

Research Article

SIFT Saliency Analysis for Matching Repetitive Structures

Yan Yang,^{1,2} Xiaopeng Hu,¹ Lijuan Xu,¹ and Fan Wang¹

¹School of Computer Science and Technology, Dalian University of Technology, Dalian, China

²School of Computer and Information Technology, Liaoning Normal University, Dalian, China

Correspondence should be addressed to Yan Yang; xuemeng62038216@163.com

Received 5 June 2017; Accepted 19 November 2017; Published 12 December 2017

Academic Editor: Oscar Reinoso

Copyright © 2017 Yan Yang 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.

The ambiguity resulting from repetitive structures in a scene presents a major challenge for image matching. This paper proposes a matching method based on SIFT feature saliency analysis to achieve robust feature matching between images with repetitive structures. The feature saliency within the reference image is estimated by analyzing feature stability and dissimilarity via Monte-Carlo simulation. In the proposed method, feature matching is performed only within the region of interest to reduce the ambiguity caused by repetitive structures. The experimental results demonstrate the efficiency and robustness of the proposed method, especially in the presence of repetitive structures.

1. Introduction

Image matching is the core operation in many computer vision tasks [1–4]. Various approaches focus on improving the appearance distinctiveness and repeatability of the features and design many distinctive descriptors to find more reliable matching pairs [5–8]. The locally defined descriptors can provide very accurate information that is invariant to scale changes and scene clutter [9]. Moreover, current research has indicated that existing feature matching methods such as Ratio-Match [7], Self-Match [10], and Mirror-Match criteria [11] are so restrictive that they fail to match repeated features [1]. All these methods make a comparison between the best and second best matches to obtain reliable matches. The only difference is the way to find the best matches and the second matches. For the repeated features or similar features from the repetitive structures of a scene, these methods can hardly find unique matches because the best matches are nearly the same as the second matches.

Identifying enough reliable matches from the repetitive structures in a scene is difficult due to the ambiguity resulting from such repetition [12–15]. To reduce the matching ambiguity and find more reliable correspondences, Xiaojie Guo and Xiaochun Cao propose a 2D-2D image matching method based on the triangle constraint [15]. And such method is successfully applied in visual attention regions for

the video copy detection [14]. They apply the Ratio-Match criteria [7] in a bimatching procedure to select initial matches or the seed vertexes and construct triangle constraints to reduce the search space for matching globally similar feature. However, the bimatching method is sometimes not able to identify any unique matches due to the low robustness of Ratio-Match criteria for the cases such as urban-like scenes where buildings have repetitive structures [1]. The ambiguity resulting from repetitive structures which are common in visual scenes remains a challenge for image matching [13].

This paper proposes a robust matching strategy which relies on the saliency analysis of image features. It is based on the computation of the distances between features and the estimation of their standard deviation via a Monte-Carlo approach. To explore more information from the repetitive structures [16], this research presents a robust matching strategy where the correspondence of salient features is found and used to determine the regions of interest for matching the repeated ones and, in the meantime, similar features outside the region are excluded, leading to the reduction of the matching ambiguities. The experimental results demonstrate the efficiency and robustness of the proposed method, especially in the presence of large portion of repetitive structures.

This paper is organized as follows. After reviewing the related work in this section, we introduce a region-based image matching strategy based on calculating the feature

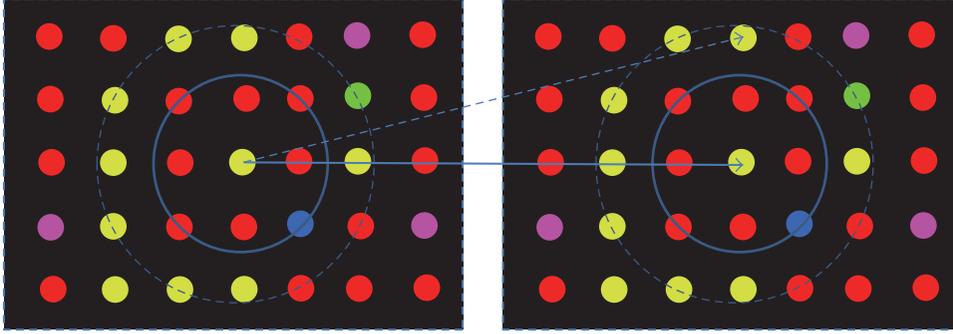


FIGURE 1: Illustration of feature saliency and matching ambiguity. Solid and dashed circle indicate different ROIs which affect the saliency of the yellow spot in the center. Solid and dashed line illustrate different candidate correspondence within different searching space in the right image for the yellow spot in the center of the left image.

saliency. Section 3 demonstrates the experimental results. Conclusions are discussed in Section 4.

2. Robust Image Matching Strategy Based on SIFT Saliency

2.1. Saliency Definition for the SIFT Feature. The local saliency of an object is usually defined as the contrast between the object and its neighborhood as illustrated by biologically plausible models such as the center-surround mechanism [17, 18]. As illustrated in Figure 1, the spots with different color are used to represent different SIFT features. The distinctiveness of the same yellow spot is different within the solid and dashed regions. The solid or the dashed regions illustrate different regions of interest (ROIs). The yellow spot is more salient in the solid circular region than in the dashed region because there are no other yellow spots in the solid circular region. As a result, the yellow spot can be easily detected in the solid circular region by using color feature only. In the dashed region, however, this method cannot simply be applied since yellow spots repeat more than once within the region.

Figure 1 illustrates that the difficulty of feature matching in a region with repetitive structures is strongly related to the repetitiveness of the feature inside the region. As a result, this paper defines a feature saliency which is determined by the difference of features in the whole image region.

Given a feature set F extracted from the image I , the saliency of a local feature $\mathbf{f}_k \in F$ is defined as

$$S(\mathbf{f}_k) = \min_{\mathbf{f}_j \in F, \mathbf{f}_j \neq \mathbf{f}_k} \left(\frac{\|\mathbf{f}_k - \mathbf{f}_j\|}{\sigma_{\mathbf{f}_k}} \right), \quad (1)$$

where the saliency S of \mathbf{f}_k is defined as the minimum difference between the features \mathbf{f}_k and \mathbf{f}_j at the locations k and j of image I . The feature difference is obtained by calculating the ratio between the Euclidean distance of two features \mathbf{f}_k and \mathbf{f}_j and the standard deviation $\sigma_{\mathbf{f}_k}$ of \mathbf{f}_k . The computation of $\sigma_{\mathbf{f}_k}$ will be explained in Section 2.2.

Equation (1) can be applied to estimate feature saliency in a ROI [14]. As demonstrated in Figure 1, such estimation is crucial for tackling the ambiguity resulting from repetitive

structures. In this paper, we focus on performing the saliency analysis in the whole image I . According to (1), if a feature is distinctive in an image, it is more likely to be salient in the region of the whole image. In Figure 1, for instance, the blue and green spots are more salient than the yellow and red spots over the whole image since there are no other spots with similar color in the same ROI. Similarly, the central yellow spot within the solid circle is salient when the solid circle can be determined as the ROI.

