A stereo matching algorithm based on SIFT feature and homography matrix*

LI Zong-yan (李宗艳)[1], SONG Li-mei (宋丽梅)[1], XI Jiang-tao (习江涛)[2], GUO Qing-hua (郭庆华)[1,2], ZHU Xin-jun (朱新军)[1], and CHEN Ming-lei (陈明磊)[1]**

1. Key Laboratory of Advanced Electrical Engineering and Energy Technology, Tianjin Polytechnic University, Tianjin 300387, China

2. School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Keiraville 2500, Australia

(Received 24 July 2015)

Aiming at the low speed of traditional scale-invariant feature transform (SIFT) matching algorithm, an improved matching algorithm is proposed in this paper. Firstly, feature points are detected and the speed of feature points matching is improved by adding epipolar constraint; then according to the matching feature points, the homography matrix is obtained by the least square method; finally, according to the homography matrix, the points in the left image can be mapped into the right image, and if the distance between the mapping point and the matching point in the right image is smaller than the threshold value, the pair of matching points is retained, otherwise discarded. Experimental results show that with the improved matching algorithm, the matching time is reduced by 73.3% and the matching points are entirely correct. In addition, the improved method is robust to rotation and translation.

Document code: A **Article ID:** 1673-1905(2015)05-0390-5

DOI 10.1007/s11801-015-5146-3

Binocular stereo vision is an important branch of computer vision[1-3]. Stereo matching has always been a focus in the field of stereo vision research[4,5]. The stereo matching algorithms can be categorized into three types: area-based matching algorithm[6,7], phase-based matching algorithm[8,9] and feature-based matching algorithm[10-13]. The algorithm of area matching has the following drawbacks: it is sensitive to the affine distortion and radiation distortion; it is lack of robustness against the impact of image noise and gray value differences or contrast differences; it is difficult to choose the size of matching window. The phase-based matching algorithm has good inhibition on high-frequency noise images. However, when the assumption is not held in the two matching images, the phase-based matching algorithm will lose its efficiency due to low magnitude of bandpass output signal. That is the problem of phase singularity. Feature-based matching algorithm forms sparse disparity map. Feature matching algorithm lays more emphasis on the space scene structure information to solve the matching ambiguity problem. Feature matching has strong robustness in many aspects when dealing with stereo vision problems.

On the basis of the former research of our laboratory[14,15], an improved stereo matching algorithm based on scale-invariant feature transform (SIFT) feature matching algorithm and homography matrix is proposed in this paper. The algorithm can improve the speed of SIFT feature matching and is robust to rotation and translation. The SIFT feature is rich in information and suitable for fast and accurately matching the feature in massive feature databases. Even though there are few objects in the scene, a large number of SIFT feature vectors can also be produced, which can easily be combined with other forms of feature vectors. The optimized SIFT matching algorithm can even meet the real-time requirement.

SIFT feature matching algorithm mainly includes the following steps: firstly, scale space is generated and feature points in the scale space are detected and extracted, and then feature points are accurately located, and direction parameters are specified for each key point; finally, the descriptor of key points is generated in order to complete feature matching between two images.

The purpose of scale space theory is to simulate multi-scale characteristics of image data. Gaussian convolution kernel is the only linear nucleus to realize the scale change. The space scale of a 2D image is defined as:

* This work has been supported by the National Natural Science Foundation of China (Nos.60808020 and 61078041), the National Science and Technology Support (No.2014BAH03F01), the Tianjin Research Program of Application Foundation and Advanced Technology (No.10JCYBJC07200), and the Technology Program of Tianjin Municipal Education Commission (No.20130324).

** E-mail: lilymay1976@126.com

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y)\,, \tag{1}$$

where $G(x,y,\sigma)$ is a scale-variable Gaussian function expressed as

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

where (x,y) is space coordinate, and σ is scale coordinate. The value of σ determines the smoothness of the image. The large scale is corresponding to the general features of image, while the small scale is corresponding to the detail features of the image. Here, feature points are detected in difference of Gaussian scale-space. The difference of Gaussian scale space can be constructed as:

$$D(x,y,\sigma)=[G(x,y,k\sigma)-G(x,y,\sigma)]*I(x,y)=$$
$$L(x,y,k\sigma)-L(x,y,\sigma)\,. \tag{3}$$

SIFT feature point is the extreme value point of scale-space. In order to find extreme value point of scale space, each sample point should be compared with 26 sample points in the adjacent scales and the adjacent position in the same scale to get the candidate feature points.

