Robust Feature Description and Matching Using Local Graph

Man Hee Lee and In Kyu Park
School of Information and Communication Engineering, Inha University
Incheon 402-751, Korea
E-mail: {maninara@hotmail.com, pik@inha.ac.kr}

***Abstract*—Feature detection and matching are essential parts in most computer vision applications. Many researchers have developed various algorithms to achieve good performance, such as SIFT (Scale-Invariant Feature Transform) and SURF (Speeded Up Robust Features). However, they usually fail when the scene has considerable out-of-plane rotation because they only focus on in-plane rotation and scale invariance. In this paper, we propose a novel feature description algorithm based on local graph representation and graph matching based, which is more robust to out-of-plane rotation. The proposed local graph encodes the geometric correlation between the neighboring features. In addition, we propose an efficient score function to compute the matching score between the local graphs. Experimental result shows that the proposed algorithm is more robust to out-of-plane rotation than conventional algorithms.**

I. Introduction

Several computer vision techniques like structure from motion [1] or visual SLAM (Simultaneous Localization and Mapping) [7] utilize a number of images or video frames as their inputs. However, to estimate camera motion or generate 3D map in those applications, they usually employ computationally efficient feature detection method because a large number of high resolution images should be used. Consequently, the output quality highly is affected by the correctness of detected features and matching results.

A few algorithms, such as SIFT (Scale-Invariant Feature Transform) [9] or SURF (Speeded Up Robust Features) [5], have been developed for feature detection and description, which have good performance in various situations, such as in-plane rotation, scale and illumination change, and blurring. In addition, FAST (Features from Accelerated Segment Test) [11] can detect corners in real-time. However, if the scene has out-of-plane rotation caused by large viewpoint changes, these algorithms match many outliers which lead to incorrect camera motion estimation. Note that out-of-plane rotation is often observed in real situations.

In order to handle out-of-plane rotation correctly, in this paper, we propose a novel local graph and graph matching based feature description algorithm. The proposed algorithm generates the graph based descriptor in feature domain as well as easily calculates descriptor's difference using the proposed score function.

This paper is organized as follows, In Section II, related works are introduced. Section III describes the proposed local graph generation algorithm. The feature description using the proposed local graph is described in Section IV. Section V provides the experimental results. Finally, we give a conclusive remark in Section VI.

II. Related Works

A. Feature Detection or Feature Description

Agrawal *et al.* [2] proposed the scale-invariant center-surround detector that is computed over multiple scales of the original image resolution. Then, they utilized the simplified bi-level kernels for fast computation using slanted integral image. In addition, they used second moment matrix to remove weak features which lie along edges or lines. Rosten *et al.* [11] proposed a fast and high quality corner detector, which used the segment-test algorithm for detecting features. Then, they applied a machine learning approach to implement the algorithm efficiently for working in real time.

Calonder *et al.* [6] proposed the fast and accurate interest point descriptor for real-time applications using short binary strings as a feature point descriptor. They directly computed binary strings from image patches and then evaluated Hamming distance for the matching score between two binary descriptors. However, it is not rotation and scale invariant. Alahi *et al.* [3] proposed the binary keypoint descriptor inspired by the retina. They utilized a series of Difference-of-Gaussian (DoG) functions over retinal sampling pattern to generate binary descriptor.

B. Simultaneous Feature Detection and Description

Lowe [9] proposed the scale-invariant feature detection based on multi-scale DoG function and feature description based on histogram of oriented gradient. However, high dimensionality of the descriptor makes it run slowly. Bay *et al.* [5] proposed the speeded-up version of SIFT using integral images and approximated integer Gaussian filter. In addition, they did not resize the original image for multi-scale but simply increased the size of filter. If it is applied to integral image, Hessian's determinants are calculated in constant time. Rublee *et al.* [12] proposed a fast feature detection and description algorithm based on the oriented FAST detector and rotated BRIEF descriptor. Because BRIEF is rotation variant, they calculated the orientation of FAST features using the intensity centroid scheme. Leutenegger *et al.* [8] proposed the fast and low cost feature detection and description algorithm. They used the scale space keypoint detection based on FAST

