

## Research Article

# A Novel Fast and Robust Binary Affine Invariant Descriptor for Image Matching

**Xiujie Qu, Fei Zhao, Mengzhe Zhou, and Haili Huo**

*School of Information and Electronics, Beijing Institute Technology, Beijing 100081, China*

Correspondence should be addressed to Xiujie Qu; [quxiujie@bit.edu.cn](mailto:quxiujie@bit.edu.cn)

Received 20 November 2013; Revised 3 January 2014; Accepted 23 January 2014; Published 4 March 2014

Academic Editor: Wuhong Wang

Copyright © 2014 Xiujie Qu 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 the current binary descriptors have disadvantages of high computational complexity, no affine invariance, and the high false matching rate with viewpoint changes, a new binary affine invariant descriptor, called BAND, is proposed. Different from other descriptors, BAND has an irregular pattern, which is based on local affine invariant region surrounding a feature point, and it has five orientations, which are obtained by LBP effectively. Ultimately, a 256 bits binary string is computed by simple random sampling pattern. Experimental results demonstrate that BAND has a good matching result in the conditions of rotating, image zooming, noising, lighting, and small-scale perspective transformation. It has better matching performance compared with current mainstream descriptors, while it costs less time.

## 1. Introduction

The local feature descriptor is the core of many computer vision technologies, such as object recognition, image retrieval, and 3D reconstruction. How to design a local feature descriptor that has excellent performance and low complexity is an important and difficult research. Many scholars have proposed a variety of descriptors in this area, such as SIFT (Scale Invariant Feature Transform) [1–3], which has good scale invariance and robustness to illumination and viewpoint changes. Because of constructing scale space and many other steps to improve accuracy, SIFT has high computational complexity. All of these make it prohibitively slow.

For these problems, many improved algorithms were proposed. SURF (Speeded Up Robust Features) [4] has similar matching rate with much faster performance by describing key points with the responses of few Haar-like filters compared with SIFT. PCA-SIFT [5] reduced dimensionality of the descriptor from 128 to 36 by using the principal component analysis, but the time of descriptor formation was increased. The GLOH descriptor [6] used the circular pattern in log-polar coordinates in order to enhance the robustness and uniqueness of the descriptors, but it is also

more expensive to compute than SIFT. ASIFT descriptor [7] achieved matching feature points well when there were large changes in viewpoint.

To some extent, above-mentioned descriptors improved the performances, but they still have high computational complexity because of the local histogram statistics. Besides, each dimension of the descriptor is a decimal number, which makes them need a lot of memory. All of these make them difficult to be achieved on a low-power, low-memory application. In recent years, many binary descriptors appear, such as BRIEF (Binary Robust Independent Elementary Features) descriptor [8], ORB (Oriented FAST and Rotated BRIEF) descriptor [9], BRISK (Binary Robust Invariant Scalable Keypoints) descriptor [10], RIFF (Rotation-Invariant Fast Features) descriptor [11], and FREAK (Fast Retina Keypoint) descriptor [12]. These binary descriptors need lower memory than SIFT-like descriptors, and the similarity of descriptors can be evaluated by computing the Hamming distance, which is very efficient to be computed. But one of their shortcomings is that the shape of pattern cannot be changed. In other words, there are only circular pattern and rectangle pattern, which makes them have no affine invariance. In this case, the high false matching rate between different viewpoints images will appear. Another shortcoming is that

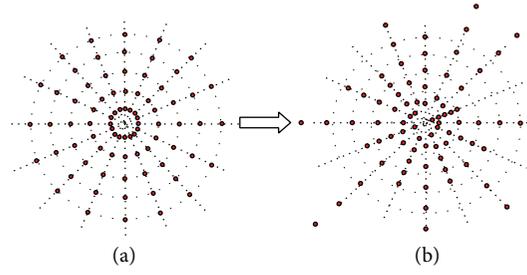


FIGURE 1: (a) is a circular pattern and (b) is an irregular shape one. The red balls are the sampling points.

