

# Image Matching Based on A Local Invariant Descriptor

Lei Qin

Institute of Computing Technology  
Chinese Academy of Sciences  
Beijing, China  
Email: lqin@jdl.ac.cn

Wen Gao

Institute of Computing Technology  
Chinese Academy of Sciences  
Beijing, China  
Email: wgao@jdl.ac.cn

**Abstract**—Image matching is a fundamental task of many computer vision problems. In this paper we present a novel approach to match two images in presenting significant geometric deformations and considerable photometric variations. The approach is based on local invariant features. First, local invariant regions are detected by a three-step process which determines the positions, scales and orientations of the regions. Then each region is represented by a novel descriptor. The descriptor is a two-dimensional histogram. Performance evaluations show that this new descriptor generally provides higher distinctiveness and robustness to image deformations. We present the image matching results. The matching results show good performance of our approach for both geometric deformations and photometric variations.

## I. INTRODUCTION

Image matching is to estimate correspondences between two views of the same scene, taken with different viewpoints, orientations or focal lengths. Image matching is a crucial step in many image processing tasks. For example, the technique has been applied in image mosaic, target localization, automatic quality control, structure from motion. But it is still a challenging task, since the content of images to be matched may have a wide range of different appearances, and may be taken under different imaging conditions (lighting, scale, rotation, and viewpoints). Even partial occlusion by moving objects (pedestrian, vehicles) can exist.

Recent research shows local information is effective to describe image content [1], and well adapted to image matching, as they are robust to clutter and occlusion and don't require segmentation [2], [3], [4], [8]. Schmid and Mohr [1] demonstrate that local information is sufficient for general image recognition under occlusion and clutter. They use Harris corners [7] as interesting points, and extract rotationally invariant descriptors from the patches around interesting points. The descriptor guarantees that the rotated images can be well matched. Lowe [8] uses the local extrema of DoG (difference of Gaussian) in scale-space as the interesting points. He proposes a distinctive local descriptor, which is computed by accumulating local image gradients in orientation histograms. Tuytelaars and Van Gool [4] construct small affine invariant regions around corners and intensity extrema. Their method looks for a specific structure "parallelogram" in images. Among the above methods, [1] and [8] are rotation and

scale invariant. [4] is affine invariant.

Many methods presented are intended for wide-baseline matching [3], [5], [6]. Baumberg [6] uses the multi-scale Harris feature detector, and orders the features based on the scale-normalized feature strength. The neighborhoods of feature points are normalized using an iterative procedure based on isotropy of the second gradient moment matrix. Mikolajczyk *et al* [5] propose the Harris-Laplace method. They detect the Harris corners in multiple scales, then select points at which the Laplacian measure attained the local extrema in scale dimension. They extend their work to affine invariant in [2]. Schaffalitzky and Zisserman [3] present a method for obtaining multi-view matching given unordered image sets. They use two kinds of features: invariant neighbourhoods and "Maximally Stable Extremal" regions [12]. They use the complex filters as descriptor.

In this paper we want to solve the problem of matching images in the presence significant geometric deformations and considerable photometric variations. We introduce an approach to detect local invariant regions and propose a novel descriptor called the Angular Radial Partitioning Intensity Histogram (ARPIH). Our method is based on representing the images by a set of scale and rotation independent feature descriptors. Features are extracted through two stages: first, local regions which transform invariantly with scale and rotation are detected in each image; second, a distinctive and invariant descriptor is computed for each region. The descriptors are used to match regions between images.

The paper is organized as follows. Section 2 presents the novel descriptor. Section 3 provides experimental results comparing the novel descriptor to steerable filters [2] on feature matching experiments. Section 4 presents the robust image matching strategy and the image matching results. Section 5 draws some conclusions.

