Journal of Electronic Imaging

JElectronicImaging.org

Feature matching for illumination variation images

Zhenfeng Shao
Min Chen
Chong Liu

Journal of Electronic Imaging 24(3), 033011 (May/Jun 2015)

Feature matching for illumination variation images

Zhenfeng Shao,[a] Min Chen,[b,*] and Chong Liu[a]
[a]Wuhan University, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, No. 129 Luoyu Road, Wuhan 430079, China
[b]Southwest Jiaotong University, Faculty of Geosciences and Environmental Engineering, West Section, High-Tech Zone, Chengdu 611756, China

Abstract. Illumination variability is one of the most important issues affecting imagery matching performances and still remains a critical problem in the literature, although different levels of improvement have been reported in recent years. This study proposes an illumination robust image matching method. There are three steps in the proposed method: first, local regions are extracted and matched from the input images by using a multiresolution region detector and an illumination robust shape descriptor; second, an algorithm is proposed to estimate the overlapping areas of images and enhance them based on the region matches; finally, general feature detectors and descriptors are combined to process the previous results for illumination robust matching. Experimental results demonstrate that the proposed matching method provides significant improvement in robustness for illumination change images compared with traditional methods. *© 2015 SPIE and IS&T* [DOI: 10.1117/1.JEI.24.3.033011]

Keywords: image matching; illumination variation; overlapping area; enhancement.

Paper 14701 received Nov. 9, 2014; accepted for publication Apr. 22, 2015; published online May 21, 2015.

1 Introduction

Image matching is a fundamental issue in scientific applications including target tracking,[1] image registration,[2] 3-D reconstruction,[3] etc. In general, image matching methods can be mainly divided into two categories: area-based and feature-based. Although satisfactory matching accuracy can be expected with area-based algorithms in which imagery is used in the form of a matrix of gray values,[4] they are rather sensitive to image intensity changes and geometric deformations. These disadvantages can be overcome by feature-based methods using local features including patches, corners, junctions, and edges. More specifically, the feature-based methods consist of three steps: feature detection, description, and matching. In feature detection, there are numerous well-known methods including corner detectors,[5–7] blob detectors,[8,9] region detectors,[10,11] etc. In feature description, the scale-invariant feature transform (SIFT)[8] and histograms of oriented gradient[12] are more widely used because of the robustness on scale, blur, and rotation of images. In feature matching, many distances like L1 distance, L2 distance, and histogram intersection distance[13] can be used in practice. The above three steps largely determine the performance of image matching. It should be noted that so far, no feature detectors or descriptors are fully invariant to the change of illumination.[14] Several methods, like SIFT, adopt the strategy of normalization to increase the illumination invariance. More recently, the iterative SIFT (ISIFT)[14] was proposed to cope with this problem by improving the matching framework. ISIFT utilizes an initial matching step to estimate the geometric transformation and extract the overlapping areas of the input images. After that, a histogram matching method is used to normalize the illumination of the two images (see Fig. 1).

Although the ISIFT method greatly improved the matching performance, it is still likely to be invalid given extreme illumination conditions because of two reasons. First, the overlapping area extraction in the ISIFT still depends on the initial SIFT matching results. A large illumination variation between the input images or poor illumination conditions may lead to a failure in the initial matching. Second, the illumination normalization method in ISIFT is based on histogram matching. At least one image with good illumination is necessary. Therefore, the performance of ISIFT is expected to be poor under extreme illumination conditions.

In order to overcome the above problems, a novel illumination robust image matching method is proposed in this paper. First of all, local regions are extracted and matched from the two input images (reference image and test image). Subsequently, a local region-based method is proposed to estimate the overlapping areas and enhance the images. Finally, the scale-invariant difference of gaussians (DoG) detector and SIFT descriptor are adopted to compute and compare features, and the random sample consensus (RANSAC) algorithm[15] is used to reject outliers from the initial matches.

The contributions of this paper lie in the following two items:

(1) An overlapping area calculation and enhancement (OACE) method is proposed to estimate the overlapping areas of images using only one pair of feature matches and it is capable of enhancing images under poor illumination conditions;
(2) An illumination robust image matching framework is proposed combining the OACE method and a popular feature-based matching method.

The rest of this paper is organized as follows. Section 2 presents the proposed matching method in detail. Comparative experimental results and their analysis are provided in Sec. 3. Section 4 concludes this paper and highlights potential improvements in the future.

*Address all correspondence to: Min Chen, E-mail: minchen@home.swjtu.edu.cn

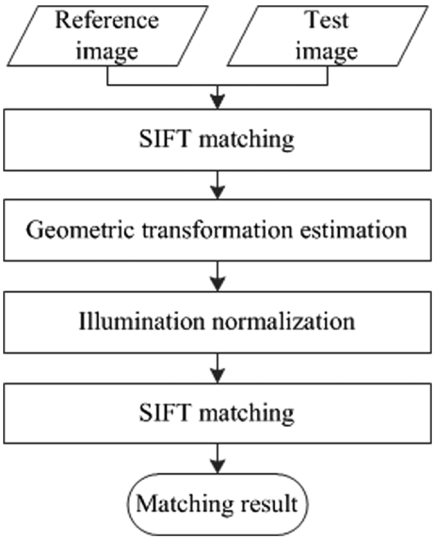


Fig. 1 Framework of the iterative scale-invariant feature transform (ISIFT) method.

2 Methodology

In this study, the image matching process is summarized in Fig. 2.

2.1 *OACE Method*

The proposed OACE model is defined as

$$(O_1, O_2) = \mathrm{M}_{\mathrm{OACE}}(I_1, I_2), \tag{1}$$

where I_1 and I_2 are matrices denoting the input image pair; O_1 and O_2 are matrices denoting the output enhanced images; and $\mathrm{M}_{\mathrm{OACE}}$ is the proposed model. More specifically, $\mathrm{M}_{\mathrm{OACE}}$ is the composition of three successive functions:

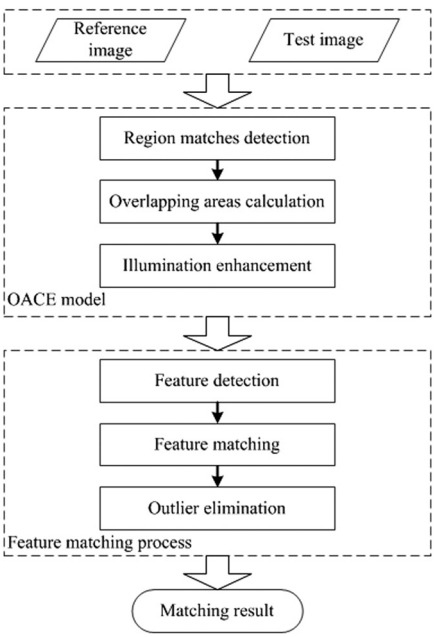


Fig. 2 Framework of the proposed image matching system.

$$\mathrm{M}_{\mathrm{OACE}}(I_1, I_2) = \mathrm{E}\{\mathrm{H}[\mathrm{D}(I_1, I_2)]\}, \tag{2}$$

where $\mathrm{D}:(I_1, I_2) \rightarrow (I_1, I_2, R)$ indicates a transformation for detecting region matches. R is the set of region matches. $\mathrm{H}:(I_1, I_2, R) \rightarrow (O_{o1}, O_{o2}, I_1, I_2)$ represents a function to calculate the overlapping areas of the images I_1 and I_2 based on the elliptical region matches. O_{o1} and O_{o2} denote the overlapping areas. It should be noted that only those images with a small viewpoint change or with planar objects are considered in this paper. Therefore, the geometric transformation between images I_1 and I_2 can be regarded as homography. $\mathrm{E}:(O_{o1}, O_{o2}, I_1, I_2) \rightarrow (O_1, O_2)$ is an image enhancement function. O_1 and O_2 represent the output enhanced images.

