

## Research Article

# Robust Image Matching Algorithm Using SIFT on Multiple Layered Strategies

**Yong Chen,<sup>1</sup> Lei Shang,<sup>1</sup> and Eric Hu<sup>2</sup>**

<sup>1</sup> School of Energy Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

<sup>2</sup> School of Mechanical Engineering, The University of Adelaide, Adelaide 5005, Australia

Correspondence should be addressed to Yong Chen; [ychencd@uestc.edu.cn](mailto:ychencd@uestc.edu.cn)

Received 4 June 2013; Revised 18 October 2013; Accepted 1 November 2013

Academic Editor: Gradimir Milovanović

Copyright © 2013 Yong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As for the unsatisfactory accuracy caused by SIFT (scale-invariant feature transform) in complicated image matching, a novel matching method on multiple layered strategies is proposed in this paper. Firstly, the coarse data sets are filtered by Euclidean distance. Next, geometric feature consistency constraint is adopted to refine the corresponding feature points, discarding the points with uncoordinated slope values. Thirdly, scale and orientation clustering constraint method is proposed to precisely choose the matching points. The scale and orientation differences are employed as the elements of  $k$ -means clustering in the method. Thus, two sets of feature points and the refined data set are obtained. Finally,  $3 * \delta$  rule of the refined data set is used to search all the remaining points. Our multiple layered strategies make full use of feature constraint rules to improve the matching accuracy of SIFT algorithm. The proposed matching method is compared to the traditional SIFT descriptor in various tests. The experimental results show that the proposed method outperforms the traditional SIFT algorithm with respect to correction ratio and repeatability.

## 1. Introduction

Feature points extraction and registration is an important section in computer vision realms, such as target recognition, image stitching, 3D reconstructing, and target tracking. Recently, a new feature vector descriptor based on local invariant information [1–3] has been widely used in computer vision fields. The main idea of image registration is to extract lots of feature points and generate feature vectors with local information. Firstly, lots of feature points are extracted by different methods, for instance, Harris corner operator, SUSAN detection operator, and SIFT descriptor. Then, the feature descriptor regarding each candidate point is generated. Moreover, the constraint rules are used to check whether these feature descriptors are correctly matched pairs. Thus, corresponding pairs of feature points are obtained, achieving the goal of image registration.

In order to test the performances of feature descriptors widely used in computer vision fields, lots of experiments were developed by Mikolajczyk and Schmid [4]. In these experiments, sift method reveals more satisfactory performance and better robustness than other descriptors.

Lowe [2] presented SIFT descriptor which was invariant to several transformations including rotation changes, translation changes, and so on. Due to its multiple merits mentioned above, SIFT descriptor has been widely utilized in target tracking and recognition and other computer vision realms. In these fields mentioned above, firstly, we use SIFT descriptor to extract stable feature points and generate feature descriptors with local context of image. Then, we need to find corresponding pairs of feature points via various matching methods. It is evident that the corresponding points with high precision are the basis of further application. Therefore, it is not difficult to see that improving the matching performance is of importance.

In the past, many scholars have presented various types of improved matching algorithms. Wang et al. [5] proposed a new method based on slope value and distance. The distances and slope values of matching points are calculated. Then, the maximum of statistics is found. Moreover, certain value with regard to maximum is used to filter out these mismatching points. Though the method mentioned has achieved satisfactory results in eye ground image matching, the performance

is unsatisfactory in image matching with scale changes and rotation transformation. Li and Long [6] presented a new distance criterion based on comparing modulus. According to this method, the limit of the eigenvector is required before the normalization processing. Then accurate matching pairs between pictures can be found according to the difference between the modulus of the relative matching pair and that of the most similar one. However, the threshold value is difficult to choose. A new matching approach based on Euclidean distance and Procrustes iteration method was proposed by Liu et al. [7]. Firstly, these points are filtered by Euclidean distance. Then, the results are furthermore refined using Procrustes iteration method. Nevertheless, the process of method is complicated to some extent.

In order to improve the accuracy of image matching, a new feature descriptor generation method based on feature fusion was presented by Belongie et al. [8]. SIFT feature and shape context feature are extracted. Meanwhile, weight function is used to implement the process of feature fusion. However, the process of generating shape context descriptor is complicated and the weight function is difficult to choose. An improved SIFT matching was presented by Bastanlar et al. [9]. Firstly, preprocessing is performed by low-pass filtering and downsampling the high-resolution image. Then, the scale ratios of points are obtained. In the following step, the maximum of the statistics distribution function is found. Then, two threshold values regarding to the maximum of the statistics are used to filter out these error points. The effectiveness of the proposed method has been demonstrated in that paper. Mikolajczyk and Schmid [4] presented an extendable SIFT descriptor-GLOH (Gradient Location-Orientation Histogram) to enhance the distinctiveness and robustness of the new feature descriptor. The region with regard to keypoints is divided into several relative subregions and the original standard coordinate is projected into polar coordinate. Also, the histograms with respect to each subarea in 16 gradient directions are obtained, generating a 272-dimension feature descriptor. Then, PCA (principal component analysis) method is adopted to reduce feature descriptor dimensions. However, this method needs lots of salient images to generate projection matrix. The matching results of this approach depend on the choices of salient images. Retrofitted SIFT algorithm was developed on a new similarity measure function based on trajectories generated from Lissajous curves [10]. In their tests, the retrofitted SIFT descriptor improves the correct rate. Generally, the similarity measure function makes it possible to quantitatively analyze the temporary change of the same geographic position.

Inspired by all the approaches mentioned above, a new matching rule with multiple layered strategies is presented in this paper. Firstly, Euclidean distance between feature descriptors is used to discard the mismatching pairs. Furthermore, geometric feature consistency constraint is adopted to filter out corresponding pairs with abnormal slope values, refining these points from the first matching section. Then, the process of data clustering scale and orientation differences is done. All these feature points are classified into two parts after data clustering. After accomplishing the process of distance constraint between two center points of clustering

classes, a refined data set can be obtained. Finally,  $3 * \delta$  rule of the refined data set is used to search all these remaining discarded points.

## 2. Introduction of SIFT Descriptor

SIFT descriptor based on multiple scale spaces was presented by Lowe in 2004. The approach is divided into four sections, which will be shown as follows [2].

**2.1. Space Extreme Detection.** First of all, images between two adjacent octaves are downsampled by a factor of 2. Multikernel Gaussian functions are adopted to smooth images belonging to different octaves. Thus, the Gauss pyramid is established. The DoG space pyramid is generated by the difference of Gauss pyramid between two adjacent scales belonging to the same octave. Then the DoG space pyramid is established. Consider the following:

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

$$\text{DoG}(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma), \quad (2)$$

where  $\sigma$  represents the scale factor and  $I(x, y)$  is the input image. Also,  $*$  is the convolution operation in  $x$  and  $y$ . Meanwhile,  $G(x, y, \sigma)$  is the representative of Gaussian function with different scale space kernels.

In order to detect extreme points from the scale space, the pixel point is compared with its neighbor points in a  $3 * 3 * 3$  cube consisting of three adjacent intervals belonging to the same octave. This pixel point is chosen as the candidate point on condition that it is a local extreme with regard to the extreme detection cube.