## II. INVARIANT DESCRIPTOR

### A. Local invariant region

The task of image matching by local information requires detecting local image regions, which are invariant under scale and rotation transformations of the image. A three-step algorithm [5] is used to achieve the invariant regions  $(x, y, scale, alpha)$ : 1) locating interesting points  $(x, y)$ ; 2)

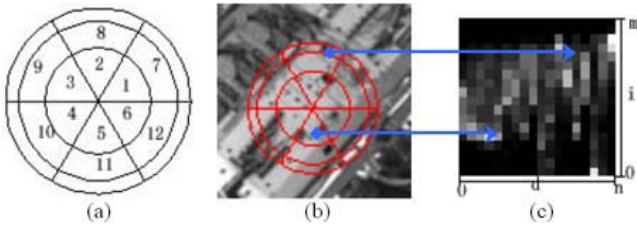


Fig. 1. ARPIH descriptor construction. (a) An image region is divided into 18 sub-regions. We use three bins for  $\rho(0.57, 0.85, 1)$  and six bins for  $\theta(1/3\pi, 2/3\pi, \pi, 4/3\pi, 5/3\pi, 2\pi)$ . The number in the sub-region is its index. The indexes of the third ring aren't shown for clarity. Two sub-regions in an image patch (b) map to two different columns in the ARPIH descriptor (c). There are 18 bins in the intensity portion

associating a characteristic scale to each interesting point (*scale*); 3) assigning an orientation for each invariant region (*alpha*) [8].

### B. ARPIH Descriptor

Given an invariant image region, we compute a novel descriptor. The proposed descriptor is the Angular Radial Partitioning Intensity Histogram (ARPIH). An ARPIH descriptor is a two-dimensional histogram encoding the intensity distribution in an image region, and the geometry relation between the sub-regions. One axis of the histogram is an index into the region of space subdivided radially around the interesting point. The second axis is based on the image intensity. The index of sub-regions is sorted in ascending order both in  $\rho$  and  $\theta$  as shown in Fig.1(a). We use bins that are not uniform in  $\rho$ , which are selected by experiments. Each column of the ARPIH descriptor (Fig.1(c)) is the intensity histogram of pixels in the corresponding sub-region. An example is shown in Fig.1. The invariance to affine illumination changes (changes of the form  $I \rightarrow aI + b$ ) is achieved by normalized the range of the intensity within the local region [13].

While Belongie et al. [9] also use the angular radial partitioning histogram, ARPIH differs in that it combines the radius and angle into an index into the subregion. Belongie et al. map a subregion to a bin, and store the number of edge points of the subregion in the bin; we map a subregion to a "slice" of ARPIH, and store the intensity histogram of the subregion in the "slice".

### C. Comparison of ARPIH Descriptor

Let  $P_i$  and  $Q_j$  represent two image regions.  $COST_{ij} = COST(P_i, Q_j)$  denotes the cost of matching these two regions. We use  $\chi^2$  test statistics [9].

$$COST_{ij} = \frac{1}{2} \sum_{n=0}^N \sum_{m=0}^M \frac{[h_i(n, m) - h_j(n, m)]^2}{h_i(n, m) + h_j(n, m)} \quad (1)$$

where  $h_i(n, m)$  and  $h_j(n, m)$  denote the values of the ARPIH descriptor at  $P_i$  and  $Q_j$ , respectively.

## III. EVALUATION UNDER CONTROLLED TRANSFORMATIONS

In this section we compare the performance of ARPIH descriptor and steerable filters [2]. In a recent evaluation [11],

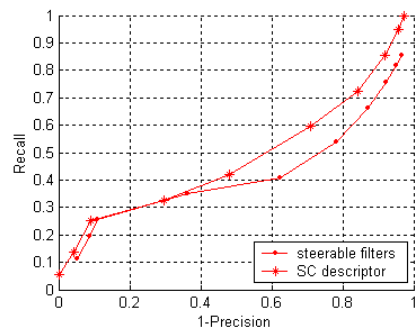


Fig. 2. Recall VS. 1-Precision curve on a matching experiment where target images are rotated by  $45^\circ$  and scaled down by 2 times.

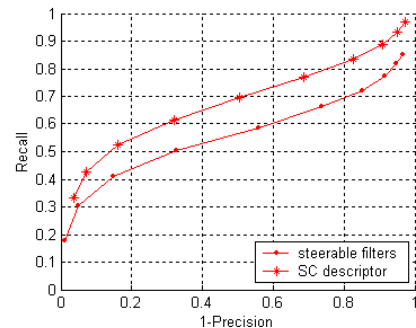


Fig. 3. Recall VS. 1-Precision curve on a matching experiment where target images are distort to simulate a  $30^\circ$  viewpoint changes.

steerable filters had obtained better performance than those of several filter-based descriptors. Thus they can represent the "state of art" of currently used filter-based descriptors.