2.1.1 *Region matches detection*

The overlapping areas of images are calculated based on local region matches in the OACE model. Several techniques, such as image segmentation and region feature detection, can be adopted for obtaining local regions. In this paper, the latter is adopted because image segmentation is still an unsolved problem and the segmentation accuracy cannot be guaranteed. In region feature detectors, the maximally stable extremal region (MSER) detector is selected to detect local regions due to its two advantages of high efficiency and high matching score in illumination change conditions.[16]

It is important to note that the original MSER algorithm is not scale invariant. In this paper, a multiresolution strategy[17,18] is used to improve the scale invariance of MSER. First, a scale pyramid is constructed by blurring and subsampling with a Gaussian kernel. Then MSERs are detected separately at each resolution image using the method proposed in Ref. 10. Finally, similar MSERs detected from different scale images (termed duplicate MSERs) are removed by eliminating fine-scale MSERs according to the criterions as follows:

(1) The distance between the centroids of the two MSER ellipses should be smaller than four pixels;

(2) The value of $\mathrm{abs}(S_1 - S_2)/\max(S_1, S_2)$ should be less than 0.2, where S_1 and S_2 are the sizes of the two MSER ellipses;

(3) The value of $\mathrm{abs}(\theta_1 - \theta_2) \cdot \max(L_1, L_2)/\pi$ should be less than 4, where θ_1 and θ_2 are the directions of the major axes of the two ellipses, L_1 and L_2 are the perimeters of the two ellipses.

After that, the shape descriptor proposed in Ref. 17 is used to describe features to improve the matching score of MSERs under illumination change conditions.

2.1.2 *Overlapping Areas Calculation*

The overlapping areas are calculated based on the MSER matches:

$$(O_{o1}, O_{o2}, I_1, I_2) = \mathrm{H}(I_1, I_2, R), \tag{3}$$

where I_1 and I_2 represent the input images and R denotes the set of MSER matches. Each MSER feature contains five parameters: the center coordinates (x-axis and y-axis), major axis length, minor axis length, and major axis orientation of the MSER.

O_{o1} and O_{o2} are the overlapping areas on the input images.

An example of one pair of MSER matches in the reference image and the test image is shown in Fig. 3.

In Fig. 3, the shaded areas in the two images denote the overlapping areas. The elliptical region in each image indicates one pair of MSER matches. o_1, l_1, w_1, θ_1 and o_2, l_2, w_2, θ_2 are the center points, major axis, minor axis, and orientations of the two elliptical regions, respectively. The range of $\theta_i (i = 1,2)$ is from 0 to 2π. In the reference image, β is the angle from the positive direction of the major axis to o_1D and in the test image, γ is the angle corresponding to β in the reference image.

The scale changes between the reference image and the test image in the directions of major axis and minor axis are λ_1 and λ_2, respectively. They are defined as

$$\begin{cases} \lambda_1 = l_1/l_2, \\ \lambda_2 = w_1/w_2. \end{cases} \tag{4}$$

In the reference image, the angle from the positive direction of the major axis to o_1D is

$$\beta = \begin{cases} \arctan\left(\frac{y_1}{|AD|-x_1}\right) - \theta_1, & \text{if } \arctan\left(\frac{y_1}{|AD|-x_1}\right) \geq \theta_1 \\ \arctan\left(\frac{y_1}{|AD|-x_1}\right) - \theta_1 + 2\pi, & \text{otherwise} \end{cases}, \tag{5}$$

where x_1 and y_1 are image coordinates of o_1, $|AD|$ is the width of the reference image, and θ_1 is the orientation of the region major axis.

Under a viewpoint change, the scale transforms in different orientations are variable. In consideration of the anisotropy of scale change, the angle corresponding to β in the test image is defined as

$$\gamma = \begin{cases} \arctan(\lambda_1/\lambda_2 \cdot \tan\beta), & \text{if } 0 \leq \beta < \pi/2 \\ \arctan(\lambda_1/\lambda_2 \cdot \tan\beta) + \pi, & \text{if } \pi/2 < \beta < 3\pi/2 \\ \arctan(\lambda_1/\lambda_2 \cdot \tan\beta) + 2\pi, & \text{if } 3\pi/2 < \beta \leq 2\pi \\ \beta, & \text{if } \beta = \pi/2 \text{ or } 3\pi/2 \end{cases}. \tag{6}$$

The scale change in the direction of o_1D is

$$\lambda' = g(\lambda_1, \lambda_2, \beta) = \lambda_1 + 2 \cdot (\lambda_2 - \lambda_1) \cdot \beta/\pi. \tag{7}$$

The coordinates of the point corresponding to D in the test image can be calculated as Eqs. (8) and (9) once the orientation and the scale of o_2d are determined.

$$\begin{cases} x_d = x_2 + \text{dis}_{o_2 \to d1} \\ y_d = y_2 - \text{dis}_{o_2 \to d2} \end{cases}, \tag{8}$$

where x_d and y_d are the coordinates of the point corresponding to D in the test image, and x_2 and y_2 are the coordinates of o_2 in the test image, $\text{dis}_{o_2 \to d1}$ denotes the distance from o_2 to the upright line through point d. $\text{dis}_{o_2 \to d1} > 0$ if o_2 lies on the left side of the upright line. Otherwise, $\text{dis}_{o_2 \to d1} \leq 0$, $\text{dis}_{o_2 \to d2}$ denotes the distance from o_2 to the horizontal line through point d. $\text{dis}_{o_2 \to d2} > 0$ if o_2 lies below the horizontal line. Otherwise, $\text{dis}_{o_2 \to d2} \leq 0$.