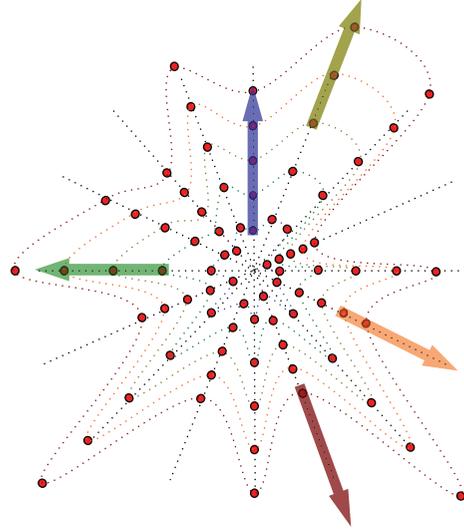


FIGURE 2: Multidirectional rotation invariant builder. The arrows represent the five local orientations. We can see that the pattern is irregular.

one pattern only has one local orientation, and they are not fast enough because of the histogram statistics. In this paper we propose a new binary descriptor with low computational complexity. The orientations of descriptor are found by the five concentric circles, and the affine invariant regions are found by 16 rays which are emanating from the feature point. Experimental results show that the binary affine invariant descriptor has good matching result in the conditions of rotating, image zooming, noising, lighting, and small-scale perspective transformation. In this paper we call this binary descriptor BAND (Binary Affine Invariant Descriptor).

## 2. Method

We describe 3 key steps of BAND in this section, namely, affine invariant radius builder, multidirectional rotation invariant builder, and simple random sampling pattern builder.

**2.1. Affine Invariant Radius Builder.** There are many algorithms that extract the local affine invariant regions, such as MSER (Maximally Stable Extremal Regions) [13], EBR (Edge-Based Regions), IBR (Intensity Extrema-Based Regions) [14], and Salient Region [15]. Here we refer to the idea of IBR. With the aim of extracting local affine invariant regions, the 16 rays

which are emanating from the feature point are extracted. The gray value of each sampling point located in the rays is obtained through quadratic interpolation. Here we define the affine invariant radius  $R_m$  as

$$R_m = T \cdot \max(R_1, R_2),$$

$$R_1 = L \min \left( \text{abs} \int_c^i (g_i - g_c) dr \right), \quad (1)$$

$$R_2 = L \max \left( \text{abs} \left( \int_c^i (g_i - g_c) dr \right) \right),$$

where  $R_m$  is the affine invariant radius in the direction of  $p$ ,  $i$  is the Euclidean arc length along the ray  $R_m$ ,  $g_i$  is the gray value of the radius at the position  $i$ ,  $\text{abs}(x)$  is the absolute value of the  $x$ ,  $L \min(y)$  is the minimum value of the distance between local minimum of the  $y$  and the feature point,  $L \max(y)$  is the maximum value of the distance between local maximum of the  $y$  and the feature point, and  $T$  is a constant, and we use a value of 0.76, which is determined by experiments.

By calculating the affine invariant radius, the original circular pattern is deformed into irregular shape pattern. It is shown in Figure 1.

**2.2. Multidirectional Rotation Invariant Builder.** In order to get the rotation invariant property, existing algorithms need

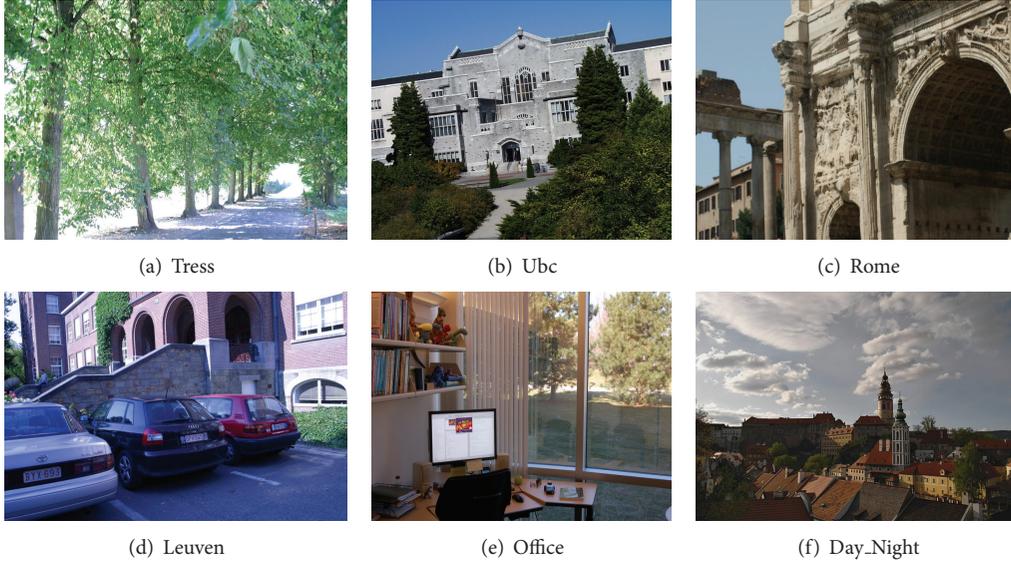


FIGURE 3: Datasets used for evaluation: blur (Trees), JPEG compression (Ubc), rotation (Rome), and brightness changes (Leuven, Office, and Day\_Night).

