Matching Interest Points Using Affine Invariant Concentric Circles

Han-Pang Chiu   Tomas Lozano-Perez
CSAIL, Massachusetts Institute of Technology
{chiu, tlp}@csail.mit.edu

Abstract

We present a new method to perform reliable matching between different images. This method finds complete region correspondences between concentric circles and the corresponding projected ellipses centered on interest points. It matches interest points exploiting all the available luminance information in the regions under affine transformation. Experiments have been conducted on many different data sets to compare our approach to two SIFT-based local descriptors. The results show the new method is more effective in natural scenes without distinctive texture patterns. It also offers increased robustness to partial visibility, object rotation in depth, and viewpoint angle change.

1. Introduction

Image matching is an essential aspect of many approaches to problems in computer vision, including object recognition, stereo matching, and motion tracking. A prominent approach to image matching has consisted of identifying "interest points" in the images, finding photometric descriptors of the regions surrounding these points and then matching these descriptors across images. The development of these methods has focused on finding interest points and local image descriptors that are invariant to changes in image formation, in particular, invariant to affine transformations of the image regions. The idea is to compute local descriptors from constructed "affine invariant image regions" [5, 6] around interest points for matching. Then they compute a cost of the match based on the similarity of the local descriptors sampled from small image patches around them.

Many local descriptors such as steerable filters [7], moment invariants [8], and SIFT-based descriptors [3, 4] have been developed. Image matching based on affine-invariant interest point detectors and local photometric descriptors has been shown to work well in the presence of changes in viewpoint, partial visibility and extraneous features. However, not surprisingly, the accuracy of matching decreases with substantial changes in viewpoint and also when the regions around the interest points lack distinctive textures.

So systems that rely on matching interest points across images using local descriptors typically employ some set of additional filters to "verify" the putative matches based on local descriptors. A good example of this can be found in Schaffalitzky and Zisserman's [9] method for multi-view matching, which involves first increasing the correlation neighborhood size of a putative match, followed by intensity based affine registration, followed by growing using affine registration and followed, finally, a robust fit for the epipolar geometry. The need for better descriptors to reduce false matches is also evident in Mikolajczyk and Schmid's comparative study [1], where they found the ranking of accuracy in different methods is relatively insensitive to the choice of interest point detector but more dependent on the representation used to model the regions around interest points.

In this paper, we explore an alternative method for matching interest points that does not rely on local photometric descriptors but, instead, involves building complete region correspondences centered on interest points under an affine transformation and that exploits all the available luminance information in the regions. Our proposed method is a computational compromise between the local descriptor comparison and full region registration methods that search over transformation parameters. We find it performs better than matching based on the popular SIFT-based descriptors [3, 4], particularly for more natural images than the highly textured images normally used to benchmark interest point detection and feature descriptors. It is also more robust to variations in viewpoint and allows for a more powerful handling of occlusion.

2. Concentric Circles and Ellipses

Concretely, we assume that a set of "interest points" has been identified in each image and that these image points are likely to correspond to different views of the same points in the scene. The problem is then to decide which pairs of interest points in fact correspond to the same scene point.

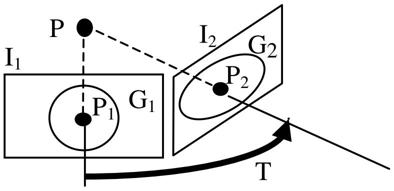

Figure 1. Characterizing regions for image matching.

This situation is illustrated in Figure 1, which shows one image point P_1 in an initial (reference) image I_1 and another image point P_2 in a second (transformed) image I_2. We want to compute a cost of the match of P_1 to P_2 based on the similarity of the colors in the image patches around them. We will assume that P_1 and P_2 correspond to point P in the scene, which is on a 3D planar surface in the scene. We want to compute a cost that is insensitive to the viewpoint change (T) between the images, approximated by an affine transformation. In our approach, we will define the region G_1 to be a set of concentric circles around p_1 and attempt to find a set of concentric ellipses around p_2 that minimize the image differences and that satisfy the conditions required for an affine transformation. This set of ellipses defines G_2.