**2.2. Keypoints Localization.** The next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. Low contrast points and unstable points with strong edge responses are discarded to improve the robustness of keypoints.

Firstly, Taylor expansion of the scale-space formula with regard to each candidate point is adopted. The specific steps are shown as follows. These candidate points with low contrast values will be discarded from the candidate points. Consider

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X, \quad (3)$$

where  $X = (x, y, \sigma)^T$  is the offset from this point. The accurate position of extreme keypoint  $\widehat{X}$  is found by calculating the derivative of function  $D$  regarding point  $X$ . Furthermore, the derivative of function  $D$  is set to be zero, which is shown in

$$\widehat{X} = - \left( \frac{\partial^2 D}{\partial X^2} \right)^{-1} \frac{\partial D}{\partial X}. \quad (4)$$

Substituting (4) into (3), we have

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

Candidate points with the absolute value of formula (5) less than some certain threshold (0.04 in this paper) should be abandoned.

In order to filter out the unstable points with strong responses along edges, Hessian matrix is used:

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

If the value of  $\text{Tr}^2(H)/\text{Det}(H)$  is less than certain threshold TH, it is reserved as one of the candidate points. Therefore, the unstable objects with strong responses along edges will be discarded.

**2.3. Feature Descriptor Generation.** In this stage, each keypoint is assigned a principal orientation by calculating the orientation histogram of vicinity with respect to the keypoint. This allows for the representation of each keypoint relative to this orientation, achieving invariance to image rotation. Then, the maximum value in the orientation histogram is obtained. In order to get a more precise orientation assignment, the peaks of the histogram are interpolated via the adjacent points. The original standard coordinate is rotated according to the main orientation, making the feature vector invariant to rotation changes. Each keypoint vector is established by calculating the gradient magnitude and orientation at each sample point in a region around the feature point region. Finally, a feature descriptor with 128 elements is obtained with respect to each feature point.

**2.4. Matching.** In [2], Euclidean distance criterion is selected as the rule to distinguish the matching extent between two feature vectors from the reference image and the unregistered one. The nearest neighbor point is defined as the keypoint with minimum Euclidean distance with regard to the invariant descriptor vector. In order to obtain a more accurate matching result, the ratio of the first closest point to the second closest point is used. If this ratio is lower than a certain value, it is a corresponding point with regard to certain feature point. The value of the first-closest to second-closest neighbor is advised to be in the range of [0.4, 0.8] [2]. In our experiments, the ratio is kept unchanged in the same group of tests.

### 3. Multiple Layered Strategies

In this section, a new method based on multiple layered strategies is presented. Firstly, Euclidean distance is used to discard the mismatching pairs. In addition, geometric feature consistency constraint is proposed to further filter out these points. In this section, the error points with abnormal slope values are discarded. A new method based on scale and orientation differences constraint is used to refine these matching points. Then the process of image matching based on multiple layered strategies is accomplished.

**3.1. Euclidean Distance Constraint.** After the process of keypoint searching and feature descriptor generation, Euclidean

distance rule is used to filter the original data. In this section, the Best-Bin-First (BBF) searching method is adopted, which returns the closest neighbor with high probability.

In [2], the range of threshold is set to be in the range of [0.4, 0.8]. Threshold (0.75) is used in this paper as the first matching judge criterion. Consider the following:

$$\frac{d_{\min}}{d_{\text{next\_min}}} < \text{threshold}, \quad (7)$$

where  $d = \sum_{i=1}^N (a_i - b_i)$ ,  $\mathbf{a} = (a_1, a_2, \dots, a_N)$ ,  $\mathbf{b} = (b_1, b_2, \dots, b_N)$ ,  $N$  is the dimension of feature the vector used in performance tests. In (7), vector  $\mathbf{a}$  is one of the feature descriptors extracted from the reference image and  $\mathbf{b}$  is one of the feature descriptors from the unregistered image.

**3.2. Geometric Feature Consistency Constraint.** Geometric feature consistency constraint means the relationship between the keypoints from the reference image and the unregistered image, including parallel feature, perpendicular attribute, and similar characteristic. The purpose of image matching is to figure out the transformation parameters between the reference image and the unregistered image, and then correct the coordinate of input image into the reference coordinate. In image matching section, the similar attribute still exists in certain region with regard to feature points, which is called geometric feature consistency constraint.

In our past experiments, corresponding feature points which have been matched by Euclidean distance still have lots of mismatching pairs.

It is evident that the slope value of correct matching points converges into a data set. Meanwhile, these feature points have evidently uncoordinated slope values in certain region, which belongs to distortion statistics data. Hence, a conclusion is that these distinct error points have uncoordinated slope values. Suppose that the correct points have similar slope values in the region between the reference image and the unregistered image, then the points which conform to the constraint rule can be discarded.

Feature points from reference image after Euclidean distance are stored in  $\text{image1\_set}$  as follows:

$$\text{image1\_set} = \{(a_1, b_1), (a_2, b_2), \dots, (a_m, b_m)\}. \quad (8)$$

Feature points from the unregistered image after Euclidean distance rule are stored in  $\text{image2\_set}$  as follows:

$$\text{image2\_set} = \{(a_1^*, b_1^*), (a_2^*, b_2^*), \dots, (a_m^*, b_m^*)\}, \quad (9)$$

where  $m$  is the number of feature points.

The process of geometric consistency constraint is shown as follows.

- (1) Firstly, the absolute value of slope with regard to corresponding feature points is obtained as  $k_i = \text{ABS}((a_i^* - a_i)/(b_i^* - b_i))$ . All these results will be stored in another data set:  $k\_set = \{k_1, k_2, \dots, k_m\}$ .
- (2) Then, the maximum value of all these slope values is obtained. Suppose  $k_{\text{maximum}}$  is the maximum value.



FIGURE 1: Images (size of image: 425 \* 340) used in this demonstration.

- (3) Threshold  $W$  with regard to the maximum value  $k_{\text{maximum}}$  is used to delete error points. All these points conforming the rule ( $k_i < W \cdot k_{\text{maximum}}$ ) will be discarded from the original set. The results further refined by geometric feature consistency constraint are stored in data base:  $k_{\text{sieved}} = \{k_1, k_2, \dots, k_t\}$ .

**3.3. Feature Points Constraint Based on Scale and Orientation Differences.** As we know sift feature point is a structure which contains coordinate location, orientation, scale, and descriptor information. Differences between two corresponding principal orientations of matching points indicate the rotation transformation relationship between the reference image and the unregistered image [11]. Supposing that all these matched points are correct, the differences should keep constant theoretically. Hence, this characteristic could be used to filter out the mismatching feature points from original data set, refining these corresponding pairs of feature points. However, in fact, affected by several factors, including the accuracy of transformation model between two images, the calculation errors, and the accuracy of matching feature points, the differences between two corresponding principal orientations should converge at a certain value. That is, the differences between the corresponding feature points from the reference image and the unregistered image should converge at a certain point. The distribution characteristic of orientation differences is similar to Gaussian distribution function and lots of statistics converge at certain center point.