Based on the geometric relationship shown in Fig. 3, Eq. (8) can be expressed as

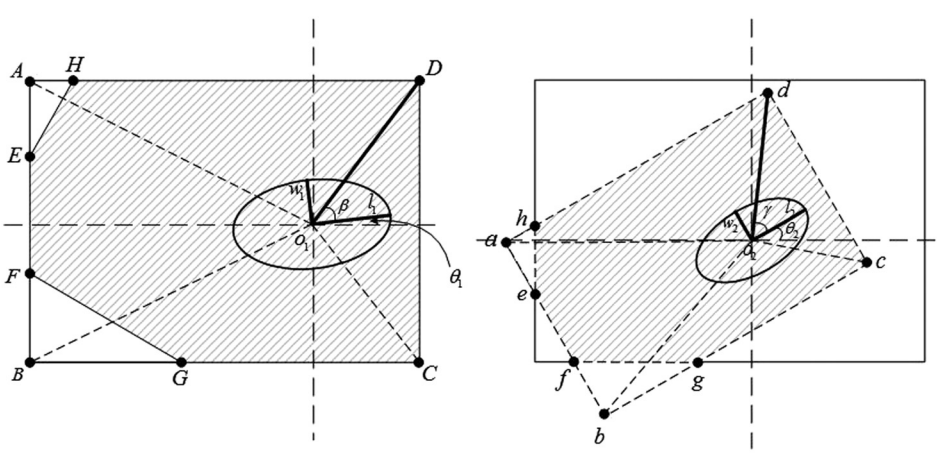
$$\begin{cases} x_d = x_2 + [y_1^2 + (|AD| - x_1)^2]^{1/2} \cdot \cos(\gamma + \theta_2)/\lambda' \\ y_d = y_2 - [y_1^2 + (|AD| - x_1)^2]^{1/2} \cdot \sin(\gamma + \theta_2)/\lambda' \end{cases}. \tag{9}$$

Similarly, in Fig. 3, points of a, b, and c in the test image corresponding to the vertexes A, B, and C in the reference image can be computed. Then the overlapping areas of the two images are determined.

The overlapping areas can be obtained by the least squares method with the use of all MSER matches. However, only the best MSER match is used in the proposed method. An index, defined as match strength (MS), is presented to evaluate the matches based on the feature size and the feature distance in the Euclidean space.

$$\text{MS}_i = (1/D_i)^p \cdot \{T[(A_{il} + A_{ir})/2]\}^q, \qquad i = 1, \ldots, n, \tag{10}$$

where D_i is the Euclidean distance between the two features of the i'th match, A_{il} and A_{ir} are the sizes of the two feature regions, and T is a linear normalization transform, defined as



(a) (b)

Fig. 3 Computation of the overlapping areas: (a) the reference image and (b) the test image.

$$T(x) = \frac{x \cdot \left[\max\left(\frac{1}{D_i}\right) - \min\left(\frac{1}{D_i}\right)\right]}{\left[\max\left(\frac{A_{il}+A_{ir}}{2}\right) - \min\left(\frac{A_{il}+A_{ir}}{2}\right)\right]} + \min\left(\frac{1}{D_i}\right). \quad (11)$$

p and q are two parameters controlling the relative importance between the distance and the size. In practice, the values of parameters p and q can be selected empirically. In this paper, roughly overlapping areas can meet the demand of the subsequent image enhancement. Similarity between MSERs is more important than size. Therefore, a larger p is acceptable. After the MS value of each pair of matches is acquired, the matched pair with the maximum MS is regarded as the best match.

Two images with rotation and scale change shown in Figs. 4(a) and 4(b) are used to test the performance of the proposed overlapping area calculation method. Figure 4(c) is the registered image of Fig. 4(a) based on the reference image Fig. 4(b). The derived overlapping area in the test image using the proposed method is shown in Fig. 4(d). The overlapping area is calculated by using only one pair of MSER matches (the best match). The comparison between Figs. 4(c) and 4(d) indicates that our proposed method is valid for the overlapping area computation. Furthermore, it is noteworthy that the proposed method is also robust for affine variation since it also considers anisotropy.

2.1.3 *Illumination Enhancement*

An improved illumination enhancement method is proposed based on the overlapping areas according to the algorithm in Ref. 19. This method is capable of generating compatible illumination conditions for the overlapping areas which are actually useful for image matching. Since the mean value of the image gray scale reflects the overall sensing of human vision to the image, a nonlinear transformation implemented to the image is defined as

$$I'^2(x,y) = \frac{\sum_y \sum_x I_{\text{overlap}}(x,y)}{127.5^2 \cdot S} \cdot [I_{\text{overlap}}(x,y) - 127.5]^2 + 255 - \frac{\sum_y \sum_x I_{\text{overlap}}(x,y)}{S}, \quad (12)$$

where $I_{\text{overlap}}(x,y)$ denotes a pixel in the overlapping area (the input of the transformation). $I'(x,y)$ is the transformed pixel value. S is the number of pixels in the overlapping area. In this paper, the input data is an 8-bit image, the upper bound of the grayscale value is 255, and the mean value of the numerical range is 127.5.

Based on Eq. (12), a pixel gray value gain is defined as

$$\alpha = \begin{cases} 0, & \text{if } I_{\text{overlap}}(x,y) = 0 \\ I'(x,y)/I_{\text{overlap}}(x,y), & \text{otherwise} \end{cases}. \quad (13)$$

In order to improve the local and global performance at the same time, the transformed pixel value is stretched by multiplying an area gray value gain η:

$$\eta = \frac{127.5 \cdot S}{\sum_y \sum_x I'(x,y)}. \quad (14)$$

The enhancement procedure mentioned above will decrease the gray difference between the adjacent pixels

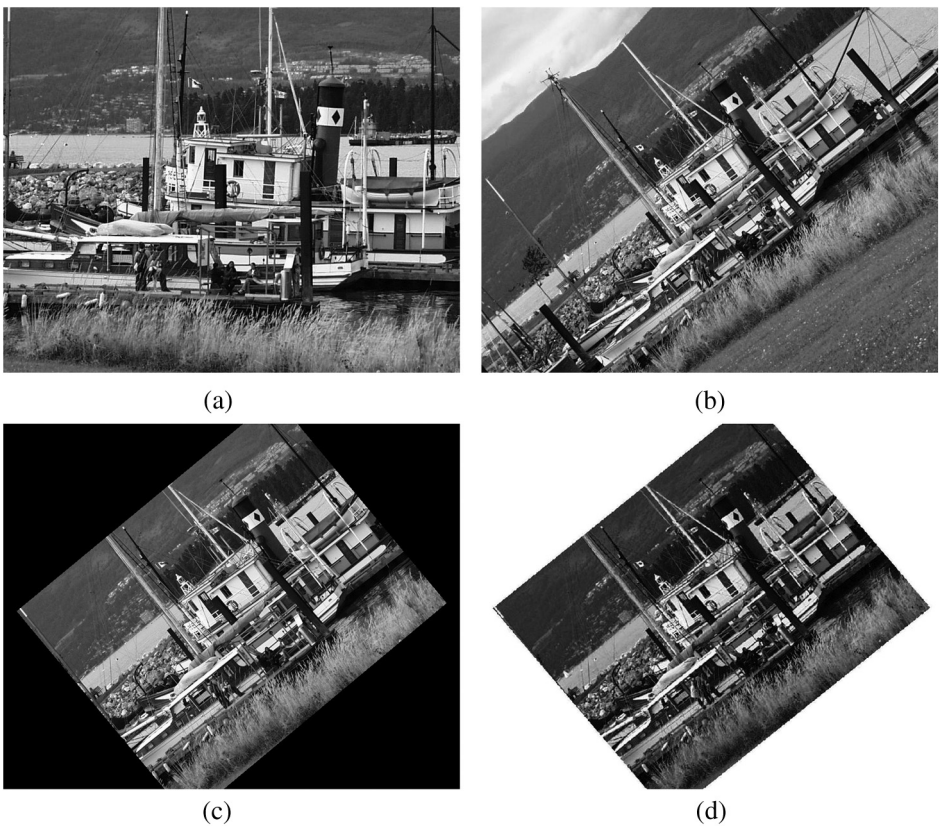


Fig. 4 Result of the overlapping area calculation. Image in (a) is the test image. Image in (b) is the reference image. Image in (c) is the registered image of (a) based on the reference image. Image in (d) is the overlapping area calculated by using the proposed method.