Let us now focus on the relationship between concentric circles and ellipses under affine transformation. We know a perspective transformation of a smooth surface can be locally approximated by an affine transformation. Under a pure affine transformation, the projections of concentric circles will be ellipses with the same center, which is the center point of the projected concentric circles. An affine transformation has six degrees of freedom including the 2 by 2 affine matrix A and the offset two-component vector t. Assume x_1 and x_2 are image coordinates of matched points on the circle in I_1 and the projected ellipse in I_2 respectively. It can easily be shown that this following form represents an affine transformation and the transformed circles are co-incident.

$$x_2 = Ax_1 + t$$

Since a circle in I_1 should be an ellipse in I_2 under any transformation, we define that the origins of the polar coordinates in I_1 and I_2 are P_1 and P_2 in Fig. 1 respectively, thus the two translation degrees of freedom (represented by vector t) can be fixed. The point x_1 on the circle with radius R and angle θ_1 in I_1 has the form ($R\cos\theta_1$, $R\sin\theta_1$) and each point in the polar coordinate system of I_2 has the form ($r\cos\theta_2$, $r\sin\theta_2$). So the final equation to generate the projection of concentric circles in I_2 becomes.

$$[\mathrm{r}\cdot\cos\theta_2 \quad \mathrm{r}\cdot\sin\theta_2]^T = A\cdot[\mathrm{R}\cdot\cos\theta_1 \quad \mathrm{R}\cdot\sin\theta_1]^T$$
$$A \in \Re^{2\times 2} \tag{2}$$

There are eight parameters in this equation (2). But if P_1 and P_2 are obtained by an affine invariant point/region detector, the detector has already constructed the affine-invariant elliptical regions E_1 and E_2 around P_1 and P_2. And the normalized matrices A_1 and A_2 can be derived. So, we can normalize E_1 and E_2 in an affine invariant way around center points P_1 and P_2 respectively. Then we can set the affine matrix $A = A_2A_1^{-1}$ which projects elliptical region E_1 onto E_2.

So in our implementation, we can just match concentric circles G_1 and the resulting projected ellipses G_2 in Fig. 1 based on A and a set of known (R, r, θ_1, θ_2), that is, the set of points (r, θ_2) on an ellipse in the transformed image corresponding to the set of points on a circle of radius R and associated angle θ_1 in the original image.

3 Matching Method

Our approach aims to find an explicit correspondence for all the pixels in the regions around a pair of interest points. We address the problem of changes in illumination by normalizing intensity in the R, G, and B channels [2]. For each pair of interest points that we want to match, we generate polar-sampled (m angles) templates T_1 and T_2 originated from the interest points in the reference image I_1 and transformed image I_2 respectively as shown in Fig. 2. The problem of matching concentric circles in I_1 and the corresponding projected ellipses in I_2 becomes that of computing color difference cost between the line in T_1 and the path in T_2, generated by the affine transformation equation in Section 2.

Note that as Fig. 3 makes clear, corresponding columns in the template cannot be compared directly. There is an unknown stretching along the θ axis as well as an unknown rotation.

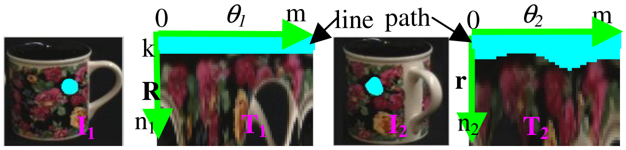

Figure 2: There are n_1 rows and n_2 rows in polar-sampled (m angles) templates T_1 and T_2 respectively.

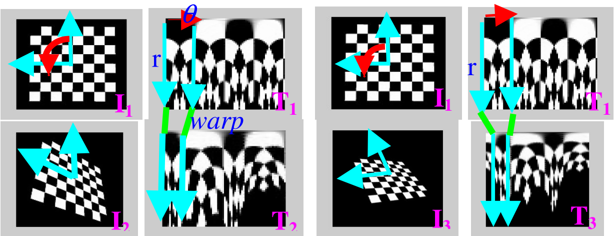

Figure 3: Two corresponded rays in reference image I_1 and transformed image I_2 are marked respectively. The warping situation exists in the polar sampled templates T_1 and T_2 generated from I_1 and I_2. Polar sampled templates T_1 and T_3 generated from I_1 and I_3 can't be compared directly due to an unknown image rotation.