Meanwhile, the differences between scales of corresponding points converge at a certain center point [12]. The distribution characteristic is similar to that of orientation differences. In order to demonstrate the characteristic, we do several experiments in advance. Here, only one of the test templates is specified in detail. In this test, Figure 1(a) is the reference image and Figure 1(b) exhibits the unregistered image.

Figure 2(a) shows the distribution characteristic of the scale differences between corresponding feature points. Figure 2(b) represents the distribution characteristic of orientation differences between matching pairs. We can get the

distribution map of scale and orientation in two-dimensional discretion map as shown in Figure 2(c).

After Euclidean distance constraint and geometric feature constraint, these corresponding points show large extent of convergence. In these figures, there are 367 points. Meanwhile, data sets of scale information converge at the point with average 0.3284 and variance 0.4383. Elements of orientation differences set converge at the point with average -0.2288 and variance 0.3965.

Evident error points have been discarded after Euclidean distance and geometric feature constraint. From the results above, it is shown that statistics reveal great convergence attribute. In Figure 2(c), abscissa is scale while ordinate stands for orientation.

Figure 3(a) is the histogram of scale difference information and Figure 3(b) shows the distribution map of orientation differences. From the figures of scale and orientation, it verifies our supposition in Section 3.3. It is evident that the distribution characteristic of these test statistics is similar to Gaussian distribution characteristic. That is, a large number of data points converge at the vicinity regarding certain point. Meanwhile, from the center point, distribution characteristic shows decay in the discretion map. All the results have been demonstrated in Matlab simulation workspace. Based on this idea, we implement the process of results refined.

The main idea of the presented approach in this paper is as follows. Firstly, Euclidean distance and geometric feature constraint are used to filter out mismatching pairs of feature points. Then, the orientation and scale factor differences of the set of pairs are calculated after the first step. Moreover, orientation and scale differences are used in data clustering. In our experiments,  $k$ -means algorithm is used to realize the data clustering. The result of data clustering shows two classes, that is, correct set- $A$  and error set- $B$ . Suppose that  $a$  is the clustering center point with regard to the correct set  $A$ . Also,  $b$  is another clustering center point of the error set  $B$ . Next, suppose that point  $p$  is one of the points in error set  $B$ . Then, the distance from point  $p$  to the clustering center point  $a$  is obtained. Consequently, the shortest distance  $T$  is obtained.

In the following step, distance  $T$  is used as the inner radius regarding correct set  $A$ . Suppose that point  $q$  is one of the

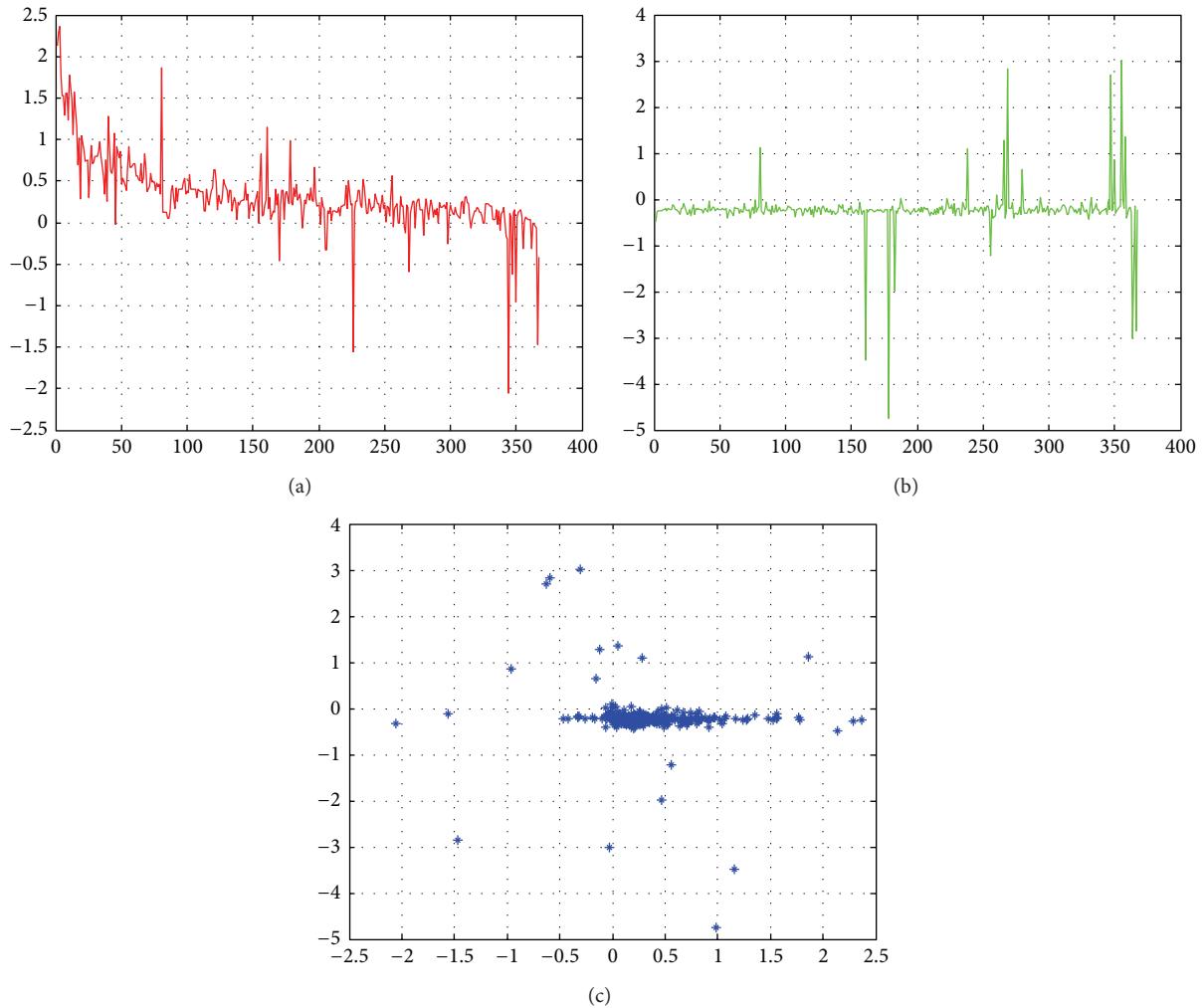


FIGURE 2: Distribution characteristics results of these points. (a) Distribution characteristic of scale differences. (b) Distribution characteristic of orientation differences. (c) Distribution map of scale and orientation.