while improving image illumination. Therefore, image edge information is extracted based on B3-spline filter as Eq. (15) to compensate and restore the decreased gray difference between pixels.

$$e = I - B * I, \tag{15}$$

where I is the input image, * is the convolution operator, and B is a 5×5 B3-spline kernel defined as

$$B = \frac{1}{256} \cdot \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}. \tag{16}$$

Therefore, the whole enhancement procedure can be expressed as

$$I_{\text{enhanced}} = \alpha \cdot \eta \cdot I_{\text{input}} + e. \tag{17}$$

One group of tests to evaluate the performance of the gray difference compensation is shown in Fig. 5. Results in Figs. 5(c) and 5(d) show that the proposed gray difference compensation method can restore detailed image information while improving image illumination.

2.2 *Image Matching*

DoG detector and SIFT descriptor[8] are adopted in the image matching process. The DoG detector consists of differences of Gaussian image construction, local extrema detection, accurate points localization, and orientation assignment. The SIFT descriptor is used to construct the local feature once the accurate location, scale, and orientation of the keypoint have been extracted. The SIFT feature descriptor is constructed by first computing the gradient magnitude and orientation at each image sample point within a region around the keypoint location using a Gaussian weighted window. These samples are then accumulated into orientation histograms summarizing the contents over 4×4 subregions to form a 128 element feature vector. Finally, the feature vector is normalized to reduce the effects of illumination change.

Feature vector matching is implemented by using the nearest neighbor distance ratio with the Euclidean distance as the basic metric after the SIFT descriptor computation. The RANSAC algorithm is used to eliminate outliers from the initial results.

3 Experimental Results and Analysis

3.1 *Dataset*

In our experiments, three datasets shown in Figs. 6–8 are used to evaluate the performance of the proposed method.

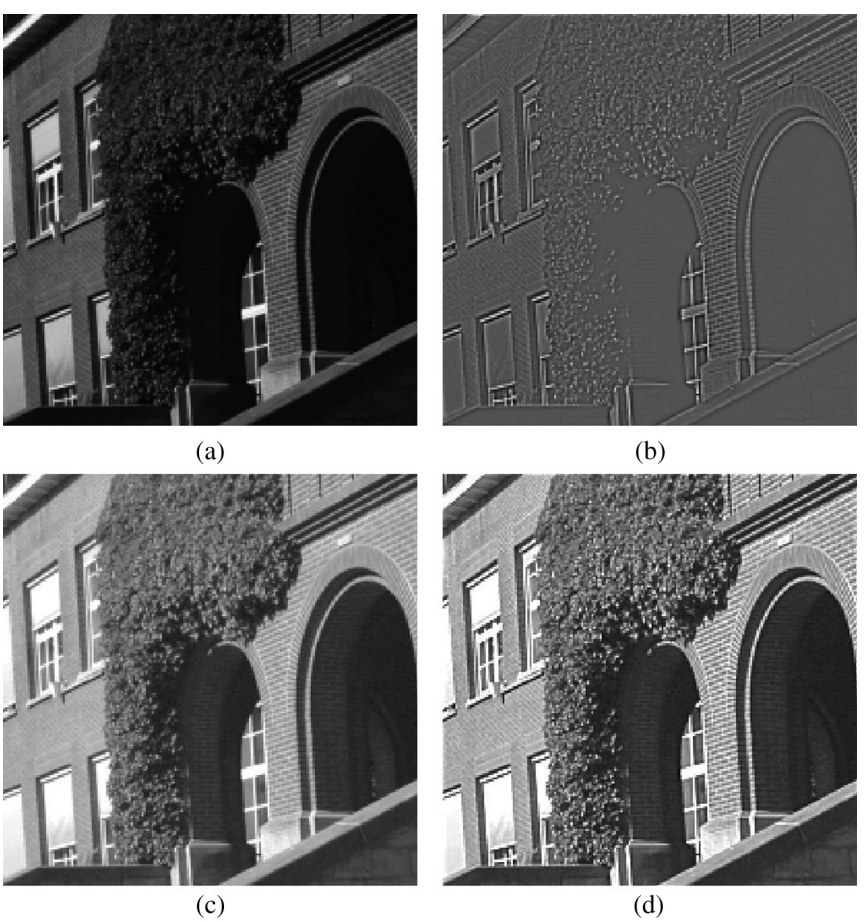


Fig. 5 Results of image enhancement. Image in (a) is the original input image. Image in (b) is the edge information extracted as in Eq. (16). Image in (c) is the enhanced image without gray difference compensation. Image in (d) is the enhanced image with gray difference compensation.

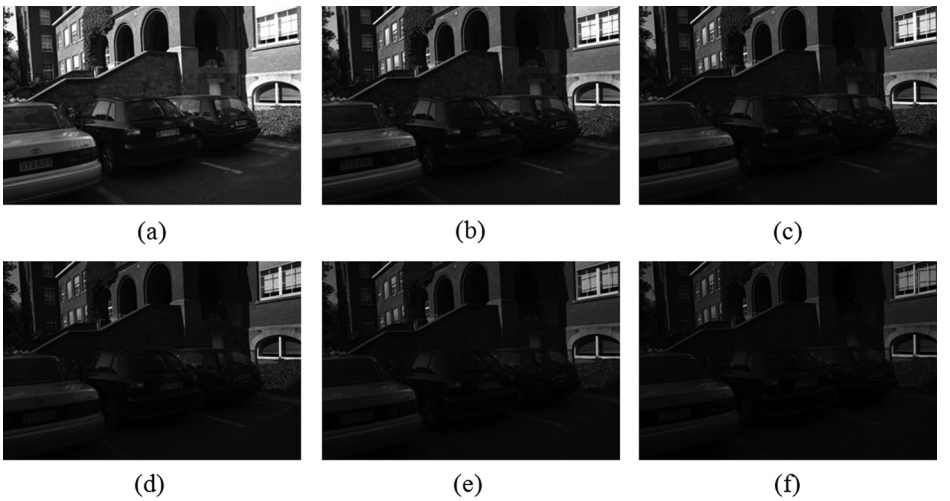


Fig. 6 Leuven: (a) image 1, (b) image 2, (c) image 3, (d) image 4, (e) image 5, and (f) image 6.