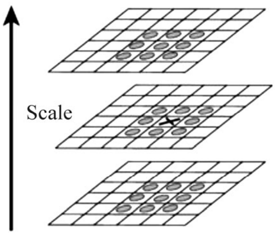

Fig.1 The detection of feature points in difference of Gaussian scale space

The candidate feature points have been already identified in above section. In order to improve noise immunity and enhance the stability of the matching, the key points of low-contrast and unstable edge response points should be removed.

Taylor expansion of the difference of Gaussian scale space function can be expressed by the following formula:

$$D(x,y,\sigma)=D(x,y,\sigma)+\frac{\partial D^{\mathrm{T}}}{\partial x}x+\frac{1}{2}x^{\mathrm{T}}\frac{\partial^2 D}{\partial x^2}x\,. \tag{4}$$

By the above formula, the precise position $\hat{x}$ of candidate points can be obtained:

$$\hat{x}=\frac{\partial D^{-1}\partial D}{\partial x^2\partial x}\,. \tag{5}$$

Take Eq.(5) into Eq.(4) to obtain $D(\hat{x})$,

$$D(\hat{x})=D(x,y,\sigma)+\frac{1}{2}\frac{\partial D^{\mathrm{T}}}{\partial x}\hat{x}\,. \tag{6}$$

In order to remove feature points with low contrast, 0.03 is selected as threshold value. If $|D(\hat{x})|\geq 0.03$, the feature point is retained, otherwise discarded. Therefore, the position and scale of feature point can be expressed as:

$$\hat{X}=(x,y,\sigma)^{\mathrm{T}}\ . \tag{7}$$

At the same time, the points of unstable edge response can be eliminated by analyzing the property of Hessian matrix of extreme point, thus the stable feature points are selected.

In order to make the descriptor rotation-invariant, the gradient direction distribution characteristics of neighboring pixels at feature point are used to specify direction parameters for each feature point.

$$m(x,y)=$$
$$\sqrt{[L(x+1,y)-L(x-1,y)]^2+[L(x,y+1)-L(x,y-1)]^2}\,, \tag{8}$$

$$\theta(x,y)=\arctan[L(x,y+1)-L(x,y-1)]/$$
$$[L(x+1,y)-L(x-1,y)]\,. \tag{9}$$

$m(x,y)$ and $\theta(x,y)$ are modulus value and direction of the gradient at point (x,y), respectively. L is the value of the point on the scale of the feature point. A neighborhood window is created using key point as the center, and the gradient direction of neighborhood pixels in the neighborhood window is added up using histogram. Histogram peak represents the main direction of gradient of feature point neighborhood, and it is used as the direction parameter θ of feature point.

So far, detecting SIFT feature points has been completed. There are three pieces of information for each feature point: position (x,y), scale σ and direction θ.

Firstly, the axis is rotated to the direction of the feature point to ensure its rotation-invariant property, as shown in Fig.2.

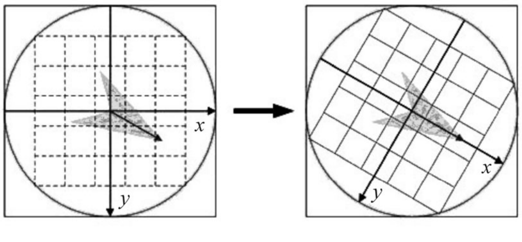

Fig.2 Rotating the axis to the direction of the feature point

Then a 16×16 sample window is created using feature point as the center, and the sample window is divided into 4×4 sub-blocks. And the relative Gaussian-weighted directions of sampling points and feature points are classified into eight-direction direction histogram. Finally the 128-dimensional feature descriptor is obtained (Fig.3).