to determine the direction of the local area and rotate the axis. In this process, the defects of these descriptors include

- (1) BRIEF descriptor ignores the rotation; therefore a lot of outliers appear in the case of large-angle rotation.
- (2) The algorithms, like SIFT and ORB, use the rectangle pattern. It is difficult to obtain the local orientation, and they have a high computational complexity.
- (3) The algorithms, like FREAK and BRISK, only consider one local orientation with a pattern, although using the circular pattern.

Based on the above three points and reference of the BRISK and FREAK, the pattern of BAND descriptor is as follows: construct five circles concentric with the feature point, and there are 16 sampling points in each circle homogeneously. Every affine invariant radius of every circle is obtained by the affine invariant radius builder.

Firstly, obtain the locations of sampling points on every circle by affine invariant radius:

$$(x_m, y_m) = \left( x_c + R_M \sin\left(\frac{2\pi p}{P}\right), y_c - R_M \cos\left(\frac{2\pi p}{P}\right) \right), \quad (2)$$

where  $m$  is the index of the sampling points surrounding the feature point,  $M$  is the number of sampling points, and  $c$  is the location of the feature point.

Secondly, according to literature [14], LBP binary string of every circle is

$$\text{LBP}_{v,w} = \begin{cases} 1, & \text{if } g_p > g_c, \\ 0, & \text{else,} \end{cases} \quad (3)$$

where  $v$  is the index of the circle,  $w$  is the location of one bit of the string,  $g_p$  is the gray value at the position  $p$  on the circle,

and  $g_c$  is the gray value of feature point. Finally, according to the rotation invariant LBP [16], we could obtain five local orientations. The rotation invariant LBP defined by Ojala is  $\text{LBP}_{P,R}^{ri}$ :

$$\text{LBP}_{P,R}^{ri} = \min \left( \text{ROR} \left( \text{LBP}_{P,R}^{ri}, i \right) \mid i = 0, 1, \dots, P-1 \right), \quad (4)$$

where  $\text{ROR}(x, i)$  performs a circular bitwise right shift operation on the number  $x$  with  $i$  times. The pattern of BAND descriptor is shown in Figure 2.

**2.3. Simple Random Sampling Pattern Builder.** This step is very simple. We get a 256 bits vector by sampling point pairs from the five LBP coding strings randomly and orderly. The bit-vector descriptor is assembled by performing all the comparisons of point pairs  $(g_a, g_b)$ , such that each bit  $L$  corresponds to

$$L = \begin{cases} 1, & \text{if } g_a > g_b, \\ 0, & \text{else,} \end{cases} \quad (5)$$

$$i = 0, 1, 2, \dots,$$

where  $g$  is the gray value of the point sampled from the five LBP strings.

### 3. Experiments and Analysis

Programming environment is Matlab 2013, Visual Studio 2010, and OpenCV 2.4.4. The datasets come from [17]. All algorithms are running on an Intel(R) Core (TM) i3 of 3.40 GHz. In order to unify the results, all of the results do not remove the mismatches and only show the best results by setting a threshold.