To address these issues, for each pair of matched interest points, we run a dynamic warping algorithm to match polar sampled templates T_1 and T_2 generated from reference image I_1 and transformed image I_2 respectively. We take one narrow horizontal patch from row 1 to k in T_1 (corresponding to a circular region of radius R in I_1) and the corresponding area in T_2 (corresponding to an elliptical region in I_2) generated by the affine equation in section 2. Then we run the dynamic warping algorithm formulated in the following equations to match them. Note that strip column C_1 on T_1 could warp to more than one strip columns on T_2, and vice versa.

$$\begin{aligned}
&\cos t(C_1, C_2) = a + d(C_1, C_2)\\
&a = \min[\cos t(C_1 - 1, C_2), \cos t(C_1 - 1, C_2 - 1), \cos t(C_1, C_2 - 1)]\\
&\cos t(1,1) = d(1,1) \;\; 1 \le C_1, C_2 \le m\\
&d(C_1, C_2) = (T_1(C_1) - T_2(C_2))^2
\end{aligned}$$

This is started with one particular column on T_1 and different starting columns on T_2 so as to handle image rotation. The one with lowest cost(m,m) is the starting column of T_2 we want. Since the transformation parameters of adjacent projected ellipses are identical, we take the next horizontal line at row K where K > k in T_1 and generate the mapped path (r, θ_2 **:** the row r along each column θ_2) on T_2 (Fig. 2). Since we know the warping, from T_1 to T_2, we can efficiently compute the cost of a neighborhood of radius R within the original circle around the interest point P_1 on reference image I_1 (Fig. 4) versus the projected ellipse on transformed image I_2 around the interest point P_2 as

$$\begin{aligned}
&\cos t = \sum_{1 \le \theta_1 \le m} (I_1(\theta_1, R) - I_2(\theta_2, r))^2\\
&where\ (\theta_2, r) = T(\theta_1, R, A)
\end{aligned}$$

Then this process can be continued with the circle of larger radius. Thus the matched region is grown by increasing the radius of the original circle.

If (I_1 (θ_1, R)-I_2 (θ_2, r)) of any angle θ_1 along the trajectory is bigger than some pre-specified threshold, we mark the position with this particular angle θ_2 along the trajectory of the ellipse as being occluded, as shown in Fig. 4. We keep track of what fraction of the rays radiating from the interest point is marked occluded and stop the matching process if this fraction exceeds 1/3. This process does a good job of stopping the growing of the matched regions in the presence of partial occlusion.

The top N candidate pairs of matched interest points with the lowest average cost will be preserved. They are the pairs of matched points P_1 and P_2 that we will return. The matched regions around P_1 and P_2 in the reference image I_1 and the transformed image I_2 respectively after the process is the final result of image matching in our method.

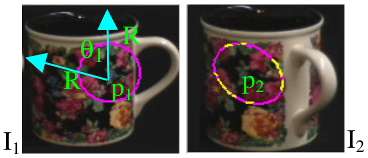

Figure 4: The concentric circular region in the reference image I_1 and the elliptical region in the transformed image I_2 that is the result of the projection of the concentric circles. The lighter colored points indicate possible occlusion positions we detected along the trajectory of the projected ellipse.

4. Experimental Results

The results of our matching method are compared with those from two SIFT-based local descriptors [3, 4] since they performed best among existing local photometric descriptors in Mikolajczyk and Schmid's evaluation [1]. To make the comparison fair, all three methods use the same input set of "interest points" obtained by the affine invariant points method in [5]. Our initial implementation takes about 1.8 seconds to process a pair of 320*240 images on a standard 2.52 GHz Pentium PC. Although this is slower than simple local descriptor comparisons; it is faster than most verification methods required to reduce the errors in the initial descriptor matches.

The main data set we used to evaluate the performance is from MIT-CSAIL Database[1]. We conducted experiments on several sequences of image frames from natural scenes of offices, streets, and corridors. We divide them into two parts: indoor scenes and outdoor scenes. Then we matched the image frames that are taken from widely separated viewpoints. We also evaluated five different data sets of controlled object images to address several issues that can affect the performance of image matching methods, including significant geometric transformation (Small_Rotation, Large_Rotation), texture (Distinctive_Texture, Similar_Texture) and occlusion. Each data set contains 10 image pairs collected from the Internet.

As shown in Fig. 5, only a handful of interest points can be detected on most images in our data sets as opposed to highly textured images (graffiti and cereal boxes) often used for evaluation. Since the number of detected interest points tends to be small, the matching accuracy per point pair is more important than the sheer quantity of the matched pairs, especially for applications such as recognition. So all three methods return the best N matches for each image pair (N is usually 5 or 10, depending on the size of the image). We report the accuracy rate (percentage of correct matches among the best N matches) to evaluate the performance of each method. The performance is displayed in Table 1.

