Robustness Evaluation of Local Descriptors

Binquan Li[*] and Xiaohui Hu

Beijing University of Aeronautics and Astronautics, Beijing, 100191, China

[*]Corresponding author

Abstract—**The descriptors should be robust to changes of images taken under various conditions in order to obtain correct recognition. In this paper, we compare the robustness of descriptors computed for local interest regions, for example, extracted by the Harris, SURF and MSER detector. Different descriptors have been compared, depending on the interest region detector. Our evaluation uses different methods of matching features to assess image transformations. We compare SIFT, SURF, SURF with regions detected by MSER, FREAK for interest regions. The results show that local descriptors are sensitive to the viewpoint change, zoom and rotation, and image blur. Illumination has less effect on the robustness of image changing. SURF with regions detected by MSER performs better and is recommended for its better robustness.**

Keywords-local descriptor; robust; extracting features; image changing

I. INTRODUCTION

Local descriptors computed for interest regions are successful in applications of pattern recognition, such as object recognition of images taken by unmanned aerial vehicle (UAV). Given the region detectors, the remaining step is whether the descriptor can extract the features correctly and efficiently. Furthermore, the local descriptors should be robust to the changes of images obtained under different conditions in order to recognize the object correctly.

Performance evaluation has been more and more important in computer vision. The performance is measured by percentage of points in two images. The higher the repeatability between two images, the more points can be matched. The more matching points exist, the better matching and recognition perform. K. Mikolajczyk and C. Schmid compared the performance of local descriptors in [1], using recall and precision as the assessment criterion. They gave comparison for image changing under different conditions. They also gave measurement of affine region detectors in [2], in which MSER consistently resulted in the highest score through many tests, proving it to be a reliable region detector. MSER also performs better in text detection [3]. In [4], Lowe presented SIFT and gave a measurement to select the neighbor obtained by comparing the distance between the closest one and the second one. The SIFT-Rank [5] descriptor was shown to improve the performance of the standard SIFT descriptor for affine feature matching. SURF was introduced in [6] A Comparison of SIFT, PCA-SIFT and SURF was given in [7, 8]. By Histogram of Oriented Gradient descriptors, the local object appearance and shape within an image can be described by the distribution of intensity gradients [9-11] or edge directions. It's similar to SIFT, but differs in that it is computed on a dense grid of uniformly spaced cells.

The evaluation of descriptors is implemented under different viewing conditions for matching and recognition of the same scene. Descriptors selected have previously shown a good performance. In this paper, the comparison is carried out for different descriptors under changes of images, such as viewpoint change, light change, zoom and rotation, and image blur. We have selected a number of descriptors that have proven to be popular, and compare them by the same test data. The evaluation criterion is the number of matches between two images. It is a stable method when it gives a stable detector and matching numbers. This paper is organized as follows. Section 2 describes the implementation details for the detectors and descriptors used in our comparison as well as the data set. In Section 3, we present the experimental results. Finally, we discuss the results in Section 4.

II. EXPERIMENTAL SETUP

A. Data Set

Figure 1 shows example images of our data set used for the evaluation. Four image transformations are evaluated: viewpoint change (Figure 1a); illumination (Figure 1b); zoom and rotation (Figure 1c); image blur (Figure 1d). Viewpoint changes introduce a transformation that can be approximated by affine transformation. This is a challenging transformation evaluated in this paper. There are also some scale, rotation and brightness changes in the test images. The descriptors are evaluated for combination of image rotation and scale change. Scale changes are in the range of 2-2.5. The rotation angle is selected in the range between 30 and 45 degrees in order to evaluate the performance for image rotation. This represents the most difficult case. The significant amount of blur is introduced by changing the camera focus.