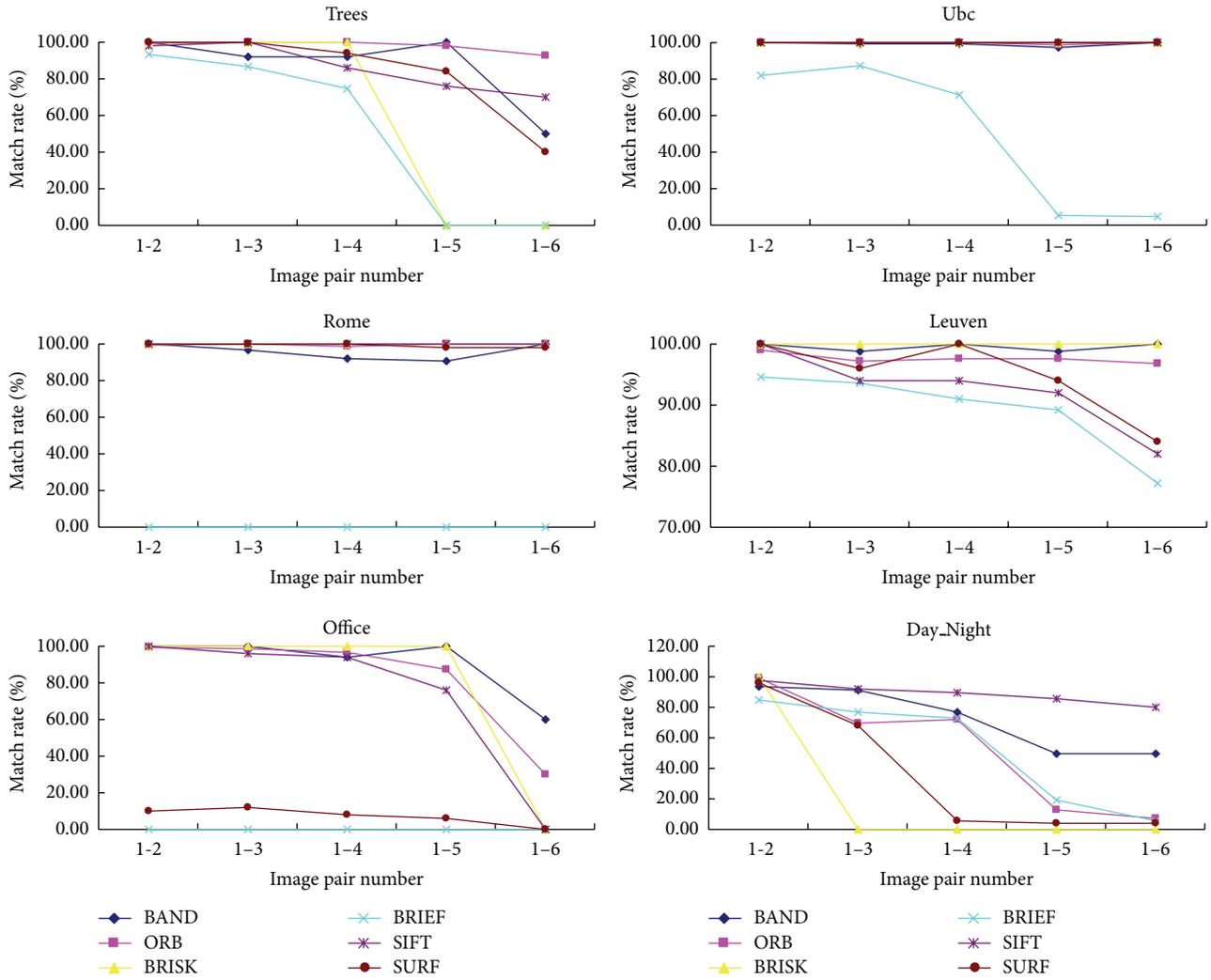


FIGURE 4: Evaluation result shows match rate for BAND, ORB, BRISK, BRIEF, SIFT, and SURF.

3.1. *Single Performance Verification of Local Affine Invariant Descriptors.* Each of the datasets contains a sequence of six images exhibiting an increasing amount of transformation. All comparisons here are performed against the first image in each dataset. Figure 3 shows one image for each dataset analyzed.

The transformations cover blur (Trees), brightness changes (Leuven, Office, and Day\_Night), JPEG compression (Ubc), and rotation (Rome). Match rate is defined as a ratio between the number of the correctly matching points and the total number of the matched points. In order to prove the easy integration of our algorithm, we use FAST [18, 19] for detecting the feature points in the sequence of the Trees and the SURF in the others. To make results be comparable, the threshold of Hamming distance for BAND, BRIEF, BRISK, and ORB is 50. We only show the best 50 matched pairs in SURF and SIFT. All of the results are shown in Figure 4.

3.2. *Images Matching for Outdoor Scenes.* We only test single performance of BAND descriptor in Section 3.1. In this

section, we would use 6-pair images showed in Figure 5 to test its comprehensive performance. As it is shown in Figure 5, all of the 6 pairs contain view point changes in different degrees, such as in scale, brightness, and rotation. Besides, all of the images contain many structures which appear repeatedly. For example, the structures in image Church and the image Fountain both are symmetrical, and the buildings in image Brussels, Venice, Semper and Rathaus have similar doors, windows and statues individually. These conditions add extra difficulty to matching.