**Transformations.** Three transformations are examined :

1. Rotate images by  $45^\circ$  and scale the size of images down by 2 times.
2. Distort images to simulate  $30^\circ$  viewpoint change.
3. Change image illumination amplitude by 50%.

**Evaluation criterion.** Invariant descriptors are extracted from the original image and transformed image, respectively. Each descriptor from the original image is compared with all descriptors from the transformed image. If the distance between a particular pair of descriptors is below the threshold  $t$ , this pair is matched. The performance of descriptors is evaluated using the Recall VS. 1-Precision criteria (obtained by varying distance threshold  $t$ ).

**Evaluation results.** Fig.2 plots the Recall VS. 1-Precision curve of the matching experiment with scale and rotation changes. The results show descriptor is better at handling rotation and scale transformation for almost all values of 1-Precision. Fig. 3 shows the experimental results of viewpoint changes. The results show that ARPIH descriptor outperforms steerable filters on all domain of 1-Precision. The Recall is about 10% better than steerable filters. While Fig. 4 gives the matching results when the image intensity changes. Both descriptors are well-suited to represent simple illumination changes.

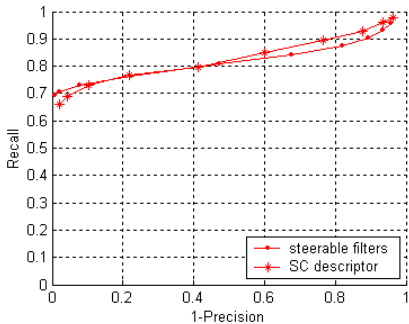


Fig. 4. Recall VS. 1-Precision curve on a matching experiment where the illumination of target images is reduced 50%

## IV. ROBUST IMAGE MATCHING

### A. Robust Matching Strategy

To robustly match a pair of images, we first determine point-to-point correspondences. We extract invariant descriptors from the two images, and select the most similar descriptor in the second image for each one in the first image based on the cost function Eq. (1). If the cost is below the threshold the correspondence is kept. All point-to-point correspondences form a set of initial matches.

The robust estimator RANSAC is used to refine the initial matches, and estimate the fundamental matrix  $F$ . After robust fundamental matrix estimation, we obtain the epipolar geometry of the image pair. The epipolar geometry can be used to predict positions for new matches. We use this prediction to increase matches. Let  $p_a$  represents an interesting point on the first image, and  $l_a = Fp_a$  represents the epipolar line of  $p_a$  on the second image. For each interesting point  $p_b$  on the second image, we compute the distance  $d_{ab}$  of point  $p_b$  to the epipolar line  $l_a$ , such that

$$d_{ab} = \sqrt{(p_b^T F p_a)^2 / ((F p_a)_1^2 + (F p_a)_2^2)} \quad (2)$$

We compute distance  $d_{ba}$  similarly. If the distance  $\max(d_{ab}, d_{ba})$  is below a threshold, these two points are considered to be matched. Since we use the robustly estimated  $F$  to guide the search for new matches, the number of matches increases efficiently.

### B. Matching Results

In this section, we evaluate our robust matching strategy on VGG's Datasets [10]. Five image transformations are examined: (1) scale and rotation changes; (2) viewpoint changes; (3) image blur; (4) JPEG compression; (5) illumination changes. These transformations include both significant geometric deformations and considerable photometric variations.

Fig. 5 shows the matching results for scale and rotation changes. The rotation angle is  $10^\circ$  and scale factor is approximately 2.4. There are 47 inliers, all are correct. Fig. 6 shows the matching results for viewpoint changes. In this transformation, we use two types of scenes, structured scene and texture scene. Matching results show our method is effective in both structured scene and texture scene. The first



Fig. 5. Example of scale and rotation changes. The two images are frames of "boat" from Oxford Univ. There are considerable rotation and scale changes between them. There are 47 matches. All are correct.