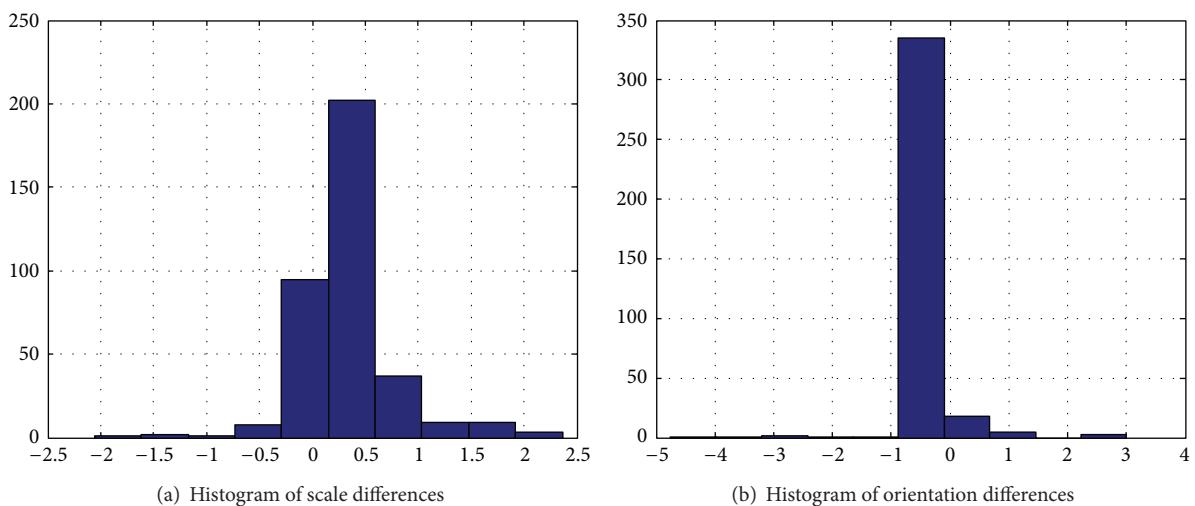


FIGURE 3: Histograms of scale differences and orientation differences.

correct set  $A$ . The distance from  $a$  to clustering center point  $a$  is obtained. Once the value is less than  $T$ , the relative point  $q$  is regarded as the element of refined set  $C$ . After the process, the other elements of set  $A$  are stored in set  $D$ .

In order to get enough correct points, we do the process of searching points which have been abandoned in previous step. Confidence interval of orientation differences with respect to refined set  $C$  is used to select these correct points from set  $D$ . After all these procedures, the process of image matching based on multiple layered strategies rule is accomplished. The confidence interval rule is shown as follows.

It is assumed that  $\mu$  is the average value of orientation differences with regard to set  $C$ ,  $\sigma$  is the standard deviation of orientation differences with regard to set  $C$ , and  $x_i$  is another point from other set  $D$ . If the absolute value of difference between  $x_i$  and  $\mu$  is less than  $3\sigma$ ,  $x_i$  is regarded as one of the elements regarding set  $C$ . Specific procedures are as follows.

(1) Firstly, Euclidean distance constraint is used to filter out mismatching points from sets of feature points as follows.

Feature points from the reference image are stored in set `image1_points`. Feature points from the unregistered image are stored in set `image2_points` as follows:

$$\begin{aligned} \text{image1\_points} &= \{A_1, A_2, \dots, A_m\}, \\ \text{image2\_points} &= \{B_1, B_2, \dots, B_m\}, \end{aligned} \quad (10)$$

where  $m$  is the number of feature points.

Each element of the sets contains the position of point, orientation, and scale factor, which is shown as follows:

$$A_i = \{x, y, \text{scl}, \text{ori}\}, \quad B_i = \{x, y, \text{scl}, \text{ori}\}. \quad (11)$$

(2) Geometric consistency constraint rule is used to filter out mismatching points from `image1_points` and `image2_points`. Then, the two sets of feature points are obtained by geometric consistency constraint. The points from reference image are stored in `set_one` =  $(A_1, A_2, \dots, A_n)$ , while those of unregistered image are stored in `set_two` =  $(B_1, B_2, \dots, B_n)$ , where  $n$  ( $n \leq m$ ) is the number of feature points after geometric consistency constraint.

(3) Two sets with regard to `image1_points` and `image2_points` are used to store scale and orientation information as follows:

$$\begin{aligned} \text{scale\_angle\_sieve\_one} &= (t_1, t_2, \dots, t_n), \\ \text{scale\_angle\_sieve\_two} &= (p_1, p_2, \dots, p_n), \end{aligned} \quad (12)$$

where  $t_i$  = {scale, orientation} and  $p_i$  = {scale, orientation}. Then, the scale differences and the orientation differences can be obtained in the next step:

$$\begin{aligned} \text{scl\_ori\_delta}_i \cdot \text{scale} &= t_i \cdot \text{scale} - p_i \cdot \text{scale}, \\ \text{scl\_ori\_delta}_i \cdot \text{orientation} &= t_i \cdot \text{orientation} - p_i \cdot \text{orientation}. \end{aligned} \quad (13)$$

Also, another set is used to store all these scale and orientation information:

$$\begin{aligned} \text{scl\_angle\_sieve\_delta} \\ = (\text{scl\_ori\_delta}_1, \text{scl\_ori\_delta}_2, \dots, \text{scl\_ori\_delta}_n). \end{aligned} \quad (14)$$

(4) In this step,  $k$ -means algorithm is adopted.

The data set `scl_angle_sieve_delta` is divided into two parts, namely, `clustering_correct` and `clustering_error`. Meanwhile, two clustering center points (`center_point_correct` and `center_point_error`) are obtained. Consider the following:

$$\begin{aligned} \text{clustering\_correct} \\ = (\text{scl\_ori\_delta}_1, \text{scl\_ori\_delta}_2, \dots, \text{scl\_ori\_delta}_{ta}), \\ \text{clustering\_error} \\ = (\text{scl\_ori\_delta}_1, \text{scl\_ori\_delta}_2, \dots, \text{scl\_ori\_delta}_{n-ta}), \end{aligned} \quad (15)$$

where  $ta$  is the number of elements with regard to `clustering_correct`.

(5) The distance from each point regarding `clustering_error` to the `center_point_correct` is calculated in this step. Thus, the shortest distance  $T$  is obtained. Then, we calculate the distance form each point regarding `clustering_correct` to the `center_point_correct`. If the distance is less than  $T$ , the corresponding element of `clustering_correct` is stored in `data_correct_scl_ori`. The corresponding positions of feature points are stored in two sets: `data_correct_one` and `data_correct_two`. Note that `data_correct_one` contains the positions of feature points from the reference image and `data_correct_two` contains the positions of feature points from the unregistered image.

The other elements of `clustering_correct` are stored in `data_sieve_scl_ori`. The corresponding positions of feature points are stored in two sets: `data_sieve_one` and `data_sieve_two`, where,

$$\begin{aligned} \text{data\_correct\_one} &= \{(x_1, y_1), (x_2, y_2), \dots, (x_{\text{num}}, y_{\text{num}})\}, \\ \text{data\_correct\_two} &= \{(x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_{\text{num}}^*, y_{\text{num}}^*)\}, \\ \text{data\_correct\_scl\_ori} \\ = \{\text{scl\_ori\_delta}_1, \text{scl\_ori\_delta}_2, \dots, \text{scl\_ori\_delta}_{\text{num}}\}, \\ \text{data\_sieve\_one} \\ = \{(x_1, y_1), (x_2, y_2), \dots, (x_{ta-\text{num}}, y_{ta-\text{num}})\}, \\ \text{data\_sieve\_two} \\ = \{(x_1^*, y_1^*), (x_2^*, y_2^*), \dots, (x_{ta-\text{num}}^*, y_{ta-\text{num}}^*)\}, \\ \text{data\_sieve\_scl\_ori} \\ = \{\text{scl\_ori\_delta}_1, \text{scl\_ori\_delta}_2, \dots, \text{scl\_ori\_delta}_{ta-\text{num}}\}. \end{aligned} \quad (16)$$