B. Matching Method

We select threshold matching method defining a match. In the case of threshold-based matching, two regions are matched if the distance between their descriptors is below a threshold. This method can return more than one match for each feature. A feature vector is matched to its nearest neighbor in the other feature set with nearest neighbor distance ratio, when the nearest neighbor satisfies a ratio test. The ratio test compares the distances from the feature vector to its first and second nearest neighbors in the other feature set. Nearest neighbor matching selects only the best match below the threshold and rejects all others. The nearest neighbor distance ratio is similar

but additionally penalizes the descriptors that have many similar matches, which improves the accuracy.
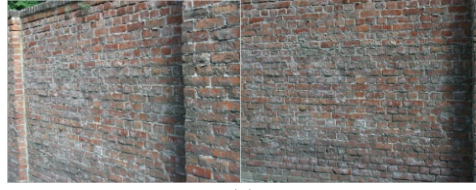
(a)

(b)

(c)

(d)

FIGURE I. DATA SET. EXAMPLES OF IMAGES USED FOR THE EVALUATION: (A) VIEWPOINT CHANGE, (B) LIGHT CHANGE, (C) ZOOM AND ROTATION, AND (D) IMAGE BLUR.EXPERIMENTAL RESULTS

In this section, we present and discuss the experimental results of the evaluation. The performance is compared for viewpoint changes, illumination, zoom and rotation and image blur.

C. Viewpoint Change

Table I shows the matching results of viewpoint change.

TABLE I. MATCHING RESULTS OF VIEWPOINT CHANGE

Descriptors	SURF		SURF with regions detected by MSER		Fast Retina Keypoint		SIFT	
Matching Strategies	SURF Points		MSER Regions		Corner Points		SIFT Points	
	2607	2088	2688	2580	6893	5256	2770	2710
Threshold	406		209		35		3	

D. Light Change

The matching results of light change are described as Table II.

TABLE II. MATCHING RESULTS OF LIGHT CHANGE

Descriptors	SURF		SURF with regions detected by MSER		Fast Retina Keypoint		SIFT	
Matching Strategies	SURF Points		MSER Regions		Corner Points		SIFT Points	
	1100	447	684	305	1341	752	1820	1690
Threshold	548		481		243		220	

The corner points are presented in Figure 2.

FIGURE II. CORNER POINTS OBTAINED BY FAST RETINA KEYPOINT

E. Zoom and Rotation

The matching results of zoom and rotation are calculated in Table III.

TABLE III. MATCHING RESULTS OF ZOOM AND ROTATION

Descriptors	SURF		SURF with regions detected by MSER		Fast Retina Keypoint		SIFT	
Matching Strategies	SURF Points		MSER Regions		Corner Points		SIFT Points	
	2791	1221	728	250	2476	984	9687	5376
Threshold	427		109		6		270	

The corner points are shown in Figure 3.

FIGURE III. CORNER POINTS OBTAINED BY FAST RETINA KEYPOINT

F. Image Blur

The matching results of image blur are described as Table IV.

TABLE IV. MATCHING RESULTS OF IMAGE BLUR

Descriptors	SURF		SURF with regions detected by MSER		Fast Retina Keypoint		SIFT	
Matching Strategies	SURF Points		MSER Regions		Corner Points		SIFT Points	
	1074	353	740	230	929	347	3553	2955
Threshold	83		82		38		479	

The regions detected by MSER are represented in Figure 4.