The dataset Leuven in Fig. 6 was provided by Mikolajczyk.[20] Images in Fig. 7 were taken by the authors at the local time from 5:30 PM to 8:00 PM. The time interval for each image was 30 min. The dataset in Fig. 8 was obtained by rotating the dataset in Fig. 7.

3.2 *Parameters Setting*

In the experiments, all parameters of each detector and descriptor are set the same as those in the original references.[6,8,14] There are also two other parameters (p and q) in the proposed method that need to be analyzed. The values of p and q can be divided into five kinds of situations:

(1) $p = 0$ and $q > 0$: only the feature size is used to evaluate the MSER match;

(2) $p > 0$ and $q = 0$: only the similarity is used to evaluate the MSER match;

(3) $p > 0$ and $q > 0$ and $p = q$: both feature size and similarity are used to evaluate the match, and the two criteria are equally important;

(4) $p > 0$ and $q > 0$ and $p > q$: both feature size and similarity are used to evaluate the match, but the former is more important than the latter;

(5) $p > 0$ and $q > 0$ and $p < q$: both feature size and similarity are used to evaluate the match, but the latter is more important than the former.

One group of tests has been performed based on the dataset Leuven in Fig. 6 to evaluate the influence of different parameters values to the matching performance. Corresponding to the five situations, the two parameters were set as $(p, q) = \{(0, 1), (1, 0), (1, 1), (2, 1), (1, 2)\}$. The matching results under different parameters values are shown in Table 1.

It can be seen from Table 1 that better matching results have been obtained when $(p, q) = (1, 1)$ and $(p, q) = (2, 1)$. It is consistent with the claim in Sec. 2.1.2 that feature similarity is more important than size and a larger p is acceptable. When the test image is Figs. 6(d) and 6(e), respectively, the matching results remain the same under

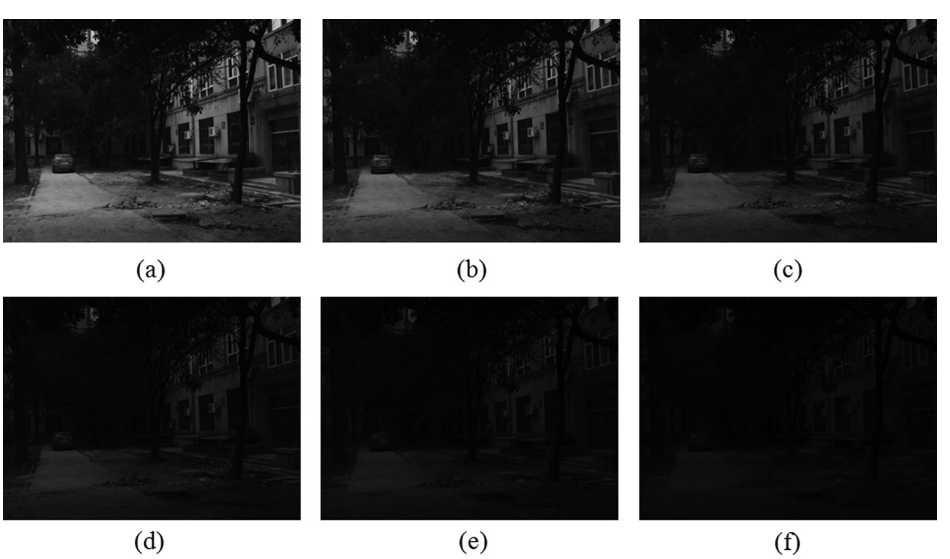


Fig. 7 Reference images: (a) reference image 1, (b) reference image 2, (c) reference image 3, (d) reference image 4, (e) reference image 5, and (f) reference image 6.

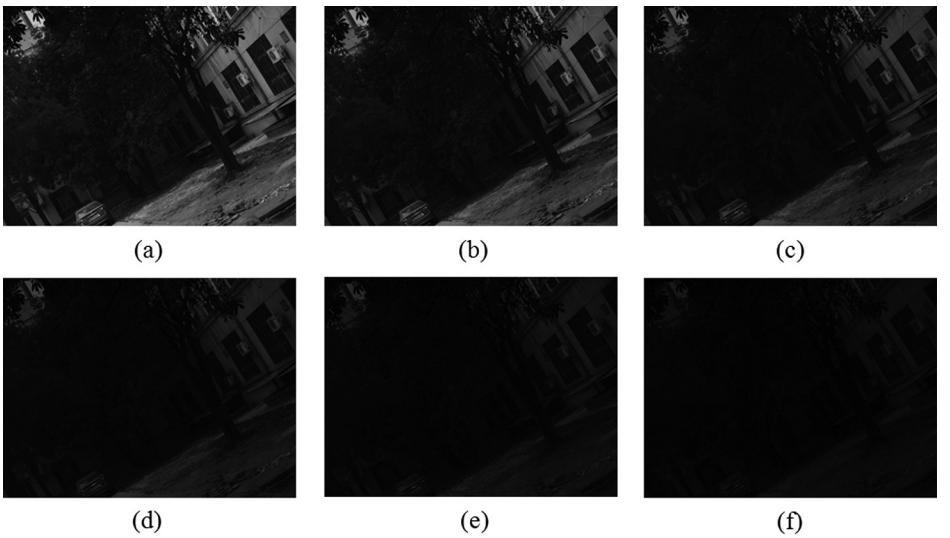


Fig. 8 Test images: (a) test image 1, (b) test image 2, (c) test image 3, (d) test image 4, (e) test image 5, and (f) test image 6.

different parameters values. This is because in the two groups of matching, fewer MSER matches have been found and the most similar features are the largest features. In the following experiments of this paper, both the two parameters are set as 1.

3.3 *Illumination Enhancement Results*

In this part, images in Figs. 6(a) and 6(f) are first selected to compare the illumination enhancement performance of the proposed OACE method with the histogram matching method in Ref. 14. Then the same comparison approach is applied to images in Figs. 6(e) and 6(f). Experimental results are shown in Fig. 9.

The results in Fig. 9 show that the histogram matching method performs well if one of the input images is under good illumination. However, no visual improvement can be observed when both the input two images are under poor illumination. On the contrary, the proposed method can significantly improve the illumination and achieve better performance in both cases.

Table 1 Parameters test results based on images in Fig. 6.

Number of correct matches on dataset Leuven. Reference image: Fig. 6(f)

	Test images				
Parameters	Fig. 6(a)	Fig. 6(b)	Fig. 6(c)	Fig. 6(d)	Fig. 6(e)
$p = 0$ and $q = 1$	372	485	549	649	647
$p = 1$ and $q = 0$	411	501	583	649	647
$p = 1$ and $q = 1$	418	535	600	649	647
$p = 2$ and $q = 1$	418	535	600	649	647
$p = 1$ and $q = 2$	402	517	585	649	647

Besides the visual evaluation, we also use quantify criteria in Ref. 21, the overall contrast $\bar{\sigma}_f$, and the overall lightness $\bar{I}$ of the image, to evaluate the two methods based on all the images in Fig. 6. The evaluation results are shown in Table 2. It can be seen that the values of $\bar{I}$ and $\bar{\sigma}_f$ of the histogram matching method decrease dramatically with the weakening of the reference image illumination. In contrast, the two parameter values of the proposed method always remain at a good level. This demonstrates that the proposed method is superior to the histogram matching method.