[1] http://web.mit.edu/torralba/www/database.html

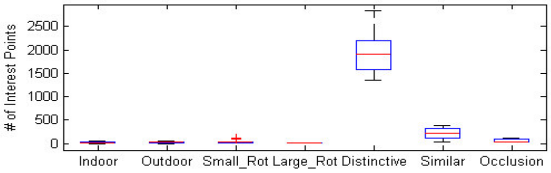

Figure 5. The Box-and-Whisker plot of the number of detected interest points in each data set using affine interest point detector

Data Set	Ours	SIFT	PCA-SIFT
Indoor	0.717/675	0.479/675	0.431/675
Outdoor	0.711/665	0.507/665	0.517/665
Small_Rot	0.92/50	0.90/50	0.92/50
Large_Rot	0.50/50	0.36/50	0.35/50
Distinctive	0.91/100	0.81/100	0.96/100
Similar	0.70/100	0.68/100	0.67/100
Occlusion	0.80/50	0.72/50	0.71/50

Table 1: Accuracy rate/total number of returned matches of the three methods on all data sets.

Our method outperforms other two methods by about 20-30% in the main data set. This result illustrates the advantage of incorporating more photometric information around matched points to verify the matches. SIFT-based descriptors occasionally produce incorrect matches due to large change in viewpoint (Fig. 6). Our method still returns correct matches in the overlapped part.

We also want to mention the occlusion data set since it is an important issue in image matching tasks. We used some pairs of images where the object in the reference image is partially covered by other objects in the transformed image. The results of our method in these cases are better than when using SIFT-based descriptors, as illustrated in Fig. 7. SIFT-based descriptors generate mismatches if the texture around an incorrectly matched point in the transformed image is similar to that of the correct match, which is occluded. It rarely happens in our methods because the matched region is extended to verify the match and we detect the sudden change in intensity difference due to the occlusion.

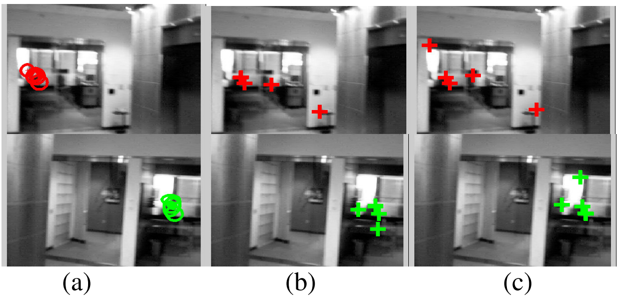

Figure 6. (a) All five matches returned by our method are correct. (b) Two of five matches returned by SIFT descriptors are the matches that the correct answer does not exist. (c) Three of five matches returned by PCA-SIFT descriptors are incorrect. Two incorrect matches are due to occlusion.

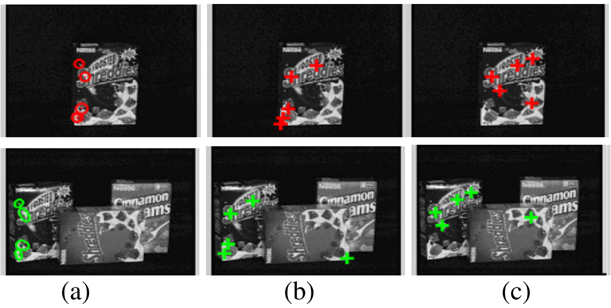

Figure 7. (a) All five matches returned by our method are correct. (b) One of five matches returned by SIFT descriptors is wrong due to similar texture. (c) One of five matches returned by PCA-SIFT descriptors is wrong due to occlusion.

5. Conclusions

In this paper we present a new method to perform reliable matching between different images. Our method constructs detailed pixel level matches between regions rather than relying on local photometric descriptors. We showed how to find complete region correspondences between concentric circles and the corresponding projected ellipses under affine transformation. It is more robust than previous methods to a variety of textures and to occlusion because it incorporates more luminance information around the interest points and because it finds a more detailed region correspondence. Experiments showed the new method performs substantially better in a variety of natural scenes than two SIFT-based descriptors. It also offers increased robustness to partial visibility, greater object rotation in depth, and more viewpoint angle change with acceptable computation cost.

References

[1] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", CVPR, 2003.

[2] B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?", ECCV, 1998.

[3] D. Lowe, "Object recognition from local scale-invariant features", ICCV, 1150-1157, 1999.

[4] Yan Ke and Rahul Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", CVPR, 506-503, 2004.

[5] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector", ECCV, 128-142, 2002.

[6] T. Tuytelaars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions", BMVC, 412-425, 2000.

[7] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters", IEEE Trans. on PAMI, 13(9), 1991.

[8] L. Van Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns", ECCV, 1996.

[9] F. Schaffalitzky and A.Zisserman, "Multi-view matching for unordered image sets", ECCV, 414-431, 2002.