2.2. Estimation of Feature Saliency via Monte-Carlo Simulation. The feature saliency defined in (1) can be estimated via a Monte-Carlo method where the degree of freedom is application-dependent. The independent variables include blur, illumination change, JPEG compression, parameters of the camera, and its pose. Given a reference image, the Monte-Carlo method for feature saliency computation is straightforward, including the following steps:

- (1) Define a set of independent image variables and their probability distributions according to the application settings.
- (2) Generate a set of simulated images randomly from the probability distributions over the independent variables.
- (3) Extract local image features from the simulated images.
- (4) Conduct a repeatability test to select robust features.
- (5) Calculate standard deviation $\sigma_{\mathbf{f}_k}$ for each local feature over the simulated images.
- (6) Calculate saliency for each local feature in the reference image by applying (1).

For example, the reference image R shown on the bottom of Figure 2 is taken from the Bikes scene of INRIA dataset [19] where it is matched against a set of test images obtained by varying blur. In the Monte-Carlo process, the Gaussian blur algorithm is used for blurring the reference image to generate N simulated images S_1, S_2, \dots, S_N shown on the left column in Figure 2. For each image, the SIFT [7] features

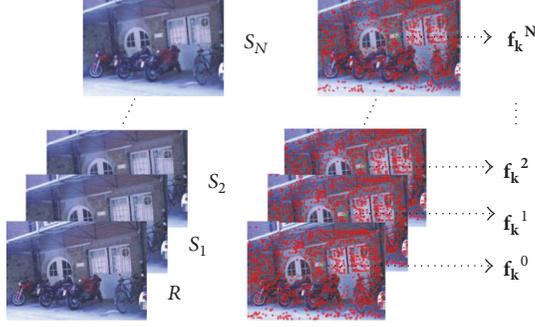


FIGURE 2: Illustration of standard deviation computation. The left column illustrates a group of Bikes-scene images with different level of blur. The right column illustrates the spots or feature key points extracted from these images.

are extracted and shown on the right column. As a result, for each SIFT feature \mathbf{f}_k^0 in the reference image, there ideally exists a corresponding SIFT feature \mathbf{f}_k^i on the i th simulated images S_k . Due to the invariance of SIFT feature, we can assume a normal distribution of the feature \mathbf{f}_k^0 . Thus, the standard deviation of the feature \mathbf{f}_k^0 in the feature space can be calculated as follows:

$$\sigma_{\mathbf{f}_k^0} = \text{Std}(\mathbf{f}_k^0) = \sqrt{\frac{\sum_{i=1}^N \|\mathbf{f}_k^i - \mathbf{f}_k^0\|^2}{N-1}}. \quad (2)$$

In practice, it is necessary to remove the feature outliers to estimate $\sigma_{\mathbf{f}_k^0}$ in (2). Techniques such as those of robust statistics can be applied to obtain a reliable estimation. To this end, we perform a repetitive test of SIFT features to remove the outliers. Features with a repeatability rate of not less than a threshold γ are regarded as stable. Only stable features can be applied in (2). In this paper, we set $\gamma = 0.6$ to achieve considerable number of feature key points for the matching process.

2.3. Saliency-Based Feature Matching. Given two feature sets F_0 and F_1 extracted from the reference image I_0 and another image I_1 respectively, the proposed saliency-based feature matching method includes the following three steps.

2.3.1. Salient Feature Selection. The salient features are selected by the aforementioned repeatability test and by thresholding the feature saliency to ensure the stability and distinctiveness of the features. The salient feature set $F_s \subseteq F_0$ in the whole image I_0 can be determined by

$$F_s = \{\mathbf{f}_s \mid \mathbf{f}_s \in F_0, S(\mathbf{f}_s) > \eta\}, \quad (3)$$

where $\mathbf{f}_s \in F_s$ is a salient feature and η is the threshold for the feature selection. As indicated by statistical hypothesis testing theories, the larger the value η , the more salient and reliable the features in F_s for matching.

2.3.2. Salient Feature Matching. Let $\mathbf{f}_s \in F_0$ be a salient feature in the whole image I_0 . Its matching feature $\mathbf{f}_s^{(1)} \in F_1$ is identified by using the following criteria:

$$\begin{aligned} \min_{\mathbf{f}_s^{(1)} \in F_1} \left(\frac{\|\mathbf{f}_s - \mathbf{f}_s^{(1)}\|}{\sigma_{\mathbf{f}_s}} \right) &= \frac{\min_{\mathbf{f}_s^{(1)} \in F_1} (\|\mathbf{f}_s - \mathbf{f}_s^{(1)}\|)}{\sigma_{\mathbf{f}_s}} \leq \lambda \\ \text{s.t. } \frac{\min_{\mathbf{f}_s^{(2)} \in F_1, \mathbf{f}_s^{(2)} \neq \mathbf{f}_s^{(1)}} (\|\mathbf{f}_s - \mathbf{f}_s^{(2)}\|)}{\sigma_{\mathbf{f}_s}} &> \alpha\lambda, \quad \alpha \geq 1, \end{aligned} \quad (4)$$

where $\mathbf{f}_s^{(2)} \in F_1$ is the 2nd nearest neighbor to the salient feature \mathbf{f}_s and $\lambda, \alpha\lambda$ are the matching thresholds. Both the distances between the best and the second best matches are compared with the thresholds λ and $\alpha\lambda$, respectively. In the above equations, the standard deviation $\sigma_{\mathbf{f}_s}$ is obtained by Monte-Carlo simulation described in Section 2.2. To find the unique matches of salient features, we impose a constraint on the above matching criteria. Such a constraint is used to verify whether there exist other similar features in the observation image due to the occlusion or other unpredictable changes. The larger the value of α , the greater the matching reliability that can be achieved. Moreover, if a given salient feature can be matched with only one candidate feature, we consider the match is unique and reliable since there is no ambiguity caused by the repetitive features from both the reference image and the observed image. To guarantee the matching quality of the salient feature, the relationship between η and λ should satisfy

$$\eta \geq \lambda. \quad (5)$$

Existing methods [7, 10, 11] share some similarity with the proposed saliency-based feature approach on finding unique matches. In those methods, the criteria for identifying the matching feature $\mathbf{f}^{(1)} \in F_1$ for feature $\mathbf{f} \in F_0$ can be formulated as follows:

$$\frac{\min_{\mathbf{f} \in F_0, \mathbf{f}^{(1)} \in F_1} (\|\mathbf{f} - \mathbf{f}^{(1)}\|)}{\min_{\mathbf{f} \in F_0, \mathbf{f}^{(2)} \in \{F_0, F_1\}, \mathbf{f}^{(2)} \neq \mathbf{f}^{(1)}} (\|\mathbf{f} - \mathbf{f}^{(2)}\|)} \leq \epsilon. \quad (6)$$

If the ratio of closest distance to second closest distance is greater than ϵ , the match is rejected as a false one with the ambiguity.