Figure 6 shows the matching results of the BAND. The matched point numbers and match rates of different descriptors are shown in Figures 7 and 8. To make results be comparable as in Section 3.1, the threshold of Hamming distance about BAND, BRIEF, BRISK and ORB is 50 invariably. We only show the best 50 matched pairs in SURF and SIFT as before.

We use SURF for detecting the feature points, and the threshold is equal to before. The 6-pair outdoor scenes are as shown in Figures 5 and 6.

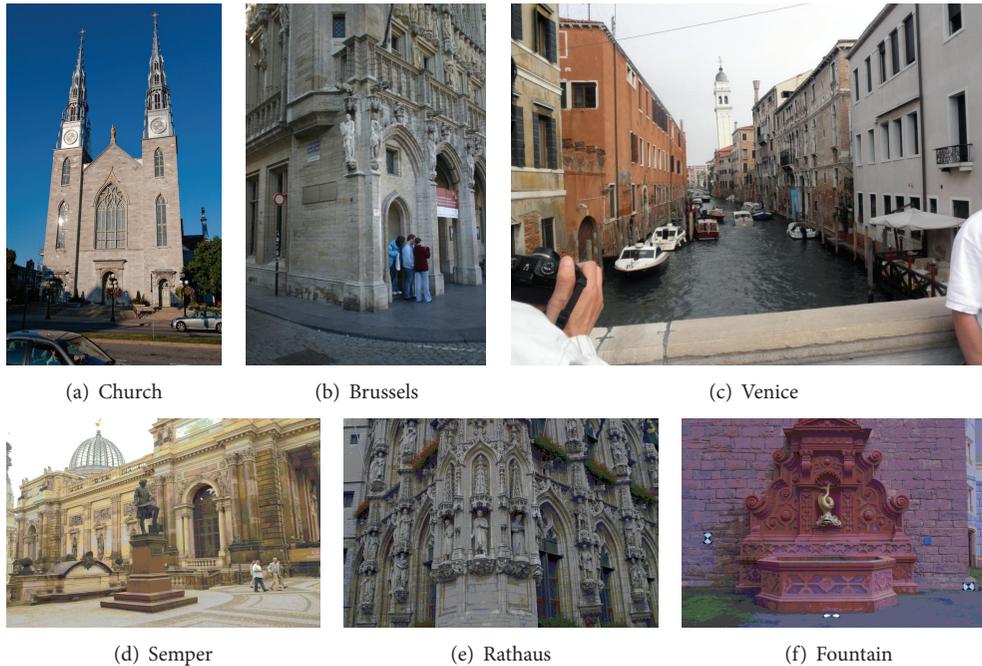


FIGURE 5: The 6-pair outdoor scenes are all the buildings and artifacts.

**3.3. Time Consumption Comparison.** In Matlab 2013 platform, we compare the time consumption between SIFT and BAND descriptor per point pair. Test result is shown in Table 1.

If we set the time of BAND describing one feature point to be 1, then we would get the time consumption of other descriptors in Table 2 clearly.

**3.4. Analysis of Experimental Results.** In the single performance verification, BAND is demonstrated as it has a better adaptability than the other descriptors. In 6 experiments, the degree of transformations is increasing. It is noteworthy that the Leuven shows a transformation in brightness from normal to dark gradually. Similarly, the Office shows a transformation in brightness from dark to normal. Differently, the Day\_Night shows a transformation in brightness from normal to dark, but some local areas become lighter than surroundings because of some lights, such as lamps, bulbs, and candles. In other words, it changes nonlinearly. Since the similarity is reduced, the direction of every line in Figure 4 shows downward. BRIEF shows the worst performance among all of the descriptors, because it does not note the variety of rotation and scale. In the light experiments, such as the Leuven, the Office, and the Day\_Night, BAND demonstrates outstanding performance. BAND can maintain a high rate of correctly matched no matter how seriously changes of lightness. In other experimental conditions, performances of BAND are similar to the mainstream descriptors. In addition, no matter what feature points we use, like FAST and SURF, BAND descriptors can work well, which shows a high adaptability.