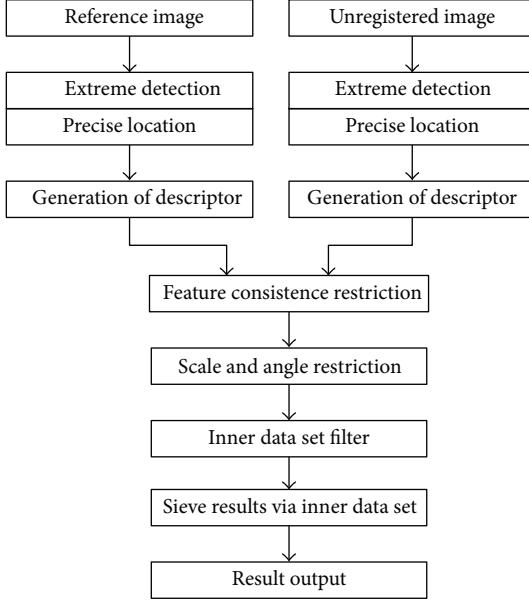


FIGURE 4: Flowchart of our proposed method.

(6) The last process is searching with the confidence interval of refined data set `data_correct_scl_ori`. All these elements from `data_sieve_scl_ori` meeting the confidence interval will be inserted into `data_correct_scl_ori`. All these corresponding feature points will be pushed into `data_correct_one` and `data_correct_two`.

Flowchart of the proposed method is presented in Figure 4.

### 3.4. Evaluation Criterion

**3.4.1. Repeatability.** To evaluate the performances of this presented approach, repeatability [13] is adopted as one of the criteria. Repeatability illuminates the stability of interest points detected via different keypoint extraction methods. With regard to these images, repeatability stands for the portion of keypoints both in the reference and unregistered images.

Suppose that  $X$  is a 3D point. Meanwhile,  $P_1$  and  $P_2$  are two relative projection matrices. It is assumed that feature point  $X_I$  is detected from the reference image  $I_I$ . This keypoint is repeated if the corresponding point  $X_i$  can be detected in the unregistered image  $I_i$ . The definition of repeatability ratio index is defined as the ratio of the number of keypoints repeated between relative images to the total number of feature points.

In the process of repeated point detection, the factor should be taken into account that the observed scene parts differ in the presence of changed imaging conditions. Consequently, these keypoints which exist in the common parts are adopted to calculate the repeatability measure. In order to find the common parts, homography matrix is used. This method is defined as follows:

$$\begin{aligned} \{\hat{x}_i\} &= \{H_{I-I}x_I, x_I \subseteq I_I\}, \\ \{\hat{x}_I\} &= \{H_{i-I}x_i, x_i \subseteq I_i\}. \end{aligned} \quad (17)$$

Furthermore, the uncertainty of detection should be considered in the repeatability measure. In fact, a repeated keypoint is not detected exactly at the position of point  $X_i$ . However, it exists in the certain neighborhood region with regard to  $X_i$  which is shown as follows:

$$R_i(\varepsilon) = \{(\hat{x}_I, \hat{x}_i) \mid \text{dist}(H_{1,i}x_I, x_i) < \varepsilon\}, \quad (18)$$

$$r_i(\varepsilon) = \frac{|R(\varepsilon)|}{\min(n_1, n_2)}, \quad (19)$$

where  $n_1$  is the number of keypoints extracted from the reference image  $I_I$ .  $n_2$  represents the number of keypoints detected from unregistered image  $I_i$ .  $R_i(\varepsilon)$  stands for the number of points defined in formula (18).

**3.4.2. Accuracy.** In order to evaluate the accuracy of our method, the transformation model approach is used to evaluate the correct ratio. In practical application, affine transformation formula has widely been employed as the judging model. Hence, in this paper, we use the following transformation model in this paper:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_5 \\ m_6 \end{bmatrix}, \quad (20)$$

where  $(x, y)$  is the position with regard to certain feature point from unregistered image and  $(u, v)$  is the position of corresponding point from reference image. The position of feature point  $p_i$  should be in neighbor region with regard to the point, defined as

$$R_i((u, v), (\hat{u}, \hat{v})) = \{\text{dist}((u, v), (\hat{u}, \hat{v})) < \varepsilon\}. \quad (21)$$

In (21),  $(\hat{u}, \hat{v})$  is the result obtained via formula (20). Points  $(u, v)$  and  $(x, y)$  are regarded as correct matching points if they meet the requirement of (21). To get these parameters, Least Square Method is adopted in this paper. The specific steps are as follows.

(1) Equation (20) is redefined as follows:

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}. \quad (22)$$

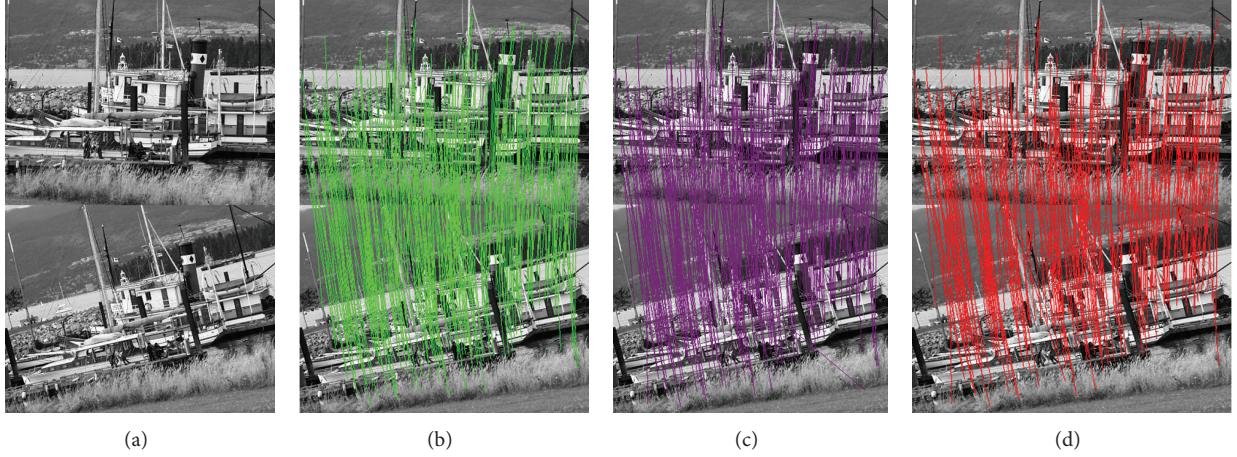


FIGURE 5: Images used in test one and results of the two methods. (a) Original unmatched images. (b) Results of proposed method. (c) Results of traditional SIFT algorithm. (d) Results of SR-SIFT.

TABLE 1: Statistics of test one.