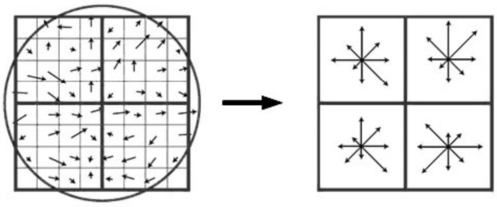

Fig.3 Obtaining the feature descriptor

C_1 and C_2 in Fig.4 are the optical centers of left camera and right camera, and P_1 and P_2 are projections of the point P in 3D space on imaging planes of left camera and right camera, respectively. P, C_1 and C_2 constitute a plane S in 3D space, and the intersecting line L_1 of left camera imaging plane and plane S goes through the point P_1. The intersecting line L_1 is called as the corresponding epipolar line of point P_2. Similarly, the intersecting line L_2 of right camera imaging plane and plane S is called as the corresponding epipolar line of point P_1.

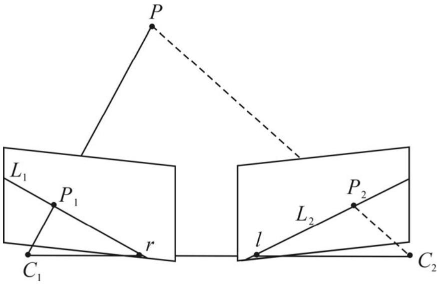

Fig.4 The schematic diagram of epipolar constraint

The fundamental matrix $\boldsymbol{F}$ and one pair of projection points of any space point on left camera and right camera imaging planes follow the constraint of:

$$\boldsymbol{m} \cdot \boldsymbol{F} \cdot \boldsymbol{m}^{'} = 0\,, \tag{10}$$

where $\boldsymbol{m}$ and $\boldsymbol{m}'$are projection points on left camera and right camera imaging planes, respectively, and $\boldsymbol{F}$ can be obtained from internal and external camera parameters. Therefore, when the fundamental matrix and one projection point are known, the epipolar constraint equation of the corresponding projection point on the other camera imaging plane can be obtained.

Due to the measurement error and uncertainty of the camera position and orientation, the corresponding point may not accurately appear in the corresponding epipolar line. Therefore, the corresponding feature point should be searched within a small neighborhood of epipolar line.

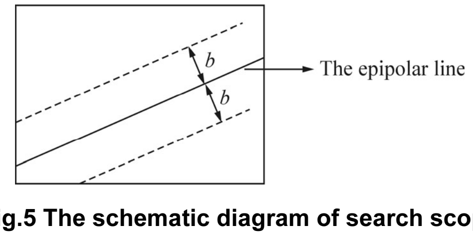

Fig.5 The schematic diagram of search scope

If the certain interval of the distance between the object and the camera is known, the search scope can be limited to a small interval of the epipolar line, as shown in Fig.6. Therefore, the epipolar constraint can greatly reduce the search space of finding the corresponding points. With epipolar constraint, the speed of matching can be improved and the number of false match points can be reduced.

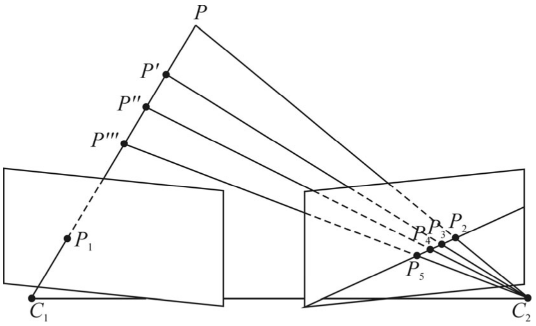

Fig.6 The schematic diagram of search scope limited to a small interval of the epipolar line

Homography $\boldsymbol{H}$ represents a mapping relationship. After giving one point of an image, the only corresponding point can be found in the other image. Assuming $\boldsymbol{c}$ is a point in the left image, and $\boldsymbol{d}$ is the corresponding point in the right image, the following relationship exists between them:

$$\boldsymbol{c} = \boldsymbol{Hd}\,, \tag{11}$$

where $\boldsymbol{H}$ is a 3×3 matrix. If four pairs of corresponding points are known, $\boldsymbol{H}$ can be obtained. The pairs of corresponding points obtained from the previous step are generally more than 4. Therefore, the least square method is used to obtain $\boldsymbol{H}$.