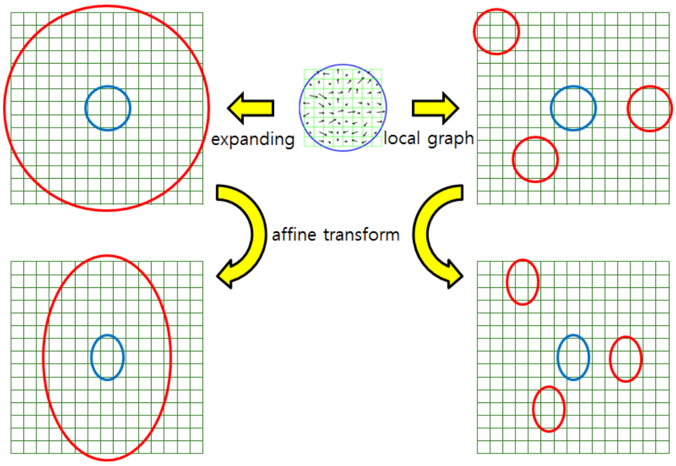

Fig. 1. An example of the up-scaled descriptor and the proposed local graph representation under out-of-plane rotation.

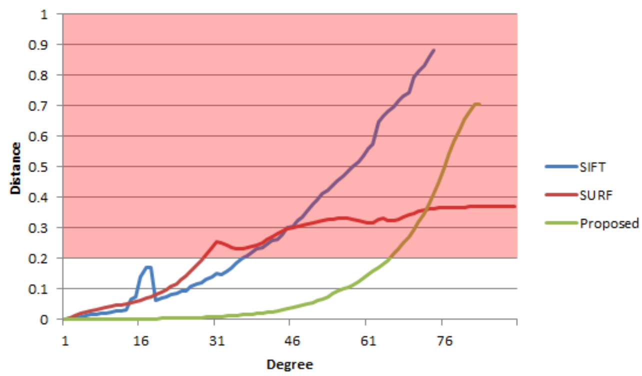

Fig. 2. Variation of descriptor's distance under out-of-plane rotation. High distance score in red shade usually matches outliers, which has less meaning in the simulation.

using scale pyramid and used circular sampling pattern to generate binary descriptor by performing a simple brightness comparison test. Alcantarilla *et al.* [4] proposed the new feature detection algorithm in nonlinear scale space using nonlinear diffusion filtering rather than Gaussian filtering. Except that, it is similar to SIFT detector and SURF descriptor.

Although conventional algorithms are robust to various imaging conditions such as in-plane rotation and scale changes, most of the previous descriptors fail if the scene has considerable out-of-plane rotation which is usually observed when the camera motion is large.

III. Local Graph under Out-of-Plane Rotation

The robustness of a descriptor under out-of-plane rotation can be improved by considering features in a larger surrounding area. In fact, it extends the support region of the descriptor. For this purpose, the simplest way is to extend the size of descriptor. However, a big-sized descriptor yields more difference than a small-sized descriptor in affine transform because the variation by an affine transform in large scale are bigger than the original scale. Therefore, the similarity of descriptors decreases in this case.

On the other hand, if graph structure encoding neighboring local features' correspondence is utilized, the topology of graph is mostly preserved in spite of underlying affine transformation. Fig. 1 shows an example that the local graph representation is more invariant than the multi-scale descriptor under out-of-plane rotation.

To evaluate the effectiveness of the proposed local graph representation, we generate a synthetic image which has four blobs with Gaussian distribution on black background. The input image is rotated (out-of-plane) from 0 to 90 degrees using perspective projection simulation. Using the rotated images and the original images, the distances of SIFT and SURF descriptors are calculated. Fig. 2 shows the plot of distance variation. As shown in Fig. 2, the distances using SIFT and SURF descriptors rapidly increases in the region with meaningful scores, which means that these descriptors are not robust to out-of-plane rotation. Unlike SIFT and SURF, the proposed graph based descriptor is more robust and shows the increased matching probability.

IV. Feature Description using Local Graph

In order to describe features using local graph in feature domain, the keypoints of the input image are detected first. To this purpose, several algorithms such as corner or blob detection can be used optionally. In this paper, FAST and SURF detectors are employed since they have good performance not only in the repeatability of features under affine transform but also with the inexpensive computational cost. Using the detected features, the local graph in each feature is constructed by exploring n closest features from the root feature. In our experiment, we set n to 16. Each node has normalized distance and relative orientation to the root feature. Because FAST does not produce the scale and orientation of detected keypoints, multi-scale oriented FAST detector [12] is employed to calculate them.