FIGURE IV. MSER REGIONS DETECTED BY MSER

The matching accuracy of different image changing by local descriptors is shown as Figure 5.
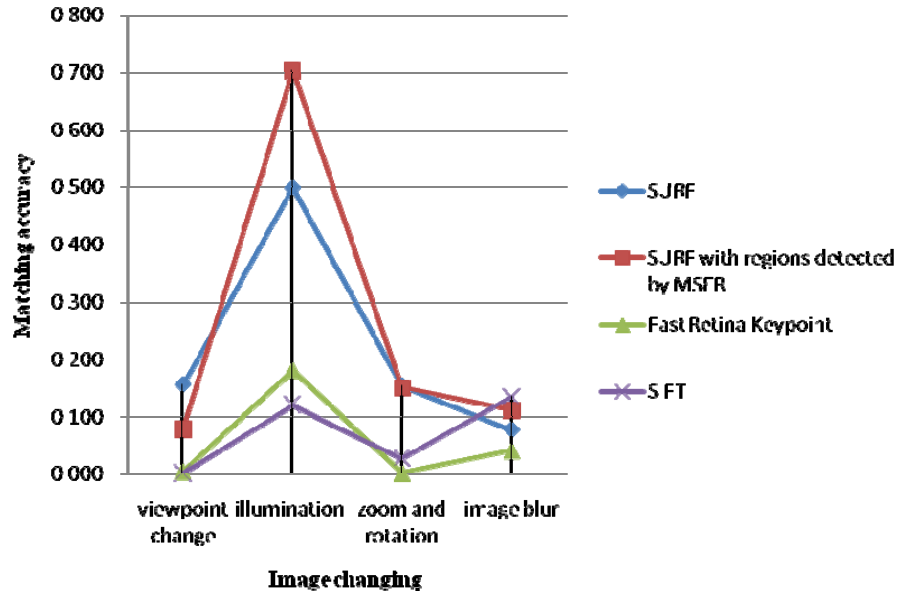

FIGURE V. MATCHING ACCURACY OF DIFFERENT IMAGE CHANGING BY LOCAL DESCRIPTORS

III. Discussion and Conclusions

In this paper, we have presented an evaluation of interest region descriptors under changes of images. The object was to compare the robustness of descriptors computed on regions with different detection techniques. The results show that viewpoint change, zoom and rotation, and image blur have great effect on the robustness of image changing. Local descriptors are less sensitive to the illumination. In the tests, SURF and SURF with regions detected by MSER obtains good results, and SURF with regions detected by MSER performs more precisely. SIFT shows its stability in the experiments and it performs common in viewpoint change. SIFT detects many key points and finds many matches, which is time consuming. We can choose a suitable algorithm according to the specific application.

References

[1] K. Mikolajzyk and C. Schmid, "A Perforance Evaluation of Local Descriptors," IEEE Trans. Pattern Anal. Mach. Intell. United States, vol. 27, no. 10, pp. 1615-1630, October 2005.

[2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A Comparison of Affine Region Detectors," Int. J. Comput. Vis. United States, vol. 65, no. (1-2), pp. 43-72, 2005.

[3] C. Z. Shi, C. H. Wang, B. H. Xiao, and S. Gao, "Scene Text Detection Using Graph Model Built Upon Maximally Stable Extremal Regions," Pattern Recognit. Lett. England, vol. 34, no. 2, pp. 107–116, January 2013.

[4] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," IJCV, vol. 60, no. 2, pp. 91–110, 2004.

[5] M. Toews and W. M. Wells, "SIFT-Rank: Ordinal Descriptors for Invariant Feature Correspondence," CVPR, pp.172–177, 2009.

[6] J. Luo and G. Oubong, "A Comparison of SIFT, PCA-SIFT and SURF," Int. J. Image Process., vol. 3, no. 4, pp. 143–152, 2009.

[7] P. M. Panchal, S. R. Panchal and S. K. Shah, "A Comparison of SIFT and SURF,", Int. J. Inno. Res. Comput. Commun. Eng., vol. 1, no. 2, April 2013.

[8] B. Herbert, E. Andreas, T. Tinne, and V. G. Luc, "SURF: Speeded Up Robust Features," Comput. Vis. Image Underst. United States, vol. 110, no. 3, pp. 346–359, 2008.

[9] R. Hu, M. Barnard and J. Collomosse, "Gradient Field Descriptor for Sketch based Image Retrieval and Localisation," Int. Conf. Image Process., vol. 119, no. 5, pp. 1025-1028, September 2010.

[10] R. Hu and J. Collomosse, "Performance Evaluation of the Gradient Field HOG Descriptor for Sketch based Image Retrieval," Comput. Vis. Image Underst. United States,, vol. 117, no. 7, pp. 790–806, July 2013.

[11] X. Y. Wang, T. X. Han, and S. C. Yan, "An HOG-LBP Human Detector with Partial Occlusion Handling," ICCV, vol. 30, no. 2, pp. 32–39, 2009.