According to (4) and (6), we can see that both our method and the existing methods [7, 10, 11] have a common point to perform the neighbor ranking to find the best matches and the second best matches, but the proposed saliency-based matching method is faster than the other approaches due to the feature selection. If less stable features are selected as salient features, there are fewer L2 distances between $\mathbf{f}_s \in F_0$ and $\mathbf{f}_s^{(1)} \in F_1$ to compute. Furthermore, the proposed matching criterion is more theoretically sound since it is based on the principles of statistical hypothesis tests by assuming a normal distribution of SIFT features due to their invariance properties. Thus, to achieve a 95.54% prediction interval, we need to set $\lambda = 2$ approximately.

2.3.3. Region-Based Feature Matching. The features generated from repetitive visual structures are usually similar, resulting in a high mismatch rate if the spatial locations of these features are not taken into consideration. Each repetitive feature has its local saliency region as illustrated in Figure 2. If the ROI can be determined and there are no more similar features in the small region, the repetitive feature becomes salient and the matching robustness is expected to be improved.

In visual search, the human visual system always directs its attention from salient object to less salient ones [20]. Inspired by such an idea, we propose identifying the robust matches of salient features first, followed by a process to apply these robust matches to determine ROIs of other less salient features, especially repetitive features. We assume that the reliable matches of repetitive structures can be found if the saliency regions of features can be determined to find unique matches. In this paper, we use the robust matches of the salient features as initial matches or seeds in the triangle constraint method (T-CM) [15] to determine a small ROI for each repetitive feature.

The quality of the seed features will greatly affect the final matching results since the region-based matching process starts with using these seeds to determine ROIs for other features. The quality includes not only the robustness of the seeds but also their number and spatial distribution. In the seeds selection, we can adjust the matching threshold for different matching methods to obtain different number of the seeds. Normally, the more the number of the seeds, the more concentrated the spatial distribution of the seeds and the smaller the area of the ROI or triangle region. According to Section 2.1, a small ROI is expected for each repetitive feature to find reliable correspondence because the feature has more chance to be different from the fewer features in the ROI. To obtain a large number of the seeds, it requires relaxing the matching threshold at the sacrifice of the matching robustness. It is still an open issue to balance the matching robustness and the number of the seeds. In most cases, the matching threshold is set by experience [7, 10, 11].

In this paper, we only focus on evaluating the quality of seed distribution with a small number to obtain a high matching robustness. When the number of seeds is small, the area coverage of all the seed features plays an important role for finding more candidate matches. We define the area coverage D_C as follows:

$$D_C = \frac{A_{cv}(F_m)}{A_{cv}(F_0)}, \quad (7)$$

where $A_{cv}(\cdot)$ is the area bounded by the convex hull of a feature set. F_m and F_0 are the sets of the seed features and all the extracted features, respectively.

Figure 3 illustrates the area coverage of the robust seed features with a small number. According to (7), the value of D_C is between 0 and 1. The larger the value of D_C , the better the area coverage of the seed features. Figure 3(a) illustrates the seed features with an ideal distribution quality. The matching ambiguity of each repetitive feature is expected to be eliminated when D_C is equal to 1 and each repetitive

feature is contained by only one triangle region constructed by the seeds. Figure 3(b) illustrates the seed features with the area coverage smaller than 1.0. There are generally two problems demonstrated in the case: (1) multiple similar features locate in the same ROI, such as four repetitive features in the dotted triangle region. The ambiguity is not resolved. (2) Not all the repetitive features are covered by the convex hull of the seed features. Thus, the ROIs for these uncovered features cannot be determined and this decreases the chance for finding more candidate matches. Moreover, Figure 3(c) illustrates the seed features with the same area coverage but a different spatial distribution is shown in Figure 3(b). Compared with Figure 3(b), the seed features in Figure 3(c) are distributed more evenly, indicating that the spatial distribution of the seed features is closely related to ROIs for each repetitive feature. To quantify the difference between the spatial distributions shown in Figures 3(b) and 3(c), we adopt the following descriptor D_A as proposed in [21].

$$D_A = \sqrt{\sum_{i=1}^{sn} \left(\frac{A_i}{\bar{A}} - 1 \right)^2}, \quad (8)$$

$$\bar{A} = \frac{\sum_{i=1}^{sn} A_i}{sn},$$

where sn is the number of the ROIs constructed by the seed features and A_i is the area of each ROI.

Considering both the area coverage and the spatial distribution, we propose an overall measure D to describe the quality of seed distribution.

$$D = \frac{D_A}{D_C}. \quad (9)$$

The smaller the value of D , the better the quality of seed distribution.

3. Experimental Results

3.1. Datasets and Procedures. Few resources exist for the purpose of testing our matching method, since most image pairs consist of only one reference image but a large number of training images for computing the proposed feature saliency and selecting the robust features. Thus, in evaluation, we generate the training set for each image pair by scale and rotation transforming or adding some noise to the given reference image from the INRIA dataset [19]. To generate a large number of test sets, we choose a scheme similar to the existing literature [11] to generate a large number of crop pairs with repetitive structures.

The INRIA dataset [19] contains 48 images of 8 scenes, taken from different imaging conditions including varying light, blur, viewpoint, and JPEG compression. The dataset provides homography functions as the ground truth for feature matching. Figure 4 shows a matching pair of the wall scene where repetitive structures are extensively present.

In the experiment, the matching threshold is set to 2 for the proposed method to achieve an about 95.54% prediction