3.4 *Image Matching Results*

The image matching experiments include two parts. In the first part, the proposed method is compared with some traditional methods based on the dataset Leuven in Fig. 6, including SIFT, ISIFT, and several other methods using the detectors[6] Harris-Laplace, Harris-Affine, Hessian-Laplace and Hessian-Affine combined with the SIFT descriptor. The sixth image in Fig. 6(f) is selected as the reference image and the other five images are test images. Three indicators: number of correct matches, ratio between the number of correct matches and minimum of total number of features detected from the image pair (termed RS), ratio between the number of correct matches and the number of total matches (termed matching precision) are used to evaluate matching methods. The experimental results are shown from Figs. 10–12.

From the above results, it can be seen that SIFT, ISIFT, and the proposed method perform better than other methods. In the following experiments, some images with worse illumination are used to evaluate the proposed method. These images are shown in Figs. 7 and 8. In this part, six groups of tests are performed. Each reference image in Fig. 7 is matched with all the test images in Fig. 8 except for the one that has the same illumination as the current reference image. Here, the proposed method is only compared with SIFT and ISIFT. The results in Fig. 13 summarize the performance of each algorithm in terms of the number of correct matches under different illumination conditions.

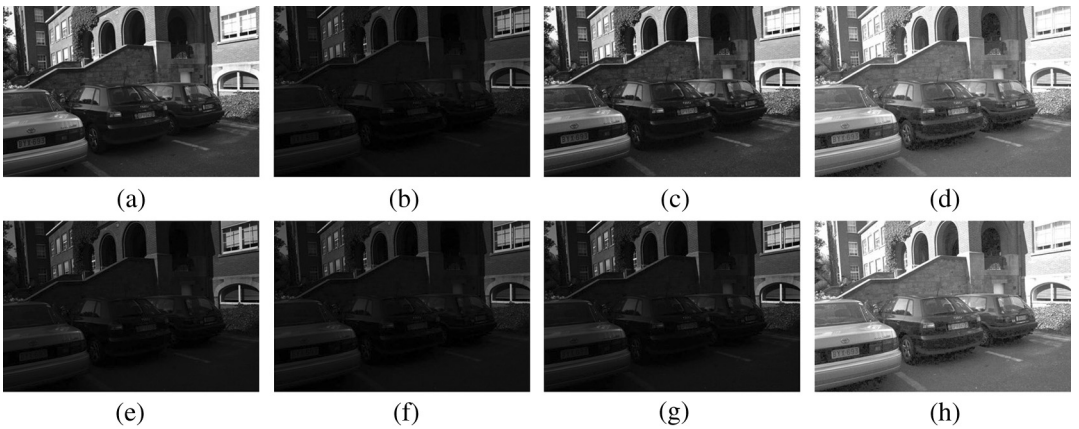


Fig. 9 Results of illumination enhancement: (a) the first image in Leuven; (b) the sixth image in Leuven; (c) the enhanced image from (b) according to the histogram of (a); (d) the enhanced image from (b) by using the proposed overlapping area calculation and enhancement (OACE) method; (e) the fifth image in Leuven; (f) the sixth image in Leuven; (g) the enhanced image from (f) according to the histogram of (e); (h) the enhanced image from (f) by using the proposed OACE method.

Table 2 Illumination enhancement results based on images in Fig. 6.

Test image	Reference Image	$\bar{I}$		$\bar{\sigma}_f$	
		Histogram matching method	Proposed method	Histogram matching method	Proposed method
Fig. 6(f) $\bar{I} = 15.9103$, $\bar{\sigma}_f = 10.1445$	Fig. 6(a)	69.3074	121.6287	33.0712	34.5427
	Fig. 6(b)	44.0530	122.2545	24.2963	35.5364
	Fig. 6(c)	33.9476	121.4127	19.9496	34.5376
	Fig. 6(d)	26.7513	121.6700	16.4666	34.9986
	Fig. 6(e)	22.7118	121.0124	14.3613	34.6113

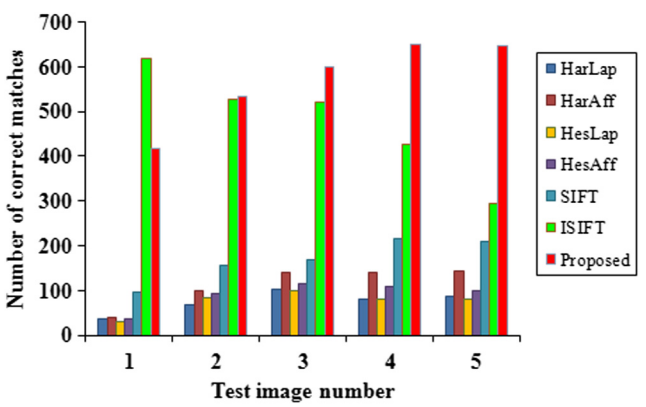


Fig. 10 Numbers of correct matches of different matching methods based on the dataset Leuven.

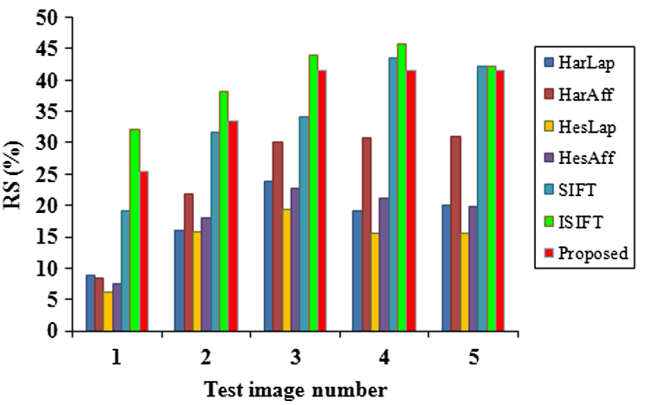


Fig. 11 RS of different matching methods based on the dataset Leuven.

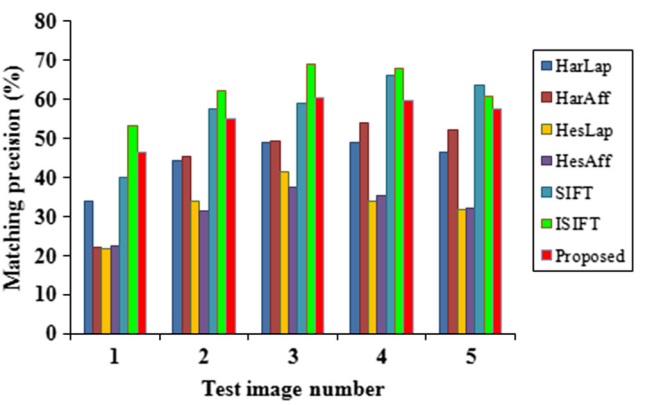


Fig. 12 Matching precision of different matching methods based on the dataset Leuven.