Suppose the obtained pair of points according to the epipolar constraint is ($\boldsymbol{a}$, $\boldsymbol{a}'$), where $\boldsymbol{a}$ is the point on the left image, and $\boldsymbol{a}'$ is the point on the right image. According to $\boldsymbol{H}$, $\boldsymbol{a}$ can be mapped into $\boldsymbol{a}''$ on the right image by $\boldsymbol{Ha}$. If the distance between $\boldsymbol{a}'$ and $\boldsymbol{a}''$ is smaller than the threshold value T (T is 3 pixels), ($\boldsymbol{a}$, $\boldsymbol{a}'$) is retained, otherwise discarded. The homography constraint is shown in Fig.7.

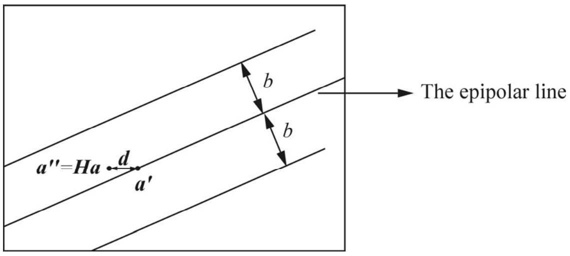

Fig.7 The schematic diagram of homography constraint

In order to verify the effect of our proposed method, the program is implemented in the environment of MAT-

LAB2014. In experiment, the resolution of left camera and right camera is 1 280×1 024. Fig.8 shows the original images captured by the cameras.

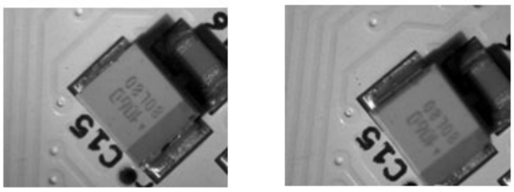
(a) (b)

Fig.8 The original images captured by cameras: (a) Left image; (b) Right image

In order to better realize SIFT feature extraction, the histogram equalization is used for original images to enhance the edge information and detail information before extracting SIFT features. Fig.9 shows left and right histogram-equalized images.

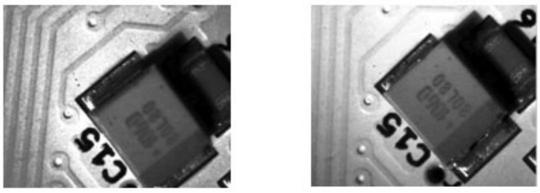
(a) (b)

Fig.9 (a) Left and (b) right histogram-equalized images

Then the SIFT feature extraction is done for treated image, and the threshold value of matching of the feature descriptors is 0.6. Fig.10 is the characteristic direction image of SIFT features. There are 263 SIFT features extracted in the left image and 635 SIFT features extracted in the right image. Fig.11 is the matching image of SIFT matching method. 34 pairs of SIFT feature points are matched in total. From Fig.11, we can see that there exist false match points.

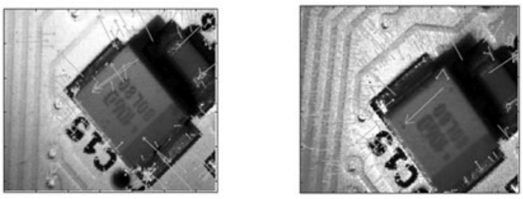
(a) (b)

Fig.10 Characteristic direction images: (a) Left image; (b) Right image

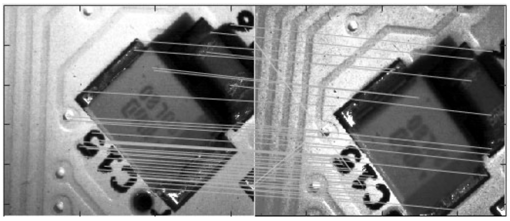

Fig.11 The matching image with SIFT matching method

The homography $\boldsymbol{H}$ obtained by the least square method is:

$$\boldsymbol{H}=\begin{bmatrix} -0.003\,5 & 5.946\,8\times10^{-4} & -0.835\,2 \\ 4.466\,6\times10^{-5} & -0.003\,8 & -0.549\,8 \\ 4.644\,1\times10^{-7} & 2.769\,9\times10^{-7} & -0.004\,5 \end{bmatrix}. \quad (12)$$

Fig.12 is the SIFT feature matching image with epipolar constraint and homography constraint. 28 pairs of SIFT feature points are matched in total. Despite the reduction of the number of matches in SIFT feature points, we can see that there is no error in the matching points from the image.