In the matching procedure, a score function to compare two local graphs is proposed as

$$s_i = \max_j \left(G_\theta \left(\theta_i - \theta_j \right) G_d \left(\left(d_i - d_j \right) / d_i \right) \right) \quad (1)$$

$$S = \sum_i s_i \quad (2)$$

In (1), i and j denote the node indices in each graph, G is Gaussian distribution function, and θ and d are relative orientation and normalized distance from the root feature, respectively. The score function computes the aggregation of the maximized similarity scores of all nodes. Since the proposed graph based descriptors are generated in feature domain, the topology of graph changes much if some features disappear due to large view change which depends on the characteristic of particular algorithms. To prevent dependency to a particular algorithm, we only aggregate the maximized similarity score of each node, excluding other small scores. Consequently, high score means high possibility of matching between descriptors.

(a) (b)

Fig. 3. Visual comparison of matching results between proposed algorithm and SURF for *graffiti* dataset. (a) Matching results of proposed algorithm. (b) Matching results of SURF.

In the matching procedure, we first find the several candidate features which are below specific threshold. Then, a pair of features with the highest similarity score is selected.

V. Experimental Results

The performance of the proposed descriptor is evaluated and compared with several other feature description algorithms. Mikolajczyk and Schmid's dataset [10] is used for the experiment, in which recall vs.1-precision plot is compared as proposed in [10]. The dataset includes different imaging conditions including in-plane and out-of-plane rotation, scale and illumination change, image blur, and JPEG compression. Corresponding features can be identified using the ground truth homography matrix. Therefore, it is possible to count the number of correct matching by reprojecting each feature onto another image using homography .

To show the robustness of the proposed algorithm under affine transformation, we compare the proposed descriptor with SIFT, SURF, ORB (Oriented FAST and Rotated BRIEF), and BRISK descriptors. The original implementations of those algorithms are provide by the authors or included in OpenCV. For evaluating recall vs. 1-precision graphs, we use *graffiti*, *bricks*, and *boat* datasets which have in-plane as well as out-of-plane rotation. Note that the proposed descriptor aims at robustness under out-of-plane rotation. We detect approximately 1,000 features in each image for all test algorithms and count the number of inlier and outlier using the ground truth homography matrix.

Fig. 3 shows the visual comparison of matching result between the proposed algorithm and SURF for the *graffiti* dateset, in which yellow and red lines indicate inlier and outlier, respectively. As shown in Fig. 3, the proposed algorithm has more inlier and less outlier over SURF algorithm.

Fig. 4 shows the recall vs. 1-precision comparison on the test datasets. In each algorithm, we change the distance threshold for the nearest neighbor matching to evaluate the performance. As illustrated in Fig. 4, the proposed algorithm outperforms other algorithms not only for the scene with out-of-plane rotation and large view changes (*graffiti* and *bricks*) but also for the scene with in-plane rotation (*boat*).

VI. Conclusion

In this paper, we proposed the robust feature description method using local graph matching which outperforms the conventional descriptors under out-of-plane rotation. In addition, we proposed an efficient score function for fast graph matching. Proposed descriptors are generated using neighboring features to increase the supporting region of each descriptor. The proposed descriptor is more robust under affine transformation and more stable than other descriptors in large view changes. Experimental results confirmed that the proposed algorithm has better matches than existing algorithms like SIFT and SURF.

References

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building Rome in a day. *Communications of the ACM*, 54(10):105–112, October 2011.

[2] M. Agrawal, K. Konolige, and M. R. Blas. CenSurE: Center surround extremas for realtime feature detection and matching. *Proc. of European Conference on Computer Vision*, pages 102–115, October 2008.

[3] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast retina keypoint. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517, June 2012.

[4] P. F. Alcantarilla, A. Bartoli, and A. J. Davision. KAZE features. *Proc. of European Conference on Computer Vision*, pages 214–227, October 2012.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[6] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary robust independent elementary features. *Proc. of European Conference on Computer Vision*, pages 778–792, September 2010.

[7] N. Karlsson, E. D. Bernardo, J. Ostrowski, L. Goncalves, P. Pirianian, and M. E. Munich. The vSLAM algorithm for robust localization and mapping. *Proc. of IEEE International Conference on Robotics and Automation*, pages 24–29, April 2005.

[8] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. *Proc. of IEEE International Conference on Computer Vision*, pages 2548–2555, November 2011.

[9] D. G. Low. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[10] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.