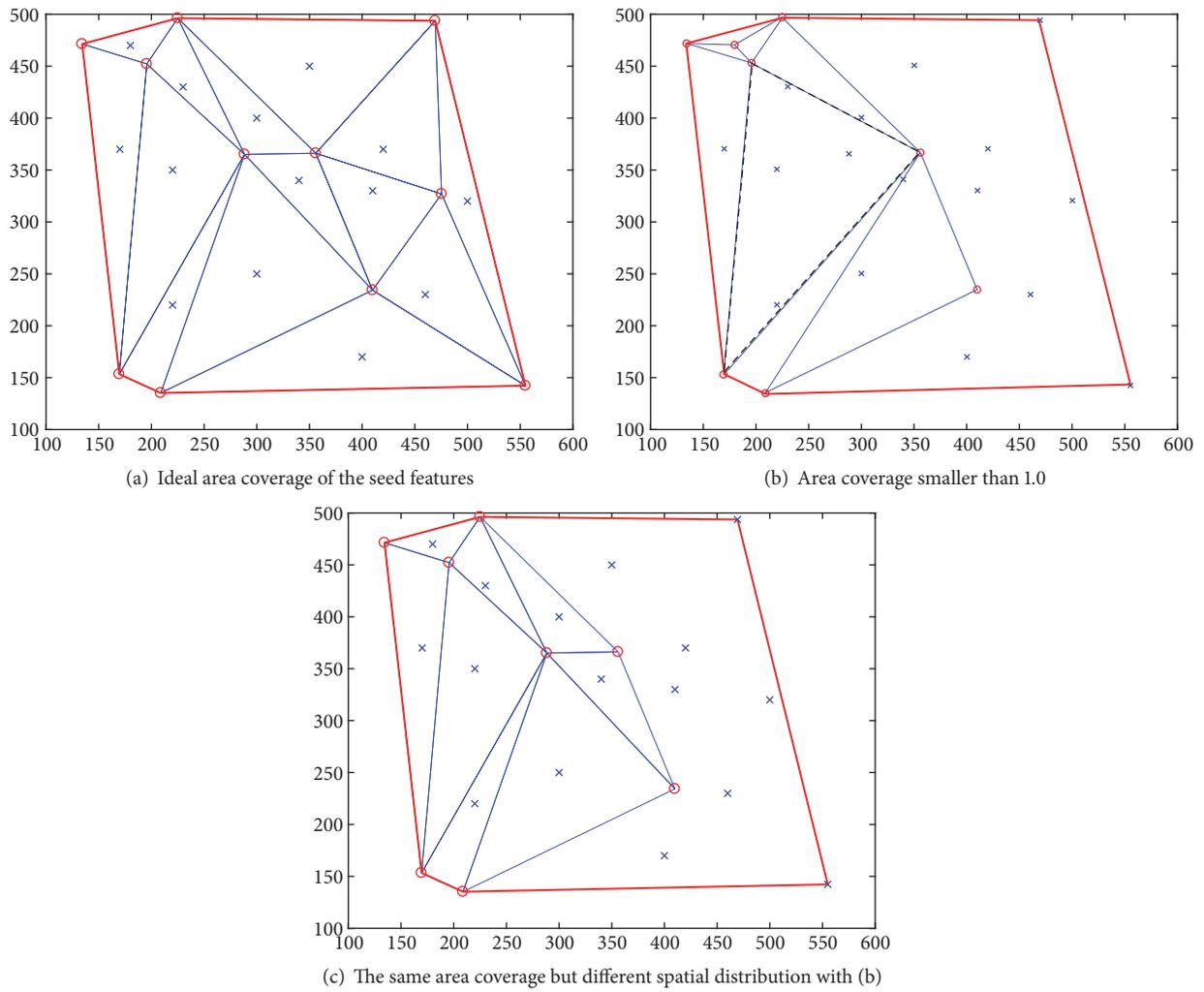


FIGURE 3: The area coverage and the spatial distribution of seed features. Circle and \times marks indicate seed features and repetitive features, respectively. The areas bounded by bold and thin solid lines represent $A_{cv}(F_0)$ and $A_{cv}(F_m)$, respectively. The triangle region with dotted lines illustrates a ROI for multiple features.

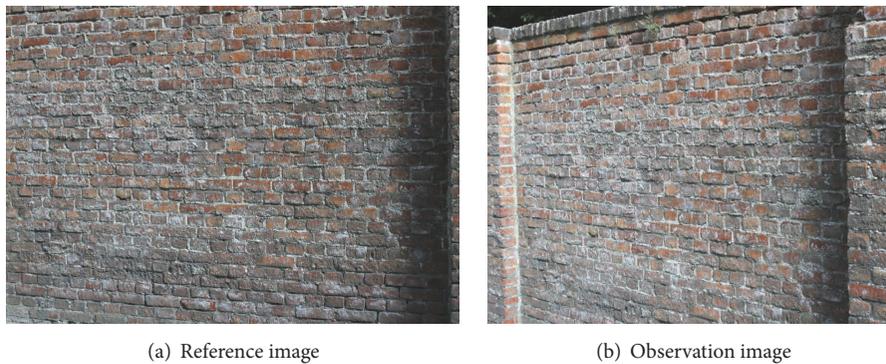


FIGURE 4: One matching pair of the wall scene with repetitive structures.

Input. An image group I of one scene where $I = I_1 \cup I^{(1)}$ and $I^{(1)} = \{I_k \mid k = 2, \dots, N\}$. The reference image is I_1 .

- (1) Select one image $I_k \in I^{(1)}$ as the observation image;
- (2) Compute the SIFT feature saliency based on the image training set $I^{(2)} = \{I_m \mid m = 1, \dots, N, m \neq k\}$ according to Section 2;
- (3) Perform the feature matching between I_1 and I_k ;
- (4) If the reference image I_1 is matched with all the images from $I^{(1)}$, go to (5); else go to (1);
- (5) Compute the average matching results according to (11)–(13).

Output. The average matching results.

ALGORITHM 1: The process of LOOCV test.

interval as discussed in Section 2.3.2. The threshold is set to 0.8 for other methods to eliminate 90% of the false matches at the sacrifice of less than 5% of the correct matches as suggested by [7].

Both the matching precision and recall rate are used to express the performance of SIFT feature matching.

$$\text{Precision} = \frac{N_{\text{correct}}}{N_{\text{correct}} + N_{\text{incorrect}}}, \quad (10)$$

$$\text{Recall} = \frac{N_{\text{correct}}}{N'_{\text{correct}}}, \quad (11)$$

where N_{correct} , N'_{correct} , and $N_{\text{incorrect}}$ are the number of correct matches, ground-truth matches, and incorrect matches, respectively. The ground-truth matches can be obtained by using the homography function H provided by the database. If the positions of two matched features are $\{\mathbf{x}_i, \mathbf{x}'_i\}$, they should satisfy

$$\|H\mathbf{x}_i - \mathbf{x}'_i\| = |\rho_i| \leq d_{\text{proj}}. \quad (12)$$

The value of d_{proj} is the distance threshold which is set to 4 pixels.

In this paper, S-Match is used to denote the proposed saliency-based feature matching method. It mainly consists of three steps including feature saliency computation, seeds selection, and region-based image matching. In the absence of imaging conditions, such as parameters of the camera and its pose, it is difficult to generate simulated images with the same conditions to compute the saliency of SIFT features. To avoid estimating the parameters of image conditions, we apply both the simulated images and the images provided by the INRIA dataset [19] for saliency computation in the Monte-Carlo method shown in Section 2.2. Thus, the Leave-One-Out Cross-Validation (LOOCV) [22] test shown in Algorithm 1 is designed. To reduce variability, all possible rounds of cross-validation are performed for each scene in INRIA dataset [19] and the validation results are averaged over the rounds.