In the experiment of matching the outdoor scenes, we can get a lot of matched points by ORB descriptor, but its

TABLE 1: The time consumption comparison between SIFT and BAND descriptor per point pair.

	SIFT	BAND
Description time (ms)	531.1	0.5463
Matching time (ms)	0.093	0.0015

TABLE 2: It shows time consumption comparison of different descriptors.

	Time consumption					
	BAND	ORB	BRISK	BRIEF	SIFT	SURF
Feature description	1	8.5	5.7	8.3	1000	65

correctly matched rate is lower than BAND descriptor. One of its negative effects is that there will be more computation in the process of removing the mismatched points. Although the amount of the matched points we get by BAND is not as much as ORB, the amount is kept at a high level and the correctly matched rate of it is satisfactory. It is better than or as good as the correctly matched rate of other descriptors. The reason is that BAND considers the affine invariance and its pattern, whose shape is variable, can better adapt to the changes of viewpoint.

In the experiment of the time-consuming comparison, description time per point of BAND is as short as 0.1% of the time of SIFT approximately. The reason is that there is no process of establishing the Gauss Pyramids or too much computation of fitting operation. BAND computes the distance between two descriptions by using Hamming distance, but SIFT uses Euclidean distance. Obviously, calculating the Hamming distance has higher efficiency and lower

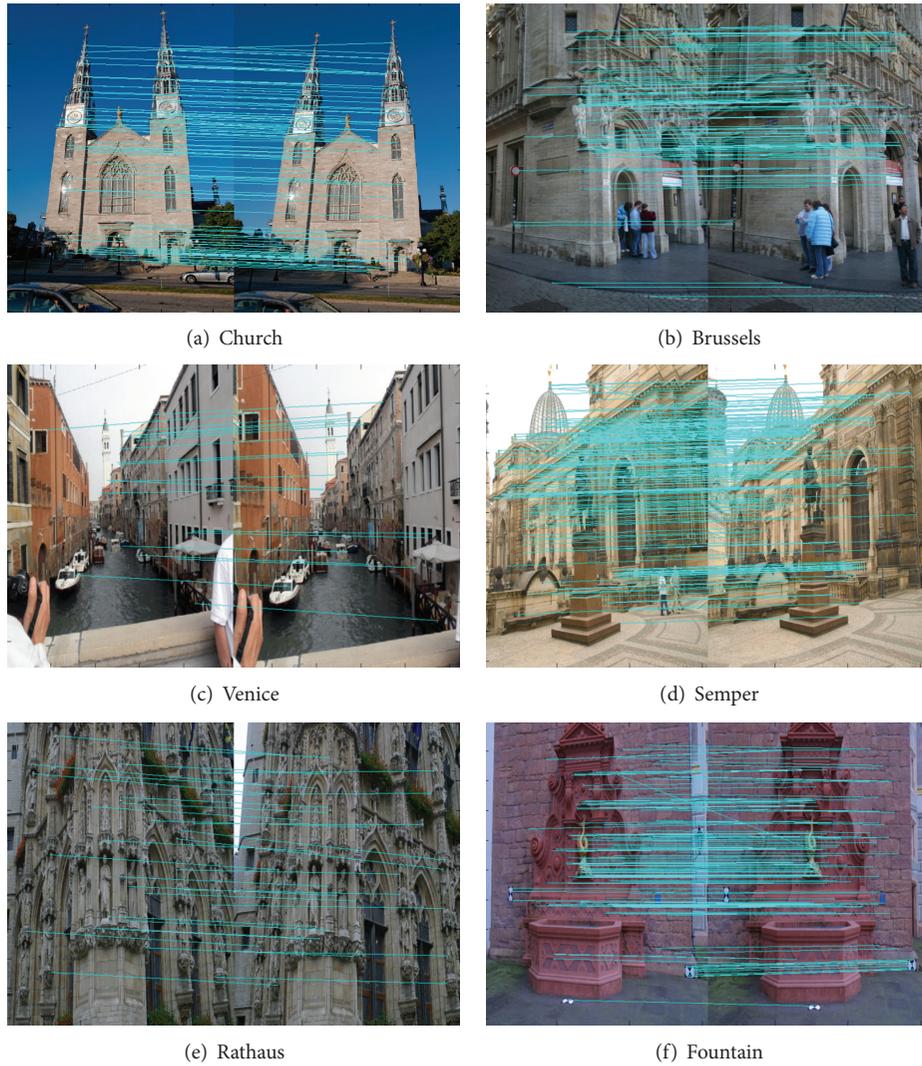


FIGURE 6: The results show the matched points, and there is no process of removing the mismatched.

