

LOCAL INVARIANT DESCRIPTOR FOR IMAGE MATCHING

Lei Qin^{1,3}, Wei Zeng², Wen Gao^{1,2,3}, Weiqiang Wang^{1,3}

¹ Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

² Department of Computer Science and Technology, Harbin Institute of Technology, China

³ Graduate School of the Chinese Academy of Sciences, Beijing, China

Email: {lqin, wzeng, wgao, wqwang}@jdl.ac.cn

ABSTRACT

Image matching is a fundamental task of many computer vision problems. In this paper we present a novel approach for matching two images in the presence of image rotation, scale, and illumination changes. The proposed approach is based on local invariant features. A two-step process detects local invariant regions. Characteristic circles associated with these regions illustrate the position and radius of the regions. Then, the regions are represented by a new image descriptor. To test the new descriptor, we evaluate it in image matching and retrieval experiments. The experimental results show that using our descriptors results in effective and faster matching.

1. INTRODUCTION

Local invariant features have been widely used in many applications, such as image matching, image retrieval, object recognition, scene reconstruction, etc [1, 5, 6, 7, 8, 11, 2, 13, 9, 10]. Local features are robust to partial occlusion, resistant to nearby clutter and can be computed efficiently. There are two considerations that should be taken in the usage of local features. The first is the selection of sparse salient image patches for subsequent computing. The second is the description of the patches. In this paper, we introduce an approach to detect local invariant region and propose a new local invariant image descriptor to represent the regions.

A number of techniques for representing local image patches have been reported in the literatures [3, 11]. Recently, Yan Ke and Rahul Sukthankar [3] use the Principal Component Analysis (PCA) to reduce the dimension of image patches. They demonstrate their method is robust and fast in the image retrieval application. They also show that the proposed compact descriptor increases the matching precision and speed. Johnson and Hebert [11] introduce an expressive descriptor *spin images* for matching range data. Their

representation is generated using a histogram of the relative position of neighborhood points to the interesting point in 3D space. The above two representations are appearance-based descriptors.

The other class of representation is feature-based descriptor such as differential descriptors [5], complex filters [6], moment invariants [14], SIFT [2, 13]. The differential descriptors are a set of image derivatives calculated up to a given order. Mikolajczyk and Schmid [5] use the differential descriptors to approximate a point neighborhood for image matching and retrieval. Schaffalitzky and Zisserman [6] introduce the complex filters. The complex filters are orthogonal, and the Euclidean distance between complex filters provides a lower bound on the Squared Sum Differences (SSD) between corresponding image patches. Van Gool [14] introduces the Generalized Color Moments to exploit the multi-spectral nature of the data. The moments describe the shape and the intensities of different color channels in a local region. Lowe [2, 13] proposes a distinctive local descriptor, scale-invariant (SIFT) features, which is computed by sampling the magnitudes and orientations of local image gradients and building smoothed orientation histograms. This description provides robustness against localization errors and small geometric distortions.

Mikolajczyk and Schmid [15] report an experimental evaluation of several different descriptors. In their experiments, SIFT descriptors obtain the best matching results. However, the SIFT descriptor is high dimensional. Thus, it is computationally expensive and not fit to the purpose of real-time applications, such as image matching and image retrieval. In order to overcome this problem, we propose a novel descriptor based on SIFT. The new descriptor is computationally efficient, and still sufficiently discriminative for successful correspondence.

The paper is organized as follows. In section 2, we introduce the Harris-Laplace detector. In section 3, our new invariant descriptor is presented and section 4 describes the robust matching algorithm. The experimental results for image matching and retrieval are given in section 5.

2. INVARIANT REGION DETECTOR

The implementation of image matching by local invariant features requires detecting image regions, which are invariant under rotation and scale transformations of the image. An algorithm is used to achieve the invariant regions $(x, y, scale, alpha)$: 1) locating interesting points (x, y) , 2) associating a characteristic scale to each interesting point $(scale)$, 3) assigning the region orientation $(alpha)$.