3.2. Feature Matching with Repetitive Structures. To show the robustness of the proposed S-Match method in presence of repetitive structures, we compare our approach with the state-of-the-art methods including Ratio-Match [7], Self-Match [10], and Mirror-Match [11]. These methods rely on

TABLE 1: Comparison results for the wall scene with repetitive structures.

Methods	N_{matches}	Precision	Recall
S-Match	378	98.41%	39.03%
Ratio-Match	918	79.41%	76.50%
Self-Match	378	95.77%	37.99%
Mirror-Match	378	95.24%	37.78%

¹ N_{matches} : the number of the SIFT feature matches.

evaluating the degree of similarity between the best matches and second best matches to find reliable matches and reduce the ambiguities caused by repetitive structures [10].

Figure 5 illustrates the experimental results of matching all the extracted features in the wall scenes shown in Figure 4. There are 2305 SIFT features extracted from the reference image to be matched. The number of the feature matches found by S-Match, Ratio-Match, Self-Match, and Mirror-Match methods is shown in Table 1.

Table 1 provides the matching precision and recall rates in Figure 5. The same number of matches is coincidentally found by the methods of S-Match, Self-Match, and Mirror-Match. In these three methods, the proposed S-Match method achieves the highest matching precision. The Ratio-Match method obtains the maximum number of matches with the best recall rate but the worst matching precision. Since the matching criteria of the other three methods are stricter than the Ratio-Match, they identify a less number of the matches with a higher matching precision but a lower recall rate.

We use the precision-recall curves for the comprehensive comparison of different matching methods. Figure 6 shows the precision-recall curves for 3 scenes from the INRIA dataset [19]. The results of the other 5 scenes in the dataset are not illustrated because all the matching methods achieve similar performance. In Figure 5, the curves can be obtained by changing λ and ϵ in (5) and (7). Each curve represents the average matching results of one method obtained from the LOOCV test. For each scene, we compare the average matching precision and recall rates for different methods. The results indicate that the proposed method achieves better matching precision for Walls, Trees, and Graffiti scenes. Along with the change of the threshold, the precision obtained by the proposed method slows down more gradually than that by the other methods. Such high precision is

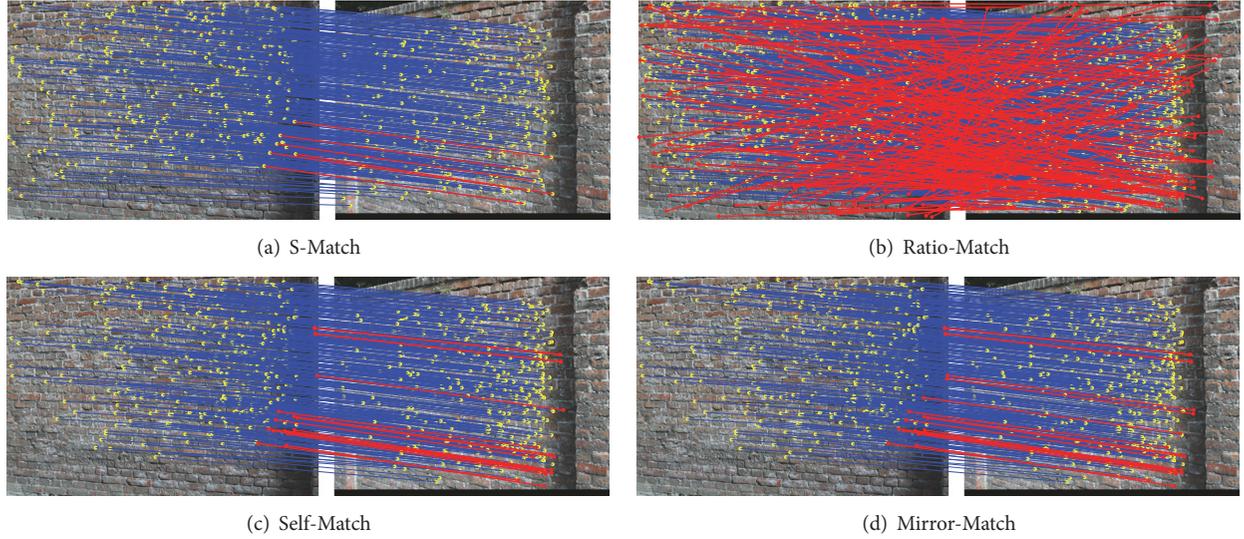


FIGURE 5: Feature matching for the wall scene with repetitive structures. The spots represent the matched features. Thin and bold lines indicate correct and incorrect matches, respectively.

Input. All the matching pairs of each scene for the LOOCV test.

- (1) Set a strict threshold for each method as stated in Section 3.1 to find the robust seeds of each matching pair;
- (2) Determine the minimum number of the seeds among all the methods;
- (3) If there is a method finding more seeds than the minimum number, go to(4); else, go to (6);
- (4) Sort the seeds based on their similarity values which are computed according to the Equation (5) and (7) in an ascending order;
- (5) Select the determined number of the seeds from the top-ranked sequence determined in step (4);
- (6) Compute the D , *Precision* and *Recall* of the seeds for each method according to (11) (13);
- (7) If the quality computation in step (6) for all the matching pairs is finished, go to (8); else go to (1);
- (8) Compute the average quality of the seeds for each scene.

Output. The average quality results for the seeds of each scene.

ALGORITHM 2: The experimental procedure of the quality computation.

preferred for region-based matching methods to select the seeds and will be used in the following section.

3.3. Seed Selection for Region-Based Feature Matching. Robust identification of reliable seeds is the foundation of region-based feature matching. In this paper, the S-Match method is used to determine the seeds. The bimatching methods using different matching criteria including Ratio-Match, Self-Match, and Mirror-Match are applied for comparison. These methods are denoted by biRatio-Match, biSelf-Match, and biMirror-Match, respectively. Figure 7 shows seed selection for tree scenes. Taking 51 seeds determined by different methods, for example, the qualities of seed distributions are different. We apply (9) to compute D to analyze the qualities of seed distributions. Table 2 shows the results which indicate that the proposed S-Match method achieves the highest quality.

We use the average precision, recall rates, and seed distributions for comparing the quality of the seeds comprehensively. The experimental procedure of computing the quality mentioned above for each scene is shown in Algorithm 2.

TABLE 2: The comparison results of D for tree scenes.

Methods	S-Match	biRatio-Match	biSelf-Match	biMirror-Match
D	3.6241	5.1749	5.2632	4.7410

Figure 8 shows some representative results of the seeds for comparing different matching methods. In Figure 8, each data point represents the average quality results for the seeds from the whole scenes but the cropped ones. For each scene, we compare the average precision, recall rate, and seed distribution for different methods. Figure 8(a) demonstrates that all the average precision of the seeds found by the S-Match method is equal to 100% approximately under the similar recall rates shown in Figure 8(b) with other methods. In most cases, the proposed method achieves a better average quality of seed distribution except that for the Graffiti scene as shown in Figure 8(c).