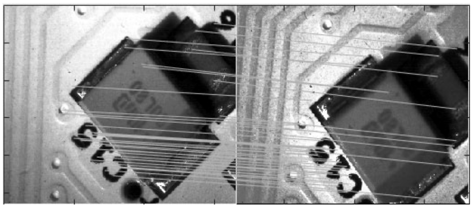

Fig.12 The matching image with our improved method

The matching time of the improved matching method and the SIFT matching method is 4 s and 15 s, respectively.

Since our improved method is based on SIFT feature matching, the improved method is robust to rotation and translation. The comparison experiment of rotation and translation between the improved method and the affine-invariant method is done. The right image is rotated at different angles and translated for different distances, then the left image and the processed right image are matched. The results are shown in Fig.13 and Fig.14.

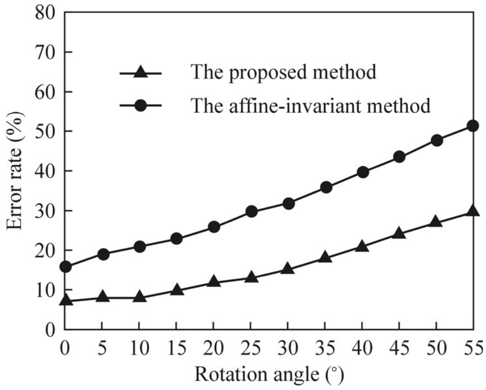

Fig.13 The relationship between error rate and rotation

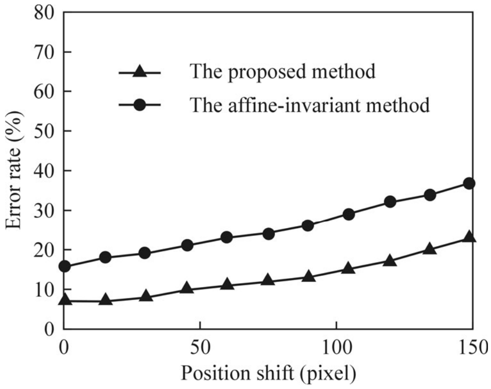

Fig.14 The relationship between error rate and position shift

An improved stereo matching algorithm with SIFT feature and homograph is proposed. Experimental results show that with the improved matching algorithm, the matching time is reduced by 73.3% while the matching points are entirely correct. In addition, the improved method is robust to rotation and translation.

References

[1] M. Cao, G.M. Zhang and Y.M. Chen, Optik - International Journal for Light and Electron Optics **125**, 366 (2014).

[2] P. Zhao and G.Q. Ni, Optics and Lasers in Engineering **48**, 505 (2010).

[3] B. Christian and N. Yang, Procedia Computer Science **39**, 146 (2014).

[4] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis and G. Karras, ISPRS Journal of Photogrammetry and Remote Sensing **91**, 29 (2014).

[5] T. Pribanic, N. Obradovic and J. Salvi, Optics Communications **285**, 1017 (2012).

[6] A. Hosni, M. Bleyer and Margrit Gelautz, Computer Vision and Image Understanding **117**, 620 (2013).

[7] Tingbo Hu, Baojun Qi, Tao Xu and Hangen He, Computer Vision and Image Understanding **116**, 908 (2012).

[8] H.Z. Jiang, H.J. Zhao, X.Y. Liang and D. Li, Optics and Precision Engineering **19**, 2520 (2011).

[9] J. Zhou, Y. Xu and X.K. Yang, Pattern Recognition Letters **28**, 1509 (2007).

[10] S. Ploumpis, A. Amanatiadis and A. Gasteratos, Image and Vision Computing **38**, 13 (2015).

[11] H. Fadaifard, G. Wolberg and R. Haralick, Graphical Models **75**, 157 (2013).

[12] Y.H. Zhao and Y.Q. Chen, Optics and Lasers in Engineering **51**, 213 (2013).

[13] Y.L. Jiang, Y.X. Xu and Y. Liu, Neurocomputing **120**, 380 (2013).

[14] L.M. Song, X.X. Dong, J.T. Xi, Y.G. Yu and C.K. Yang, Optics & Laser Technology **45**, 319 (2013).

[15] L.M. Song, Y.L. Chang, Z.Y. Li, P.Q. Wang, G.X. Xing and J.T. Xi, Optics Express **22**, 13641 (2014).