Interesting points. Our interesting points are multi-scale Harris corners of the images in scale-space. Harris corners are chosen for its high repeatability in the presence of image rotation, illumination changes, and perspective transformations [18]. However the repeatability of Harris detector degrades significantly when the images have large-scale changes. In order to cope with such changes, a multi-scale Harris detector is presented in [16].

$$C(x, \sigma_I, \sigma_D) = \sigma_D^2 G(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix}$$

Where σ_I is the integration scale, σ_D the derivation scale, $L(x, \sigma) = G(\sigma) * I(x)$ different levels of resolutions created by convoluting the Gaussian kernel $G(\sigma)$ with the image $I(x)$ and $x = (x, y)$. Given an image $I(x)$ the derivatives can be defined by $L_x(x, \sigma) = \frac{\partial}{\partial x} G(\sigma) * I(x)$. The multi-scale Harris detector is used to locate interesting points at different resolution $L(x, \sigma_n)$, with $\sigma_n = k^n \sigma_0$, σ_0 is the initial scale factor.

Characteristic scale. For an interesting point, the characteristic scale can be defined as that at which the result of a differential operator is maximized [17]. Different differential operators are comparatively evaluated in [8]. Laplacian obtains the highest percentage of correct scale detection. We use Laplacian to verify for each of the candidate points found on different levels if it forms a local maximum in the scale direction.

$$Lap(x, \sigma_n) > Lap(x, \sigma_{n-1}) \wedge Lap(x, \sigma_n) > Lap(x, \sigma_{n+1})$$

$$Lap(x, \sigma_n) > threshold$$

Orientation. The invariance to rotation can be obtained by assigning a consistent orientation to each local region. One can use the method in [4] to get a stable estimation of the dominant direction.

Fig.1 shows two images with detected scale invariant regions. The threshold of multi-scale Harris and Laplacian are 1000 and 10, respectively. 15 resolution levels are used for scale representation. The factor k is 1.2 and $\sigma_I = 2\sigma_D$.

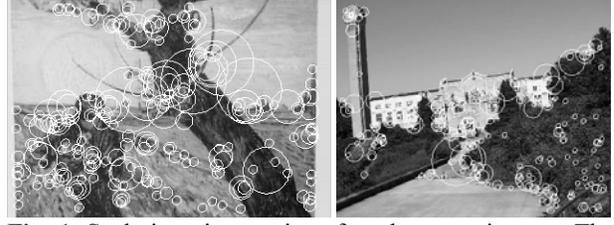


Fig. 1. Scale invariant regions found on two images. The circle center is the interesting point's location and the radius represent it's scale. There are 215 and 247 points detected in the left and right images, respectively

3. INVARIANT REGION DESCRIPTOR

Our invariant descriptor is motivated by the SIFT descriptor [13] which is based on the image gradients in each interesting point's local region. The descriptor is calculated by sampling the magnitudes and orientations of the image gradient in the local region around the interesting point and then building smoothed orientation histograms. The local region is divided in 4×4 sub-regions, each with 8 orientations, which results in a descriptor of dimension $128 = 4 \times 4 \times 8$. The 128-dimensional descriptor is normalized to unit length to eliminate the effects of illumination change.

Given 128-dimensional descriptors, we employ Principal Components Analysis (PCA) [12] to project them to low dimensional space. The number of Principal Components (n) is empirically selected. Here we use $n=20$, which get comparable results with the 128-dimensional descriptor. The dimension of our descriptor is low, which results in significant space and speed benefits.

4. ROBUST MATCHING

To robustly match a pair of images, we first determine point-to-point correspondence. We select for each descriptor in the first image the most similar one in the second image based on the Euclidean distance. If the Euclidean distance is below a threshold η , the correspondence is kept. All point-to-point correspondences form a set of initial matches. We refine the initial matches using RANdom SAMple Consensus (RANSAC). RANSAC has the advantage that it is largely insensitive to outliers. We use fundamental matrix as the transformation of RANSAC in our experiments.

5. EXPERIMENTS

We validate our algorithm by image matching experiments and an image retrieval application.

5.1. Image matching experiments

Fig. 2 shows the matching result for a real world scene, which includes significant scale, rotation change as well as a change in the viewing angle.