We use the determined seeds to perform the triangle constraint for the region-based feature matching. Similar

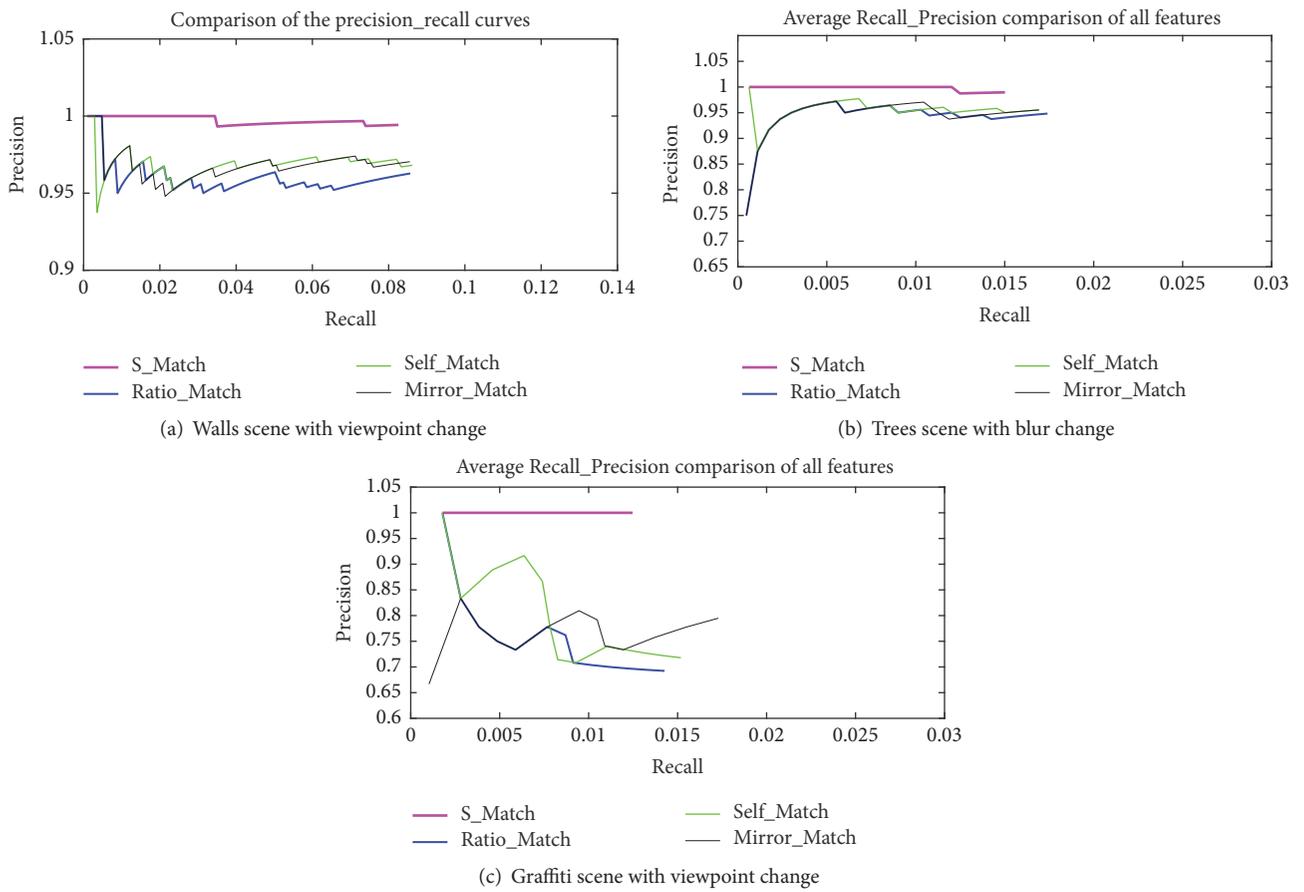


FIGURE 6: Comparison of the precision-recall curves.

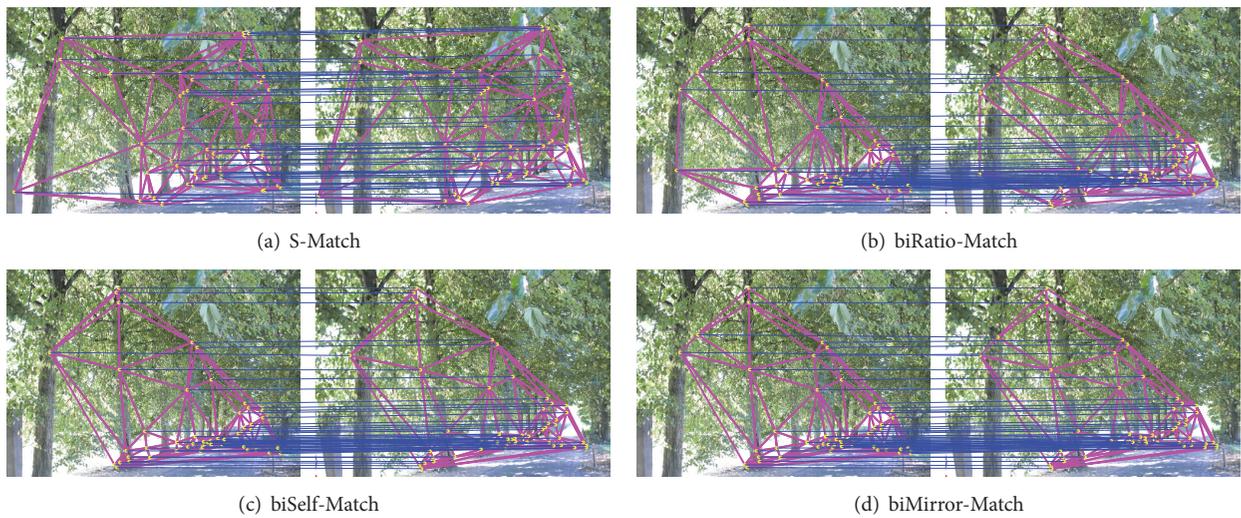


FIGURE 7: Seed selection for tree scenes. Thin and bold solid lines indicate the seeds and their convex hull, respectively.