	Reference image	Unregistered image	Results	Correct ratio	Repeatability
Traditional SIFT	1052	929	335	0.8865	0.3196
Improved approach	1052	929	363	1.0000	0.3907
SR-SIFT	1052	929	268	0.9701	0.2852

(2) Suppose that (22) is equation:  $G \cdot X = F$ . The affine transformation parameters could be obtained via step (3):

$$G = \begin{bmatrix} x_1 & y_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & y_1 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n & y_n & 0 & 0 & 1 & 0 \\ 0 & 0 & x_n & y_n & 0 & 1 \end{bmatrix}, \quad (23)$$

$$F = \begin{bmatrix} u_1 \\ v_1 \\ \cdot \\ u_n \\ v_n \end{bmatrix}, \quad X = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \\ m_6 \end{bmatrix}.$$

(3) The affine transformation parameters can be computed via the formula  $G^T G \cdot X = G^T \cdot F$ . After these processes, transformation parameters are obtained.

## 4. Experimental Results

Experimental setup: Intel(R) Dual CPU, 1.60 GHz, 1.0 GB memory.

To verify the effectiveness and feasibility of our improved method, OpenCV and VC++ are used to realize the presented algorithm. In our tests, we use Rob Hess's code of the traditional SIFT. In order to evaluate the performances of different approaches, the correct ratio and repeatability are used. In this paper, SR-SIFT in [12] will be introduced in our experiments. Then we will compare our performances with

SR-SIFT and SIFT. In these results, the upper one is reference image and the lower one is unregistered image. Besides, unregistered image is the result of reference image after transformation. In this paper, matching direction is from unregistered image to reference image. In the Euclidean distance section, threshold is 0.75. In Section 3.3, confidence interval is set to be in the range of  $[-3.3 * \text{standard deviation}, +3.3 * \text{standard deviation}]$ . All the tested pictures can be downloaded at <http://www.robots.ox.ac.uk/~vgg/research/affine>.

**4.1. Test One (Affine Transformation between Two Images Includes Rotation and Scale Changes).** Here, the image size is  $425 * 340$ . The test performances are presented in Figure 5. Figure 5(a) presents the two images unprocessed. Figure 5(b) is the performance of our proposed method and Figure 5(c) represents the result of original SIFT method. Figure 5(d) shows the matching results of SR-SIFT.

Table 1 shows that there are 1052 feature points in reference image. Feature points set extracted from unregistered image includes 929 candidate points. Number of points filtered by Euclidean distance in traditional SIFT algorithm is 335. Also, the number of points filtered by multiple layered strategies is 363. Moreover, the number of feature points from SR-SIFT is 268.

The correct ratio of the traditional algorithm is 0.8865, while the correct ratio of our improved method is 1.0000. Besides, the correct ratio of SR-SIFT is 0.9701. As for repeatability, the repeatability of traditional SIFT algorithm is 0.3196. Besides, the repeatability ratio of proposed approach is 0.3907. In addition, the repeatability of SR-SIFT is 0.2852.

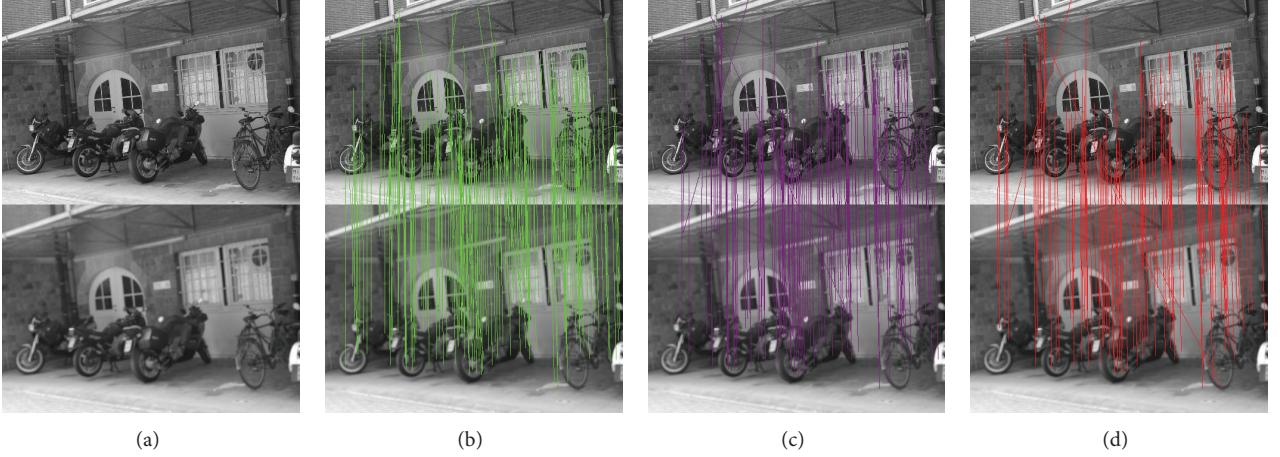


FIGURE 6: Images used in test one and results of the two methods. (a) Original unmatched images. (b) Results of proposed method. (c) Results of traditional SIFT method. (d) Results of SR-SIFT.

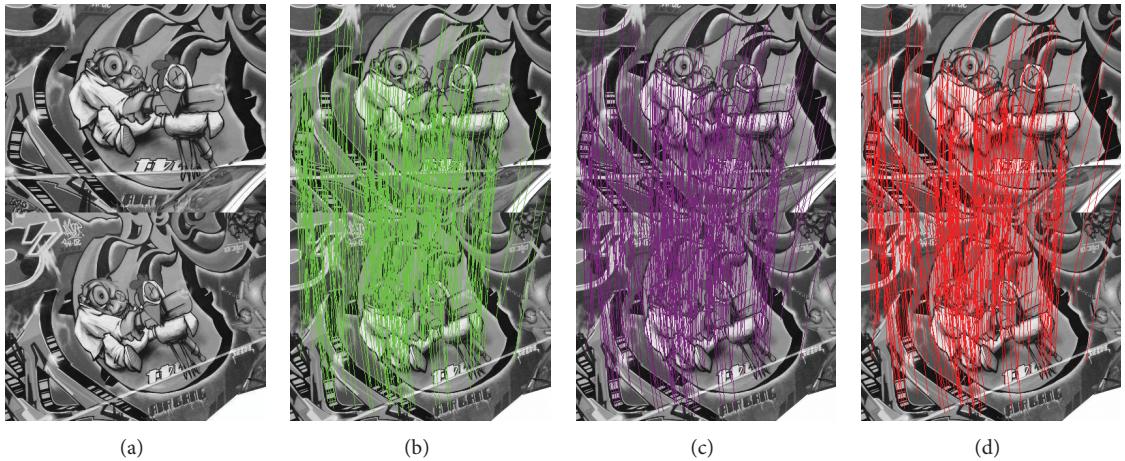


FIGURE 7: Images used in test one and results of the two methods. (a) Original unmatched images. (b) Results of proposed method. (c) Results of traditional SIFT algorithm. (d) Results of SR-SIFT.

By comparison with SIFT, the correct ratio of our method increases by 0.1135. Besides, the repeatability of our method increases by 7.11%.

The correct ratio of our method increases by 2.99% compared with SR-SIFT [12], while the repeatability increases by 10.55%.