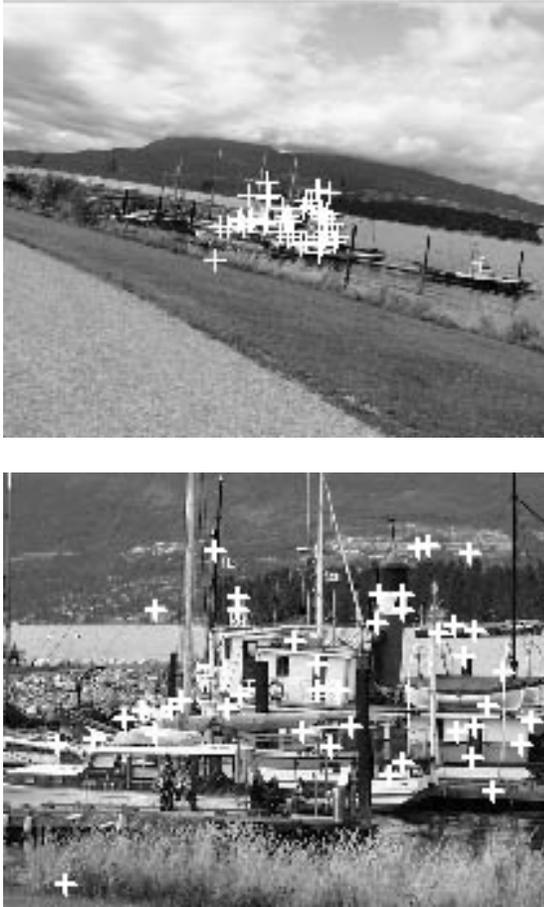


Fig. 2. Robust matching result by our algorithm. There are 46 inliers, all of them are correct. The rotation angle is 11degrees, and the approximate scale factor is 3.9.

In the following, we present the comparative evaluation result of our descriptor, SIFT and Cross-correlation on image matching tasks with geometric transformation, viewing angle change and significant intensity variation. The performance of descriptors is evaluated using the recall-precision criteria (obtained by varying Euclidean distance threshold η). Fig. 3 plots the recall-precision curve of an experiment on images with scale and rotation changes. The scale factor is 2 and the rotation angle is 45° . Fig. 4 shows the experimental results where the target images are distorted to simulate a 30° viewpoint change. While Fig. 5 gives the matching results when the intensity of target images is reduced 50%. Our approach performs significantly better than Cross-correlation but slightly worse than the SIFT.

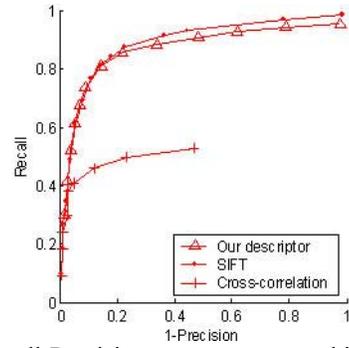


Fig. 3. Recall-Precision curve on a matching task where scale change is factor 2 and image rotation is 45 degrees

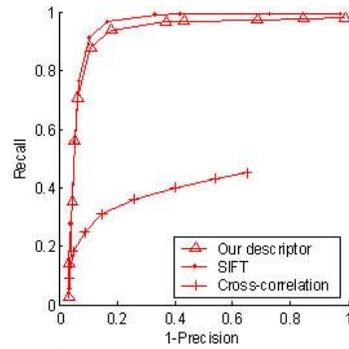


Fig. 4. Recall-Precision curve on a matching task where viewpoint change is 30 degree

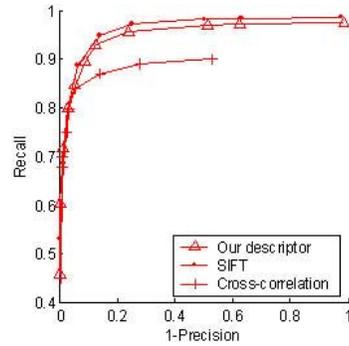


Fig. 5. Recall-Precision curve on a matching task where intensity reduction is 50%

We compare the matching time between the SIFT and our method. It takes 4.113 seconds for the SIFT to match two images, while 2.240 seconds for our method. The matching time is the mean value of matching 60 pairs of images. This comparison shows our method is significantly faster than the SIFT.

5.2. Image retrieval experiments

We evaluate the performance of our method in an image retrieval application. The experiments have been

conducted for a small image database containing 400 images. In the retrieval application, we first extract our descriptors of each image in the image database. Each descriptor of a query image is compared against all descriptors in the other image. If the distance of the two descriptors is below a threshold, they are accepted as a match. We regard the number of matched interesting points as a similarity measure between images.

Fig. 6 shows the results of image retrieval experiments. The first column displays five query images. The second column shows the corresponding image in the database, which is the most similar one. The changes between the image pairs (first and second column) include scale and rotation changes, for example pairs in Fig. 6(a) and Fig. 6(b). They also include small viewpoint variations, such as image pairs in Fig. 6(c) and Fig. 6(d). Furthermore, they include significant intensity changes (image pairs Fig. 6(e)). The retrieval results show the robustness of our approach to image rotation, scale changes, viewpoint variations and intensity changes.