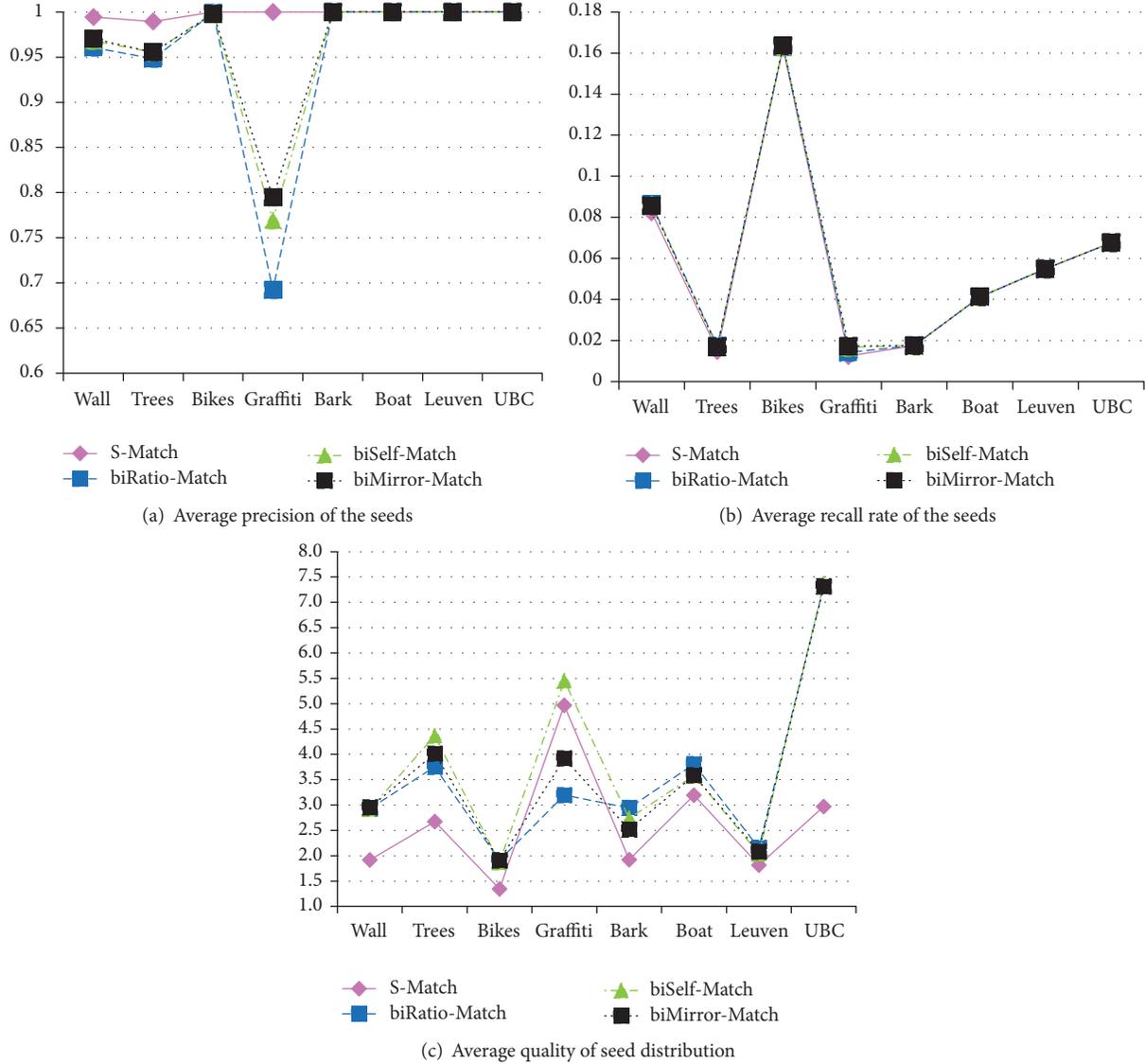


FIGURE 8: Comparison results of the seeds for different matching methods.

to the method proposed in [15], a triangle region of each unmatched feature is determined by three seed features in its neighborhood. The feature difference s_j between each unmatched feature f_i located at $\mathbf{p}_i = (x_i, y_i)$ and the candidate feature $f_j^{(1)}$ located at $\mathbf{c}_j = (x_j^{(1)}, y_j^{(1)})$ is computed as follows [15]:

$$s_j = 1.5^{-\left(\|\mathbf{p}_i - \mathbf{c}_j\|/d_{\text{proj}}\right)^2} \times \mathbf{f}_i^T \mathbf{f}_j^{(1)}, \quad (j = 1, 2, \dots, n), \quad (13)$$

where $d_{\text{proj}} = 4$ and n is the number of the candidate features extracted from the observation image. The best match for each unmatched feature is determined by the minimum value of s_j .

To validate the effectiveness of the proposed region-based feature matching method for finding more reliable matches from the repetitive structures, both the average precision and the average recall rate of the total matches are given in

Figure 9. In the figure, all the methods combined with the triangle constraint are denoted by S-Match + TC, biRatio-Match + TC, biSelf-Match + TC, and biMirror-Match + TC, respectively. We can see that the S-Match + TC method outperforms the other methods. In most cases, it achieves a higher precision and a higher recall rate simultaneously. In the case of the Bark scene, the average precisions of S-Match + TC, biRatio-Match, biSelf-Match + TC, and biMirror-Match + TC are 84.44%, 86.16%, 86.99%, and 86.91% as shown in Figure 9(a) while the average recall rates are 45.78%, 29.84%, 29.29%, and 30.04%, respectively, in Figure 9(b). The precision of the S-Match + TC method is slightly lower. However, a significantly higher recall rate is obtained. The results shown in Figure 9 indicate the proposed S-Match + TC identifies more correct matches. Especially for the Graffiti scene, the proposed method achieves the best final results among all the methods although the average quality