Experimental results in Fig. 13(a) show that the proposed method is strongly robust to illumination change and it derives more the correct matches than those derived from the SIFT and ISIFT methods under the same illumination conditions. Compared with SIFT, which only partially deals with the input image before feature extraction and comparison, ISIFT can obtain more matches based on the histogram matching in the iterative procedure. However, ISIFT does not improve the invariance of the algorithm for illumination change. ISIFT is an iterative method. A set of correspondences are obtained at the first iteration. Then the overlapping area is determined by more than

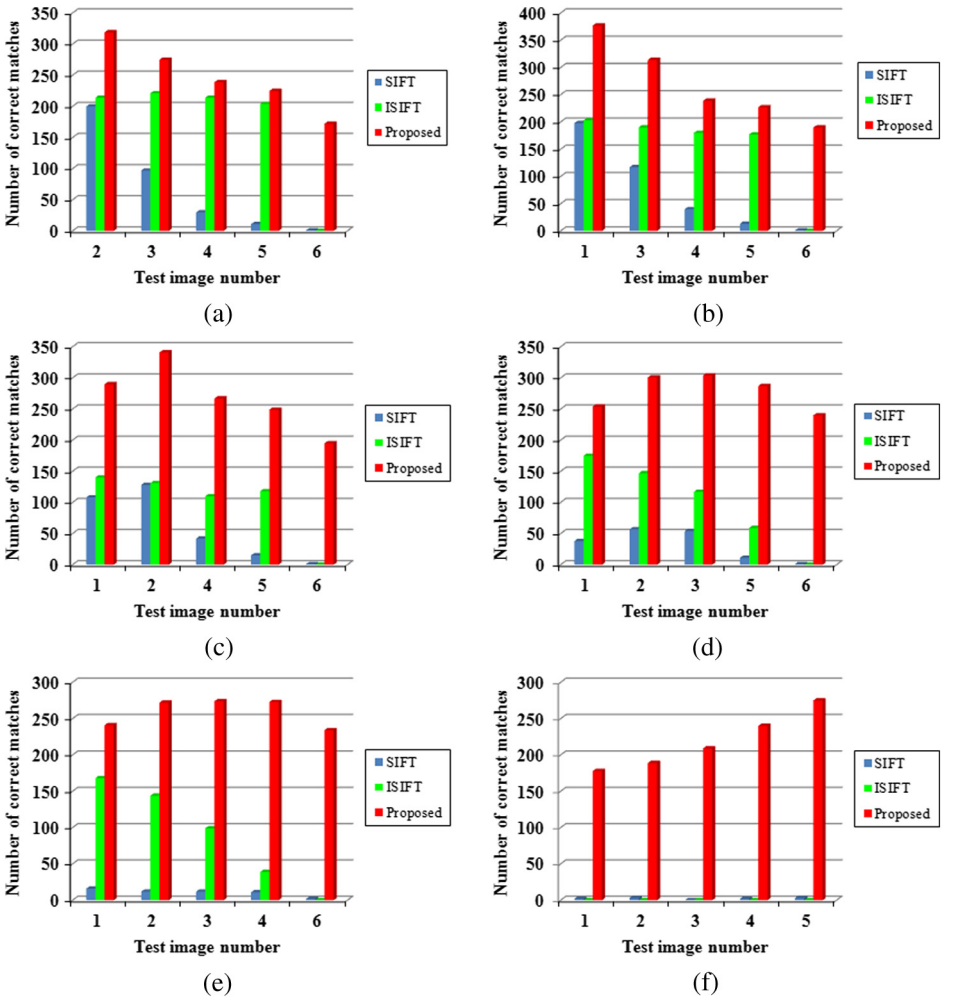


Fig. 13 Number of correct matches of different illumination condition images in the six tests with approaches SIFT, ISIFT, and the proposed method, respectively: (a) reference image 1, (b) reference image 2, (c) reference image 3, (d) reference image 4, (e) reference image 5, and (f) reference image 6.

three initial SIFT matches, and the histogram matching between the reference image and the test image is performed. With the increase of illumination change, the SIFT method can obtain fewer correct matches, and it fails when the change exceeds a certain degree. If the correct matches are less than three, the iterative algorithm of ISIFT will fail and ISIFT will fail accordingly. Comparatively, the proposed method works well even when a large illumination change exists or all the input images are under poor illumination conditions due to the use of the MSER detector and the proposed enhancement algorithm. Similarly, the proposed method performs better than SIFT and ISIFT in the five other groups of experimental results shown in Figs. 13(b) to 13(f).

In the six groups of tests, the matching results of the proposed method for the test image with the worst illumination are illustrated in Figs. 14(a) to 14(f).

In Figs. 14(e) and 14(f), both the input images are under extreme poor illumination conditions. Only one pair of MSER matches has been found between each pair of images. However, the overlapping areas of images have been computed correctly and good matching results have been obtained. This demonstrates the good property of the proposed method by which the overlapping areas of images can be estimated using only one pair of feature matches.

In order to evaluate the feature localization accuracy after image illumination enhancement, the proposed method is compared with SIFT by calculating the root mean square error (RMSE) based on the dataset Leuven in Fig. 6 as Eq. (18):

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} [(x_i' - x_i'')^2 + (y_i' - y_i'')^2]/n}, \tag{18}$$

where n is the number of the selected correspondences $(x_i, y_i) \sim (x_i', y_i')$; (x_i'', y_i'') is the transformed result of (x_i, y_i) by the homography matrix provided by Mikolajczyk.[20] The first image is selected as the reference image and the others are the test images. The computed errors are shown in Table 3.

The results in Table 3 show that both methods yield RMSE within one pixel, which can meet the requirements of general applications. In addition, the RMSE of the proposed method is smaller than that of SIFT in each pair. Therefore, the conclusion can be drawn that the illumination

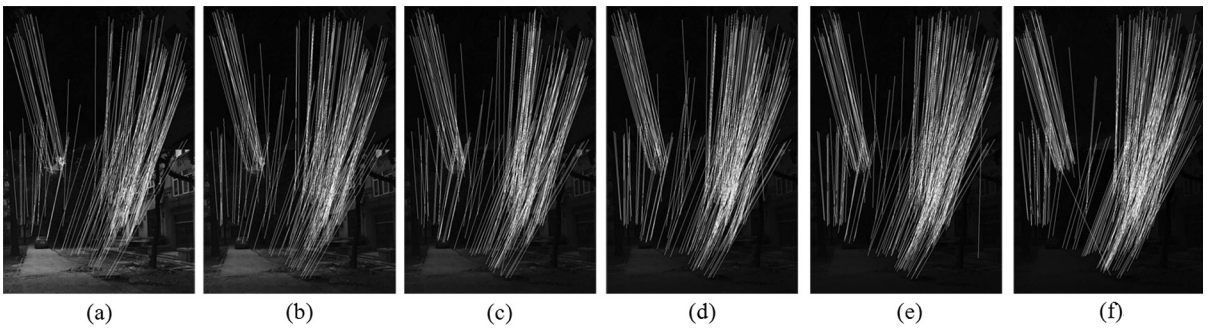


Fig. 14 Partial matching results of the proposed method among the six groups of tests. (a) the result of the reference image 1 and the test image 6; (b) the result of the reference image 2 and the test image 6; (c) the result of the reference image 3 and the test image 6; (d) the result of the reference image 4 and the test image 6; (e) the result of the reference image 5 and the test image 6; and (f) the result of the reference image 6 and the test image 5. The results of matches are drawn in white lines.