6. CONCLUSION

In this paper we present a novel method to matching two images in the presence of rotation and scale transformation, viewing angle change and significant intensity variation. We propose a new local invariant image descriptor. Our descriptor is compact, yet still distinctive, and robust to scale and viewpoint changes.

Acknowledgements

This work is supported by National Hi-Tech Development Programs of China under grant No. 2003AA142140.

References

- Schmid, C., Mohr, R.: Local Grayvalue Invariants for Image Retrieval. *IEEE PAMI*, 19 (1997) 530–534
- Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: *ICCV*, (1999) 1150–1157
- Y. Ke and R. Sukthankar: "PCA-SIFT: A more distinctive representation for a local image descriptors", *CVPR04*.
- Mikolajczyk, K.: Detection of Local Features Invariant to Affine Transformations. PhD thesis, INRIA, (2002)
- Mikolajczyk, K., Schmid, C.: An Affine Invariant Interest Point Detector. In: *ECCV*, (2002) 128-142
- Schaffalitzky, F., Zisserman, A.: Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?". In: *ECCV*, (2002) 414-431
- Tuytelaars, T., Van Gool, L.: Wide Baseline Stereo Matching Based on Local Affinely Invariant Regions. In: *BMVC*, (2000) 412-425
- Mikolajczyk, K., Schmid, C.: Indexing Based on Scale Invariant Interest Points. In: *ICCV*, (2001) 525–531
- Fergus, Perona, Zisserman. Object class recognition by unsupervised scale invariant learning. In: *CVPR* 2003.

- R.Hartley and A.Zisserman. Multiple view geometry in computer vision. Cambridge University Press,2000.
- Andrew E. Johnson and Martial Hebert. "Using Spin-Images for efficient multiple model recognition in cluttered 3-D scenes." *IEEE PAMI*, 21(1999), 433-449.
- R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2nd ed., New York: John Wiley and Sons, 2001.
- Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91-110, 2004.
- L. Van gool, T. Moons, and D. Ungureanu. Affine photometric invariants for planar intensity patterns. In *ECCV*, pp. 642-651, 1996.
- Mikolajczyk, K., Schmid. A performance evaluation of local descriptors. In *CVPR* June 2003.
- Dufournaud, Y., Schmid, C., Horaud, R.: Matching Images with Different Resolutions. In: *CVPR*, (2000) 621-618
- Lindeberg, T.: Feature detection with automatic scale selection. *IJCV*, 30 (1998) 79–116
- Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of Interesting Point Detectors. *IJCV*, 37 (2000) 151-172

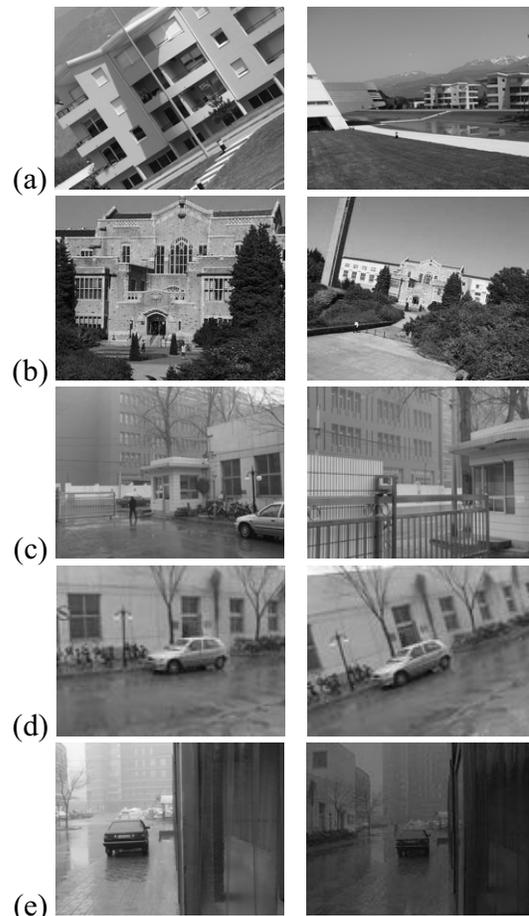


Fig. 6. The first column shows some of the query images. The second column shows the most similar images in the database, all of them are correct