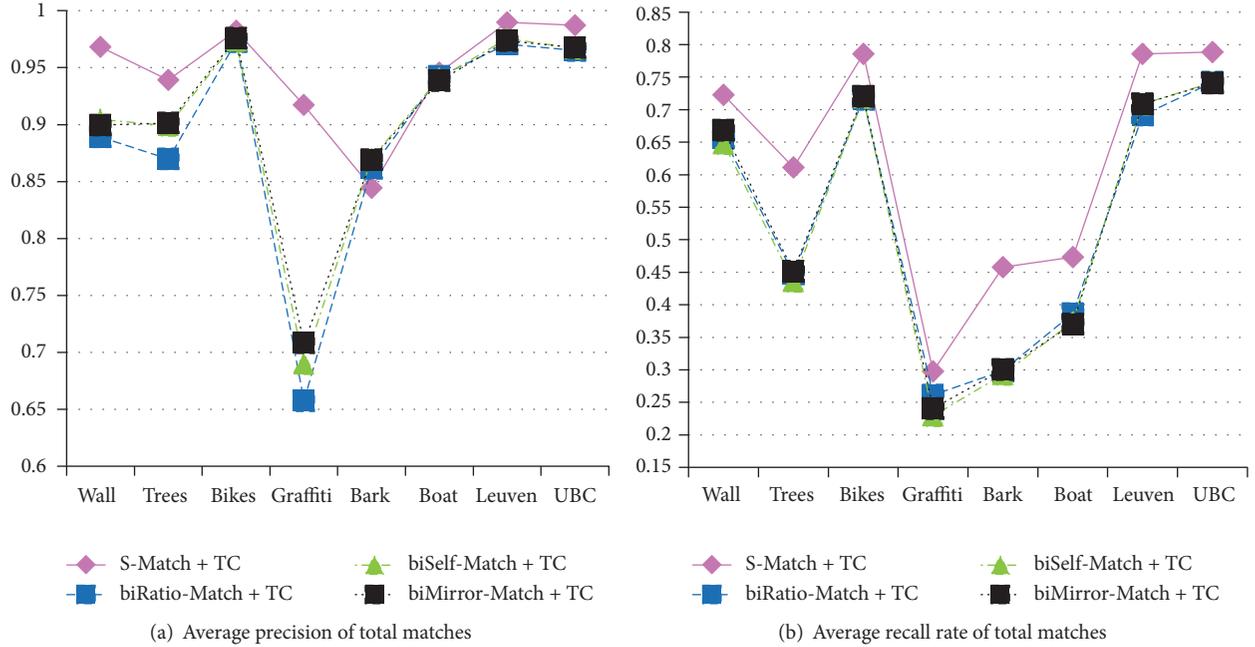


FIGURE 9: Comparison results of the total matches for different matching methods.

of seed distribution is not the best according to Figure 8(c). It indicates that the precision of the seeds also plays a critical role for the region-based feature matching.

3.4. Computational Complexity. As shown in Section 2.2, the Monte-Carlo simulation for computing the feature saliency is time consuming. However, it can be performed before pairwise matching begins in practical applications such as the landmark-based navigation [23, 24]. In these applications, a detailed understanding of the environments is usually required for selecting a small number of features.

Given the feature saliency, the proposed S-Match method has a lower computational complexity than existing state-of-the-art methods because it has the ability to preselect fewer salient and stable features as the matching candidates in the reference image by adjusting the parameters of γ and η . In the existence of a large number of repetitive structures in the scenes, it has important practical significance to reduce the computational cost by utilizing the proposed method. Table 3 lists the computational complexity of both the proposed method and other single-direction matching methods [7, 10, 11]. Given n features in the reference image and m features in the observation image, our method can be implemented in $O(n' \log(m))$, where n' is the number of the preselected salient features from the reference image. When the selected feature number n' is far less than n or m , the proposed method can be performed more efficiently.

4. Conclusions

In this paper, we present a robust image matching strategy based on SIFT saliency. The feature saliency analysis indicates the close relationship between the feature distinctiveness

TABLE 3: Comparison of computational complexity.

Algorithm	Complexity
S-Match	$O(n' \log(m))$
Ratio-Match	$O(n \log(m))$
Self-Match	$O(n \log(nm))$
Mirror-Match	$O(n \log(n+m))$

and the region of interest. We start to find robust matches for the salient features and perform the triangle constraint to reduce the ambiguities caused by the repetitive features. Experimental results show that the proposed saliency-based feature matching method outperforms the state-of-the-art methods. More correct matches from the repetitive features can be found by combining the triangle constraint method.

In the future, we will focus on estimating the ROI for each repetitive feature and improve the region-based matching performance furtherly.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by National Natural Science Foundation of China (no. 61272523), the National Key Project of Science and Technology of China (no. 2011ZX05039-003-4), General Project of Education Department of Liaoning Province (no. L201683682), and the Fundamental Research Funds for the Central Universities (no. DUT15QY33).

References

- [1] M. Amiri and H. R. Rabiee, "RASIM: a novel rotation and scale invariant matching of local image interest points," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3580–3591, 2011.
- [2] J. T. Arnfred, S. Winkler, and S. Süsstrunk, "Mirror match: Reliable feature point matching without geometric constraints," in *Proceedings of the 2013 2nd IAPR Asian Conference on Pattern Recognition, ACPR 2013*, pp. 256–260, Japan, November 2013.
- [3] J. T. Arnfred and S. Winkler, "A general framework for image feature matching without geometric constraints," *Pattern Recognition Letters*, vol. 73, pp. 26–32, 2016.
- [4] T. Avraham and M. Lindenbaum, "Attention-based dynamic visual search using inner-scene similarity: Algorithms and bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 251–264, 2006.
- [5] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [6] D.-Y. Chen and Y.-M. Chiu, "Visual attention guided video copy detection based on feature points matching with geometric-constraint measurement," *Journal of Visual Communication and Image Representation*, vol. 24, no. 5, pp. 544–551, 2013.
- [7] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, vol. 761, Prentice Hall, London, England, 1982.
- [8] X. Guo and X. Cao, "Triangle-constraint for finding more good features?" in *Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR 2010*, pp. 1393–1396, Turkey, August 2010.
- [9] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2214–2219, November 2011.
- [10] M. Kushnir and I. Shimshoni, "Epipolar geometry estimation for urban scenes with repetitive structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 12, pp. 2381–2395, 2014.
- [11] R. Lerner, E. Rivlin, and I. Shimshoni, "Landmark selection for task-oriented navigation," *IEEE Transactions on Robotics*, vol. 23, no. 3, pp. 494–505, 2007.
- [12] T. Leung and J. Malik, "Detecting, localizing and grouping repeated scene elements from an image," in *Computer Vision — ECCV '96*, vol. 1064 of *Lecture Notes in Computer Science*, pp. 546–555, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.
- [13] Q. Li, H. Zhang, and T. Wang, "Scale invariant feature matching using rotation-invariant distance for remote sensing image registration," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 2, Article ID 1354004, 2013.
- [14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [15] K. L. Lurie, R. Angst, and A. K. Ellerbee, "Automated mosaicing of feature-poor optical coherence tomography volumes with an integrated white light imaging system," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 7, pp. 2141–2153, 2014.
- [16] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [17] K. Mikolajczyk, T. Tuytelaars, C. Schmid, and et al., *Affine covariant features, Collaborative work between: the Visual Geometry Group, Katholieke Universiteit Leuven, Inria Rhone-Alpes and the Center for Machine Perception*, 2004.
- [18] P. Sala, R. Sim, A. Shokoufandeh, and S. Dickinson, "Landmark selection for vision-based navigation," *IEEE Transactions on Robotics*, vol. 22, no. 2, pp. 334–349, 2006.
- [19] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi, "Visual place recognition with repetitive structures," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 883–890, USA, June 2013.
- [20] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.
- [21] A. R. Zamir and M. Shah, "Image geo-localization based on multiplenearest neighbor feature matching usinggeneralized graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1546–1558, 2014.
- [22] B. Zeisl, T. Sattler, and M. Pollefeys, "Camera pose voting for large-scale image-based localization," in *Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015*, pp. 2704–2712, Chile, December 2015.
- [23] G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger, "Signature detection and matching for document image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2015–2031, 2009.
- [24] Q. Zhu, B. Wu, and Z.-X. Xu, "Seed point selection method for triangle constrained image matching propagation," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 2, pp. 207–211, 2006.



Hindawi

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

