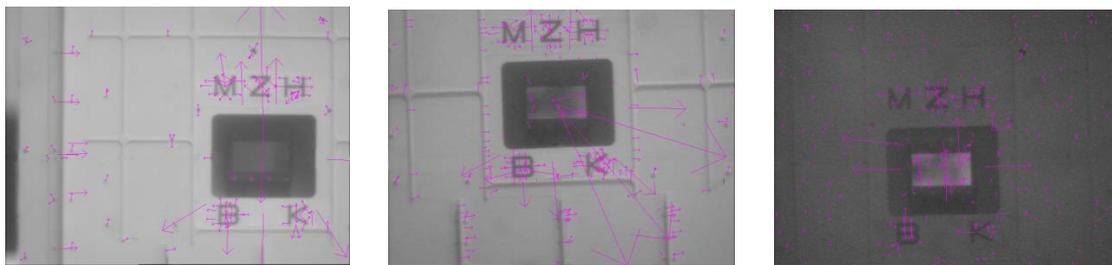


Figure 4. Traditional SIFT feature extraction

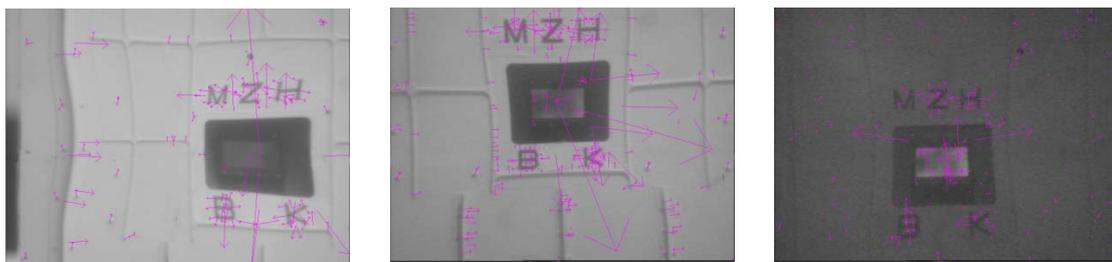


Figure 5. Traditional SIFT feature extraction under affine conditions



Figure 6. Improved SIFT feature extraction



Figure 7. Improved SIFT feature extraction under affine conditions

5. Conclusion

In this paper, the MSER algorithm substitutes the DOG operator which used in traditional SIFT algorithm, not only increasing the stability of the characteristics, but also reducing the number of feature descriptor; followed with a fan-shaped sub-region instead of the traditional square sub-region of the SIFT and the combination of Gaussian function to weighted the gradient information field to construct the SIFT feature descriptor. Taking advantage of the symmetry of the circular domain itself to STATS gradient orientation histogram, and using the coordinate rotation could save the computational cost of image rotation, and reduce the number of dimensions of the feature vector, and also has a certain recognition ability for the small target, at the same time this algorithm can be combined with local information such as the edge further enhanced the effectiveness of the algorithm. Experiments show that the algorithm not only has translational invariance, scale invariance and rotational invariance, but also has affine invariance and faster speed, and this algorithm can meet the requirements of real-time image processing compared with the traditional SIFT algorithm. However, SIFT algorithm is prone to using the multi-classification algorithm based on the minimum distance after detecting the interest points, and that will affect the robustness of the algorithm, and in addition this, there are many assumptions in the PCA model determines certain restrictions to this algorithm. Thus the research on the more robust descriptors based on Hessian-Affine detector that replaced SIFT to extract sub image area, as well as research on descriptor dimensionality reduction trial using NLPCA will be the direction of future efforts.

Acknowledgements

This work is supported by Liaoning Province Colleges and Universities Excellent Talents Support Program. (Grant No.LR201034), CALT Innovation Foundation (Grant No.20130423), State Key Laboratory of Robotics Foundation, Shenyang Institute of Automation, Chinese Academy of Sciences (Grant No.2012017) and also supported by the Liaoning Province Educational Office Foundation of China (Grant No.L2011038).

References

- [1] Zhang Liang, Wang Haili, Wu Renbiao. Matching of Interesting Points Based on Improved SIFT Algorithm. *Electronics & Information Technology*. 2009; 31(11): 2620-2625.
- [2] Liang Xingzhong. Research on Image Processing System Based on Machine Vision. m.a. Thesis. Shandong University; 2004.
- [3] K Mikolajczyk, C Schmid. *Indexing based on scale invariant interest points*. Los Alamitos: IEEE Computer Soc, 2001.
- [4] K Mikolajczyk, C Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*. 2004; 60(1): 63-86.
- [5] J Matas, O Chum, M Urban. Robust wide-base-line stereo from maximally stable extremal regions. *Image and Vision Computing*. 2004; 22(10): 761-767.
- [6] T Tuytelaars and L Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*. 2004; 59(1): 61-85.
- [7] An Ning, Lin Shuzhong, Liu Haihua, Chui Hui. Study on Method & Application of Image Processing. *Instrumentation Reported*. 2006; 27(6): 792-794.
- [8] Song Ming. Image Segmentation and the Application in Medical Image Based on Mathematical Morphology. m.a. Thesis. Yangzhou University; 2005

- [9] Fan Zhibin. Robot Object Recognize and Location Based on Image Local Invariant Features. m.a. Thesis. Beijing Jiaotong University; 2011.
- [10] D Lowe. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision*. 2004; 60(2): 91-110.
- [11] C Singh and E Walia. Fast and numerically stable methods for the computation of Zernike moments. *Pattern Recognition*. 2010; 43(7): 2497-2506.
- [12] Li Xiaoming, Zheng Lian, Hu Zhanyi. SIFT Based Automatic Registration of Remotely-sensed Imagery. *Journal of Remote Sensing*. 2006; 10(6): 829-835.
- [13] Ke Y, Sukthankar R. *PCA SIFT: A more distinctive representation for local image descriptors*. Proceedings of the conference on computer Vision and Pattern recognition. Washington, USA. 2004; 511-517.
- [14] Hu Yingfeng. Research on a three-dimensional reconstruction method based on the feature matching algorithm of a scale-invariant feature transformation. *Mathematical and Computer Modeling*. 2010; 54(3): 919-923.
- [15] Wang Wei, Li Wenhui, Wang Chengxi, Xin Huijie. A Novel Watermarking Algorithm based on SURF and SVD. *TELKOMNIKA Indonesian Journal of Electrical engineering*. 2013; 11(3): 1560-1567.
- [16] Wang Peng, Wang Ping, Shen Zhenkang, Gao Yinghui. Qu zhiguo. A Novel Algorithm for Affine Invariant Feature Extraction Based on SIFT. *Signal Processing*. 2011; Vol. 01, page: 88-93.
- [17] Geng Nan, He Dongjian, Song Yanshuang. Camera Image Mosaicing Based on an Optimized SURF Algorithm. *TELKOMNIKA Indonesian Journal of Electrical engineering*. 2012; 10(8): 2183-2196.