Table 3 Root mean square error.

	Image 2	Image 3	Image 4	Image 5	Image 6
Scale-invariant feature transform (SIFT)	0.4055	0.4751	0.5728	0.6003	0.6462
Proposed	0.3548	0.3694	0.5029	0.5628	0.4735

enhancement in the proposed method will not decrease the feature localization accuracy.

3.5 *Method Complexity Analysis*

Computation complexity is an important factor. The proposed method mainly includes four steps: MSER correspondences extraction, overlapping area calculation, illumination enhancement, and SIFT matching. In MSER extraction, since the sort can be implemented as BINSORT and the list of connected components and their areas is maintained using the efficient union-find algorithm, the complexity is almost linear.[10] In the overlapping area calculation, only the best MSER match is adopted. This step is very fast. The following illumination enhancement step is also a very fast step. Therefore, the complexity of the whole method mainly depends on the last step, SIFT matching.

A group of tests based on the dataset Leuven in Fig. 6 has been performed to evaluate the time complexity of the proposed method. The test results are shown in Table 4. The computation times mentioned in the table have all been measured on an Intel Core 2 2.1 GHz Windows XP PC. Even though the computation time is related to the methods of implementation and the image content, we believe that the table gives a reasonable indication of a typical computation time.

Table 4 Comparison of computation times of different method.

	Computation time for the five pairs of images in Fig. 6. (Unit: s)				
Matching method	Figs. 6(a) and 6(f)	Figs. 6(b) and 6(f)	Figs. 6(c) and 6(f)	Figs. 6(d) and 6(f)	Figs. 6(e) and 6(f)
SIFT	14.41	13.67	13.46	13.09	12.86
Iterative scale-invariant feature transform (ISIFT)	29.23	27.75	27.07	26.52	26.00
Proposed	19.05	18.22	18.07	17.74	17.13

From the above analysis and test results, a conclusion can be drawn that the proposed method is faster than the ISIFT method, and slightly more time-consuming than the original SIFT method. It is acceptable in many applications.

4 Conclusions and Future Work

In this paper, a novel illumination robust matching method is proposed. The multiresolution MSER detector and shape descriptor are used to compute scale and illumination invariant local region correspondences. Then an OACE method is put forward to estimate the overlapping areas and enhance images. The illumination enhancement is performed in the overlapping area to eliminate the interference of invalid areas. Finally, DoG detector and SIFT descriptor are combined to obtain matches. The proposed matching method exhibits substantial improvements compared to traditional methods for illumination change images. It should be noted that the number of regions extracted by the MSER detector strongly depends on the image content. However, it does not directly affect the final matching performance because one pair of correspondences is adequate in our OACE method. A possible future work is to improve the view invariance of the matching method based on OACE.

Acknowledgments

This work was supported by the National Natural Science Foundation Program of China (No. 61172174), the National Key Technology Research and Development Program of China (2012BAH35B02), the Science and Technology Program of Sichuan, China (No. 2015SZ0046) and the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing [No. (14) Key 01].

References

1. A. Tolga Can, O. Karalı, and T. Aytaç, "Detection and tracking of sea-surface targets in infrared and visual band videos using the bag-of-features technique with scale-invariant feature transform," *Appl. Opt.* **50**(33), 6302–6312 (2011).

2. Z. Song, S. Li, and T. F. George, "Remote sensing image registration approach based on a retrofitted SIFT algorithm and Lissajous-curve trajectories," *Opt. Express* **18**(2), 513–522 (2010).
3. B. E. Kratochvil et al., "Image-based 3D reconstruction using helical nanobelts for localized rotations," *J. Microsc. Oxford* **237**(2), 122–135 (2010).
4. A. Gruen, "Development and status of image matching in photogrammetry," *Photogramm. Rec.* **27**(137), 36–57 (2012).
5. C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, M. M. Matthews, Ed., pp. 147–152, Manchester, United Kingdom (1988).
6. K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vision* **60**(1), 63–86 (2004).
7. S. Smith and J. Brady, "SUSAN: a new approach to low-level image-processing," *Int. J. Comput. Vision* **23**(1), 45–78 (1997).
8. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
9. H. Bay et al., "Speeded-up robust features (SURF)," *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
10. J. Matas et al., "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comput.* **22**(10), 761–767 (2004).
11. T. Tuytelaars and L. V. Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vision* **59**(1), 61–85 (2004).
12. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 886–893, San Diego, California (2005).
13. A. Barla, F. Odone, and A. Verri, "Histogram intersection kernel for image classification," in *Proc. IEEE Int. Conf. on Image Processing*, pp. 513–516 (2003).
14. Y. Yu et al., "A novel algorithm for view and illumination invariant image matching," *IEEE Trans. Image Process.* **21**(1), 229–240 (2012).
15. M. Brown and D. G. Lowe, "Recognizing panoramas," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1218–1225 (2003).
16. K. Mikolajczyk et al., "A comparison of affine region detectors," *Int. J. Comput. Vision* **65**(1–2), 43–72 (2005).
17. P. E. Forssen and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1–8, Rio de Janeiro, Brazil (2007).
18. M. Chen et al., "Invariant matching method for different viewpoint angle images," *Appl. Opt.* **52**(1), 96–104 (2013).
19. J. Liu, Z. Shao, and Q. Cheng, "Color constancy enhancement under poor illumination," *Opt. Lett.* **36**(24), 4821–4823 (2011).
20. K. Mikolajczyk, "Affine Covariant Features," http://www.robots.ox.ac.uk/~vgg/research/affine/ (2007).
21. D. J. Jobson, Z. U. Rahman, and G. A. Woodell, "The statistics of visual representation," *Proc. SPIE* **4736**, 25–35 (2002).

Zhenfeng Shao is a professor of Wuhan University. He received his BS degree from Wuhan Technology University of Surveying and Mapping in 1998 and his MS and PhD degrees from Wuhan University in 2001 and 2004, respectively. He is the author of more than 50 journal papers. His current research interests include image processing, machine vision, etc.

Min Chen is a lecturer of Southwest Jiaotong University. He received his BS degree from Wuhan University in 2009 and his PhD degree from Wuhan University in 2014. His current research interests include feature detection, image matching, object recognition, etc.

Chong Liu is a PhD candidate of Wuhan University. He received his BS degree from Wuhan University in 2010. His current research interests include image processing and impervious surface extraction.