**4.2. Test Two (Transformation between Two Images Includes Blurring Changes).** Image size is  $500 * 350$ . The test results are presented in Figure 6. Figure 6(a) presents the two images unprocessed, Figure 6(b) is the performance of proposed method, and Figure 6(c) represents the result of the original SIFT method. Figure 6(d) shows the matching results of SR-SIFT.

Table 2 shows that there are 785 feature points in reference image. Feature points set extracted from unregistered image includes 612 candidate points. The number of points filtered by Euclidean distance in traditional SIFT algorithm is 204. The number of points filtered by multiple layered strategies is

232. Moreover, the number of feature points from SR-SIFT is 134.

The correct ratio of traditional algorithm is 0.9607, while the correct ratio of our method is 0.9784. Besides, the correct ratio of SR-SIFT is 0.9776. As for repeatability, the repeatability of traditional SIFT algorithm is 0.3202. Comparatively, the repeatability ratio of proposed approach is 0.3709.

As for SR-SIFT, there is an equal number of the correct ratio between our method and SR-SIFT. Besides, the repeatability of our method increases by 22.06% compared with SR-SIFT.

**4.3. Test Three (Affine Transformation between Two Images Includes Viewpoint and Rotation Changes).** The image size is  $400 * 320$ . The test results are presented in Figure 7. Figure 7(a) presents two relative images, Figure 7(b) shows the performances of proposed method, and Figure 7(c) represents the result of the original SIFT method. Figure 7(d) shows the matching results of SR-SIFT.

TABLE 2: Statistics of test two.

	Reference image	Unregistered image	Results	Correct ratio	Repeatability
Traditional SIFT	785	612	204	0.9607	0.3202
Improved approach	785	612	232	0.9784	0.3709
SR-SIFT	785	612	134	0.9776	0.1503

TABLE 3: Statistics of test three.

	Reference image	Unregistered image	Results	Correct ratio	Repeatability
Traditional sift	923	1126	329	0.8510	0.3033
Improved approach	923	1126	347	0.9915	0.3759
SR-SIFT	923	1126	278	0.9424	0.2946

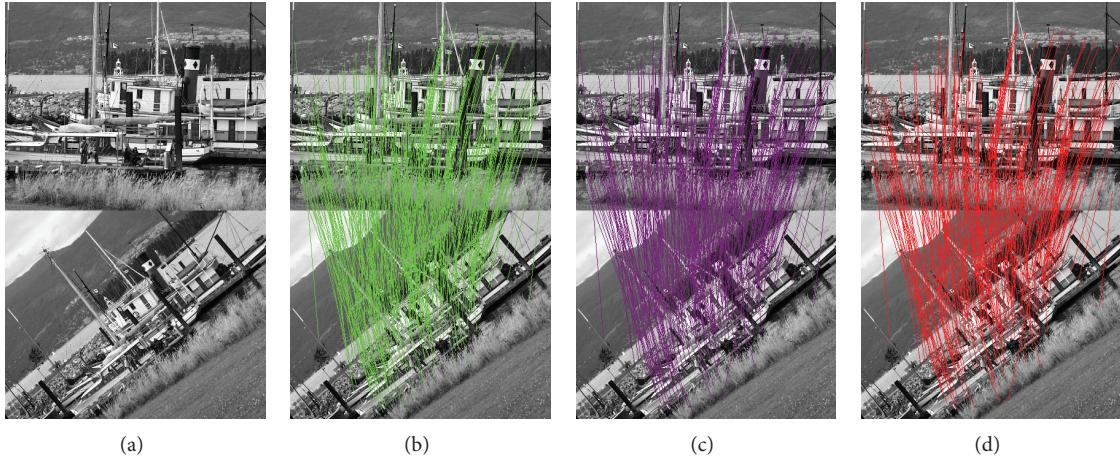


FIGURE 8: Images used in test one and results of the two methods. (a) Original unmatched images. (b) Results of proposed method. (c) Results of traditional SIFT algorithm. (d) Results of SR-SIFT.

Table 3 shows that there are 923 feature points in reference image. Feature points set extracted from unregistered image includes 1126 candidate points. The number of points filtered by Euclidean distance in traditional SIFT algorithm is 329, while the number of points filtered by multiple layered strategies is 347. In addition, the number of feature points from SR-SIFT is 278.

The correct ratio of the traditional algorithm is 0.8510. Comparatively, the correct ratio of our method is 0.9915. Besides, the correct ratio of SR-SIFT is 0.9424. As for repeatability, the repeatability of traditional SIFT algorithm is 0.3033 while the repeatability ratio of our proposed approach increases by 0.3759.

The correct ratio of our method increases by 4.91% when compared with SR-SIFT. Besides, when compared with SR-SIFT, the repeatability of our method increases by 8.13%.

**4.4. Test Four (Affine Transformation between Two Images Includes Scale and Rotation Changes).** Image size is 425 \* 340. The test performances are presented in Figure 8. Figure 8(a) presents two relative images, Figure 8(b) is the performance of proposed method, and Figure 8(c) represents the result of original SIFT method. Figure 8(d) shows the matching results of SR-SIFT.

Table 4 shows that the number of feature set of reference image is 1052, while that of unregistered image is 866. The number of points filtered by Euclidean distance in traditional algorithm is 266, while that of our method is 281. Moreover, the number of feature points from SR-SIFT is 209.

As for correct ratio, that of the traditional algorithm is 0.9060. Comparatively, the correct ratio of our approach in this paper is 0.9965. Besides, the correct ratio of SR-SIFT is 0.9523. As for the repeatability, the ratio of SIFT is 0.2667, while that of the proposed algorithm is 0.3233.

When compared with SR-SIFT, the correct ratio of our method increases by 4.42%. Besides, the repeatability of our method increases by 9.59%.

**4.5. Results Analysis.** Based on the results shown above, curves regarding correction ration and repeatability will be presented in Figure 9. Where Figure 9(a) is the correct ratio figure of the traditional algorithm and our method. Figure 9(b) shows the repeatability distribution result of the traditional SIFT, the SR-SIFT, and our improved method.

In the tests, we compare the improved approach based on multiple layered strategies with the traditional SIFT method and SR-SIFT. These transformations of test images include scale, rotation, and viewpoint changes.

TABLE 4: Statistics of test four.