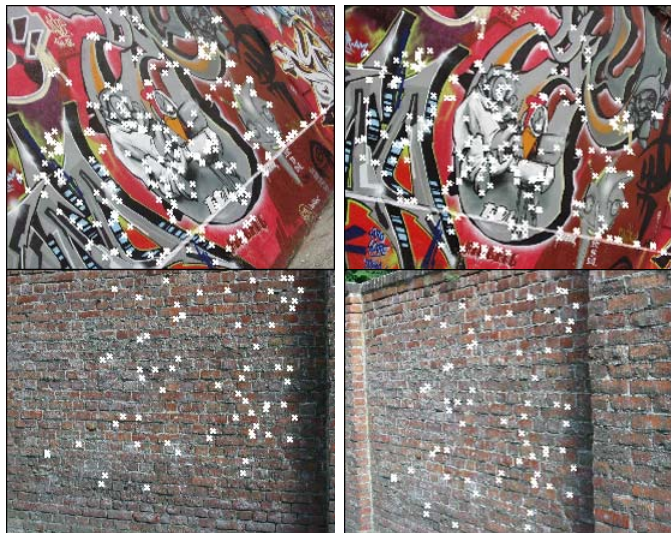


Fig. 6. Examples of viewpoint changes. The first row is structured scene, which shows 270 inliers. The second row is texture scene, which shows 68 inliers. All of them are correct.

row of Fig. 6 shows the structured scene. There are 270 inliers. The second row of Fig. 6 shows the texture scene. There are 68 inliers. All are correct.

We examine three types of photometric variations: blur, jpeg compression and intensity changes. Top row of Fig. 7 illustrates the case of image blur. The blur is visible. Our algorithm gets 79 inliers. We construct scale-space when detect local regions, so the effectiveness to blur is intrinsic to our algorithm. The experiment results validate this. The second row of Fig. 7 shows the matching results between two images with JPEG compression. Due to high compression rate, the block effect is very significant. Most of the details are lost. However, our algorithm obtains 102 inliers, all are correct. The results show our approach is very robust to JPEG compression. The third row of Fig. 7 shows the matching results when illumination changes. The illumination variation is evident on the windows and wall. There are 68 inliers, which shows our approach can handle simple illumination changes.

## V. CONCLUSION

In this paper we propose an approach to solve the problem of obtaining reliable correspondences between two images in



Fig. 7. Examples of photometric variations. Top row shows the matching results of image blur. There are 79 inliers. The second row shows the matching results of jpeg compression. There are 102 inliers. The third row shows the matching results of intensity changes. There are 68 inliers. All are correct.

presenting significant geometric deformations and considerable photometric variations. This is important because reliable correspondences are the basis of many computer vision applications. We present a novel region descriptor, which is based on the intensity distribution. The comparative evaluations show that this descriptor does well or better than the steerable filters. The robust image matching experiments show our approach is effectiveness to significant geometrical transformation as well as robust to photometric variations.

## VI. ACKNOWLEDGEMENT

This work is supported by National Hi-Tech Development Programs of China under grant No. 2003AA142140. We are grateful to Dr. WeiQiang Wang and Dr. Wei Zeng for helpful discussions, to Dr. K. Mikolajczyk for providing the image datasets, to anonymous reviewers for helpful advice.

## REFERENCES

- [1] C. Schmid, R. Mohr: Local Grayvalue Invariants for Image Retrieval. *IEEE PAMI*, 19 (1997) 530–534
- [2] K. Mikolajczyk, C. Schmid: An Affine Invariant Interest Point Detector. In: *ECCV*, (2002) 128–142
- [3] F. Schaffalitzky, A. Zisserman: Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?". In: *ECCV*, (2002) 414–431
- [4] T. Tuytelaars, L. Van Gool: Wide Baseline Stereo Matching Based on Local Affinely Invariant Regions. In: *BMVC*, (2000) 412–425
- [5] K. Mikolajczyk, C. Schmid: Indexing Based on Scale Invariant Interest Points. In: *ICCV*, (2001) 525–531
- [6] A. Baumberg: Reliable Feature Matching across Widely Separated Views. In: *CVPR*, (2000) 774–781
- [7] C. Harris, M. Stephens: A Combined Corner and Edge Detector. In: *Proc. Alvey Vision Conf.*, Manchester (1988) 189–192
- [8] D. G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60, 2 (2004), pp. 91–110.
- [9] S. Belongie, J. Malik, J. Puzicha: Shape Matching and Object Recognition Using Shape Contexts. *IEEE PAMI*, 24 (2002) 509–522
- [10] <http://www.robots.ox.ac.uk/~vgg/data5.html>.
- [11] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *CVPR*, June 2003.
- [12] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In: *BMVC* 384–393, 2002.
- [13] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In: *ICCV*, 636–643, 2001.