[11] E. Rosten, R. Porter, and T. Drummond. Faster and better: A machine learning approach to corner detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(1):105–119, January 2010.

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. *Proc. of IEEE International Conference on Computer Vision*, pages 2564–2571, November 2011.

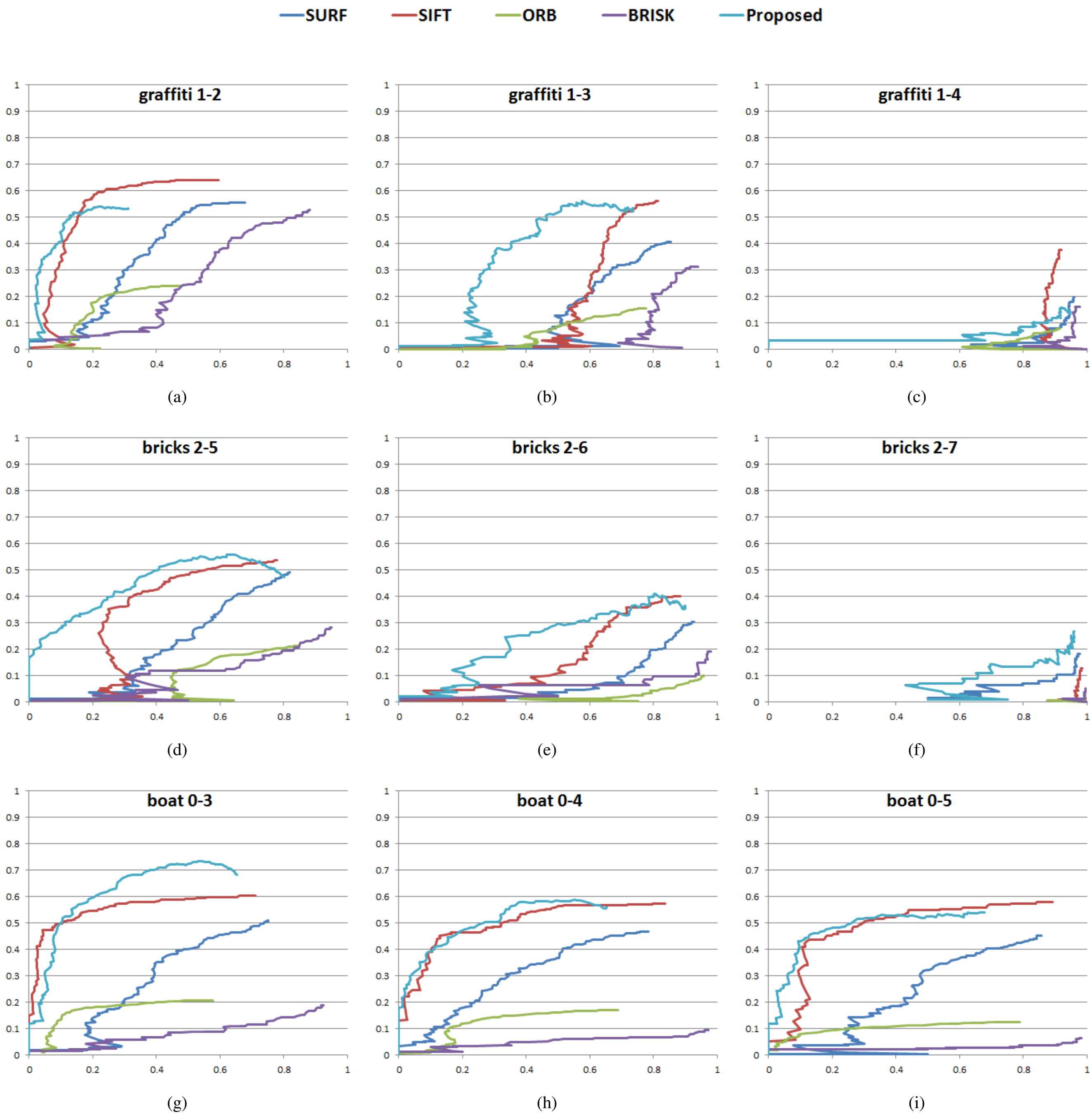

Fig. 4. Recall vs. 1-precision plots for each algorithm. (a)-(c) On *graffiti* dataset. (d)-(f) *bricks* dataset. (g)-(i) *boat* dataset.