	Reference image	Unregistered image	Results	Correct ratio	Repeatability
Traditional sift	1052	866	266	0.9060	0.2667
Improved approach	1052	866	281	0.9965	0.3233
SR-SIFT	1052	880	209	0.9523	0.2274

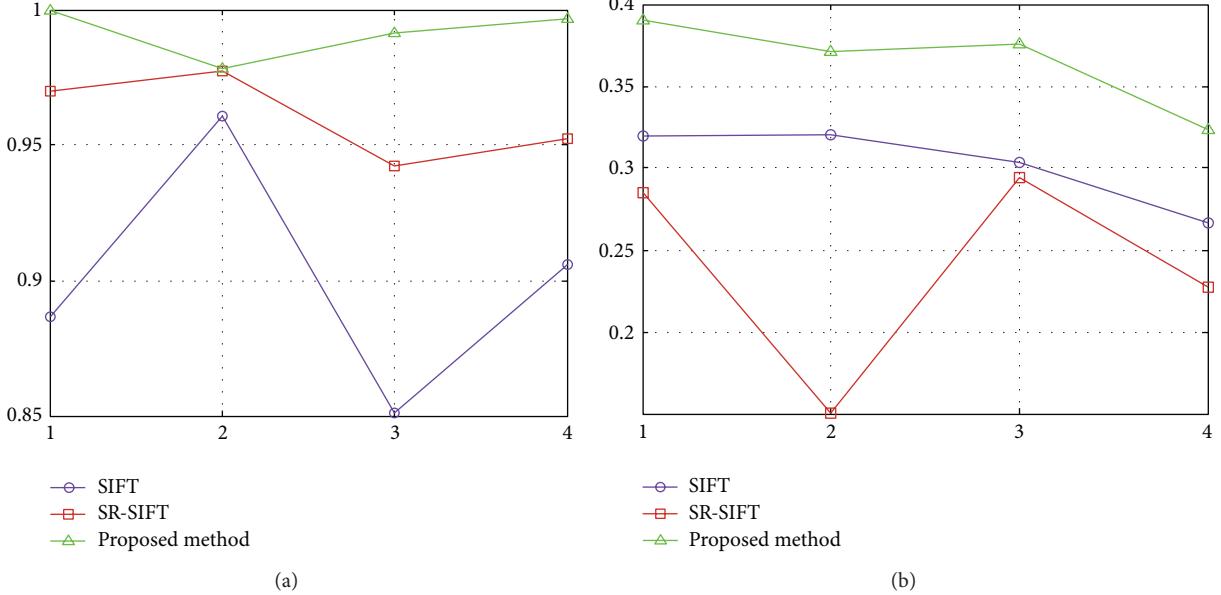


FIGURE 9: Performances of traditional SIFT algorithm and the proposed method. (a) Correct ratio results of the two approaches. (b) Repeatability ratio of SIFT and proposed method.

The SR-SIFT method proposed in [12] is based on scale restriction and shows better performances than the traditional SIFT under image registration experiments. However, the repeatability of SR-SIFT needs to be improved in practical applications, such as object detection based on feature points.

From the performances presented above, we can see that our proposed approach outperforms the traditional SIFT and SR-SIFT under these changes. These test results demonstrate the effectiveness and feasibility of our improved algorithm.

## 5. Conclusions

As for the problem of correct ratio and repeatability that is occurred in complicated image matching application using SIFT descriptor, a new approach based on multiple layered strategies is proposed in this paper. Firstly, the results are filtered by Euclidean distance. In addition, geometric feature consistency constraint is used to discard error pairs with abnormal slope. Then, a new method based on scale and orientation differences constraint is used to refine these matching points. The correction ratio and repeatability of the improved method outperform those of the traditional SIFT and SR-SIFT. Performances of experiments demonstrate the effectiveness and feasibility of the proposed algorithm. In future work, we will focus on improving the efficiency of the proposed approach.

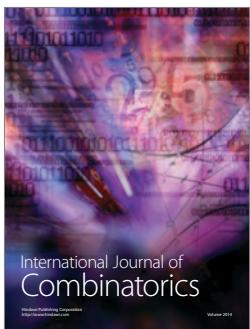
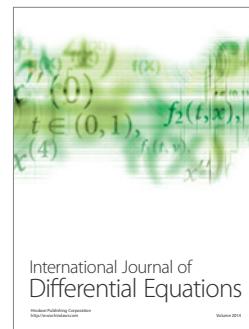
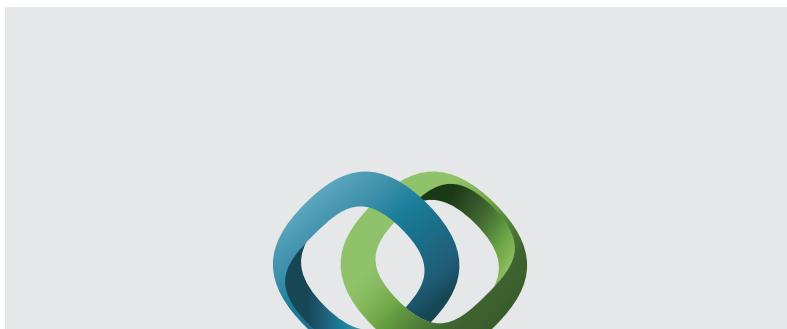
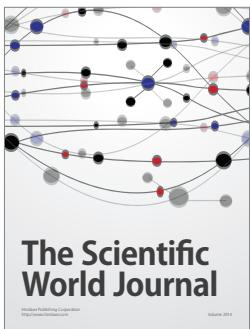
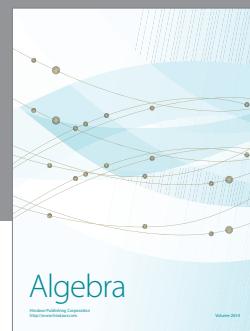
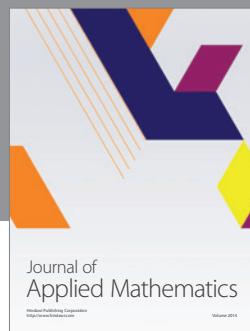
## Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant (61105030), The Fundamental Research Funds for the Central Universities of China (ZYGX2011J021), and the Scientific and Technical Supporting Programs of Sichuan Province (2013GZ0054).

## References

- [1] Y. Zhou, S. T. Zhou, Z. Y. Zhong, and H. G. Li, "A de-illumination scheme for face recognition based on fast decomposition and detail feature fusion," *Optics Express*, vol. 21, no. 9, pp. 11294–11308, 2013.
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [3] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: a SIFT descriptor with color invariant characteristics," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1978–1983, New York, NY, USA, June 2006.
- [4] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [5] Y. L. Wang, J. X. Shen, W. H. Liao, and L. Zhou, "Automatic fundus images mosaic based on SIFT feature," *Journal of Image and Graphics*, vol. 16, no. 4, pp. 2747–2751, 2010.

- [6] L. J. Li and X. Long, "LSIFT—an improved SIFT algorithm with a new matching method," in *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM '10)*, pp. V4-643–V4-646, Taiyuan, China, October 2010.
- [7] J. Liu, G. H. Zhang, and L. L. Huang, "Image registration approach based on improved SIFT," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 36, no. 9, pp. 1121–1124, 2010.
- [8] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
- [9] Y. Bastanlar, A. Temizel, and Y. Yardimci, "Improved SIFT matching for image pairs with scale difference," *Electronics Letters*, vol. 46, no. 5, pp. 346–348, 2010.
- [10] Z. L. Song, S. Li, and T. F. George, "Remote sensing image registration approach based on a retrofitted sift algorithm and lissajous-curve trajectories," *Optics Express*, vol. 18, no. 2, pp. 513–522, 2010.
- [11] Q. L. Li, G. Y. Wang, J. G. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 287–291, 2009.
- [12] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on SIFT," *Electronics Letters*, vol. 44, no. 2, pp. 107–108, 2008.
- [13] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.



Submit your manuscripts at  
<http://www.hindawi.com>