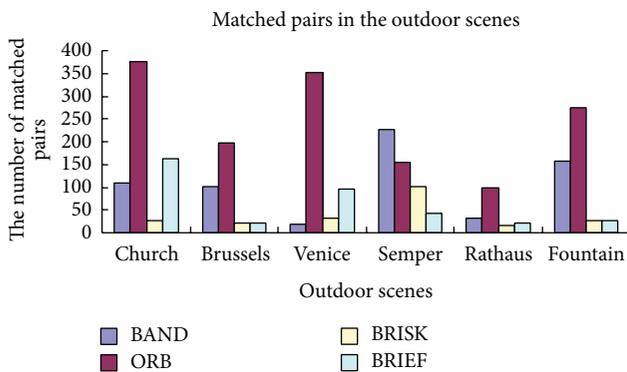


FIGURE 7: This figure shows the numbers of matched point pairs from different binary descriptors.

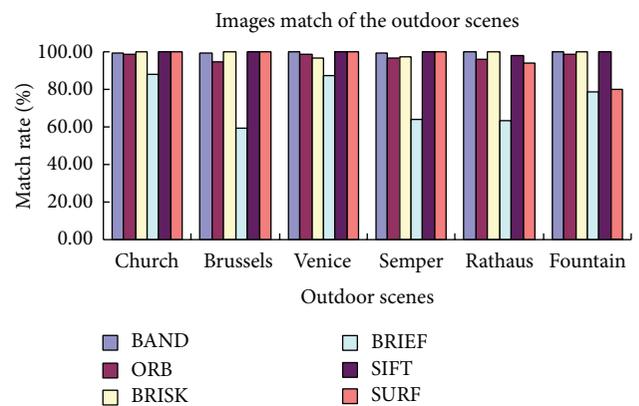


FIGURE 8: This figure shows the match rates of different binary descriptors in the 6 outdoor scenes.

computational complexity. So, together, BAND shows higher efficiency. And the comparison of the time-consuming with other descriptors also shows that BAND is faster. Therefore,

BAND has advantage in the situation that demands high computing speed.

#### 4. Conclusion

Experimental results show that BAND descriptor has significant advantages, such as low computational complexity, well adaptability, and good stability. It makes up for the disadvantages of other descriptors that have high computational complexity and have no affine invariance. BAND has a good matching result in the conditions of rotating, image zooming, noising, lighting, and small-scale perspective transformation. More specifically, it has a moderate number of matched points and a high correctly matched rate. Because of these, on the base of guaranteeing accuracy, BAND could improve the calculation efficiency and meet real-time requirement.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

#### References

- [1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, IEEE, Kerkyra, Greece, September 1999.
- [2] D. G. Lowe, "Local feature view clustering for 3D object recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. I-682–I-688, IEEE, Kauai, Hawaii, USA, December 2001.
- [3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [5] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-506–II-513, IEEE, Washington, DC, USA, July 2004.
- [6] 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.
- [7] J. M. Morel and G. Yu, "ASIFT: a new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.
- [8] M. Calonder, V. Lepetit, C. Strecha et al., "BRIEF: binary robust independent elementary features," in *Computer Vision—ECCV 2010*, vol. 6314 of *Lecture Notes in Computer Science*, pp. 778–792, Springer, Berlin, Germany, 2010.
- [9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2564–2571, IEEE, Barcelona, Spain, November 2011.
- [10] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: binary robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2548–2555, IEEE, Barcelona, Spain, November 2011.
- [11] G. Takacs, V. Chandrasekhar, S. Tsai et al., "Rotation invariant fast features for large-scale recognition," in *Applications of Digital Image Processing XXXV*, vol. 8499 of *Proceedings of SPIE*, San Diego, Calif, USA, October 2012.
- [12] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: fast retina keypoint," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 510–517, IEEE, Providence, RI, USA, June 2012.
- [13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [14] T. Tuytelaars and L. van Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.
- [15] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.
- [16] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [17] C. Strecha, W. von Hansen, L. van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [18] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision—ECCV 2006*, vol. 3951 of *Lecture Notes in Computer Science*, pp. 430–443, Springer, Berlin, Germany, 2006.
- [19] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.



# Hindawi

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

