Invariant Features from Interest Point Groups

Matthew Brown and David Lowe
{mbrown|lowe}@cs.ubc.ca

Department of Computer Science,
University of British Columbia,
Vancouver, Canada.

Abstract

This paper approaches the problem of finding correspondences between images in which there are large changes in viewpoint, scale and illumination. Recent work has shown that scale-space 'interest points' may be found with good repeatability in spite of such changes. Furthermore, the high entropy of the surrounding image regions means that local descriptors are highly discriminative for matching. For descriptors at interest points to be robustly matched between images, they must be as far as possible invariant to the imaging process.

In this work we introduce a family of features which use groups of interest points to form geometrically invariant descriptors of image regions. Feature descriptors are formed by resampling the image relative to canonical frames defined by the points. In addition to robust matching, a key advantage of this approach is that each match implies a hypothesis of the local 2D (projective) transformation. This allows us to immediately reject most of the false matches using a Hough transform. We reject remaining outliers using RANSAC and the epipolar constraint. Results show that dense feature matching can be achieved in a few seconds of computation on 1GHz Pentium III machines.

1 Introduction

A widely-used approach for finding corresponding points between images is to detect corners and match them using correlation, using the epipolar geometry as a consistency constraint [3, 13]. This sort of scheme works well for small motion, but will fail if there are large scale or viewpoint changes between the images. This is because the corner detectors used are not scale-invariant, and the correlation measures are not invariant to viewpoint, scale and illumination change. The first problem is addressed by scale-space theory, which has proposed feature detectors with automatic scale selection [6]. In particular, scale-space interest point detectors have been shown to have much greater repeatability than their fixed scale equivalents [7, 8]. The second problem (inadequacy of correlation) suggests the need for local descriptors of image regions that are invariant to the imaging process.

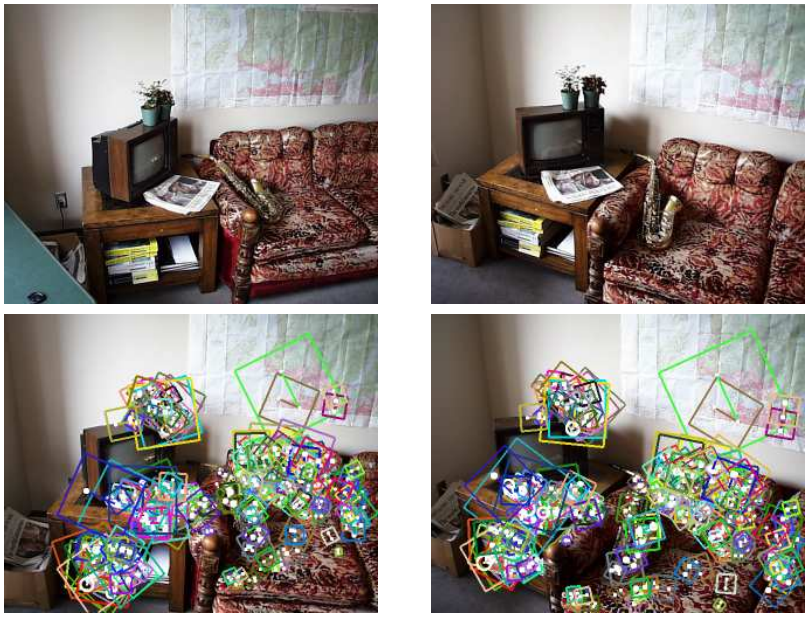


Figure 1: Matching of invariant features between images with a large change in viewpoint. Each matched feature has been rendered in a different greyscale.

Geometrical invariance can be acheived by assuming that the regions to be matched are locally planar, and by describing them in a manner which is invariant under homographies. Many authors use feature descriptors which are invariant under special cases of this group e.g. similarities or affinities. For example, Schmid and Mohr [10] use rotationally symmetric Gaussian derivatives to characterise image regions. Lowe's SIFT features [7] use a characteristic scale and orientation at interest points to form similarity invariant descriptors. Baumberg [2] uses the second moment matrix to form affine invariant features.

Our approach is to use groups of interest points to compute local 2D transformation parameters. By using different numbers of points we can form feature descriptors which are invariant to any 2D projective transformation (similarity, affinity or homography). Similar ideas have been proposed by [12], but they have suffered from the lack of scale-invariant interest point detectors in their implementation. Groups of interest points which are nearest neighbours in scale-space are used to calibrate the 2D transformation to a canonical frame [9]. The resampling of the image region in this canonical frame is geometrically invariant. Partial illumination invariance is achieved by normalising intensity in each of the R, G, B channels [5].

In addition to enabling robust matching, a key advantage of the invariant feature approach is that each match represents a hypothesis of the local 2D transformation. This fact enables efficient rejection of outliers using geometric constraints. We use broad-bin Hough transform clustering [1] to select matches that agree (within a large tolerance) upon a global similarity transform. Given a set of

feature matches with relatively few outliers, we compute the fundamental matrix and use the epipolar constraint to reject remaining outliers.

2 Interest Points in Scale-Space

Our interest points are located at extrema of the Laplacian of the image in scale-space. This function is chosen for its response at points with 2-dimensional structure, and for the fact that it can be implemented very efficiently using a Laplacian Pyramid [4]. In a Laplacian Pyramid, a difference of Gaussians is used to approximate the Laplacian. Pyramid representations have the advantage that the minimum number of samples are used to represent the image at each scale, which greatly speeds up computation in comparison with a fixed resolution scheme.

To find the maxima and minima of the scale-space Laplacian we first select samples which are extrema of their neighbours ± 1 sample spacing in each dimension. Then, we locate the extrema to sub-pixel / sub-scale accuracy by fitting a 3D quadratic to the scale-space Laplacian

$$L(\mathbf{x}) = L + \frac{\partial L}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 L}{\partial \mathbf{x}^2}\mathbf{x}$$

where $\mathbf{x} = (x, y, s)^T$ is the scale-space coordinate and $L(\mathbf{x})$ is the approximation of the Laplacian. The quadratic coefficients are computed by approximating the derivatives using pixel differences of the already smoothed neighbouring samples. The sub-pixel / sub-scale interest point location is taken as the extremum of this 3D quadratic

$$\hat{\mathbf{x}} = -\frac{\partial^2 L}{\partial \mathbf{x}^2}^{-1} \frac{\partial L}{\partial \mathbf{x}}$$

Locating interest points to sub-pixel / sub-scale accuracy in this way is especially important at higher levels in the pyramid. This is because the sample spacings at high levels in the pyramid correspond to large distances relative to the base image.

3 Invariant Features from Interest Point Groups

Once 'interesting points' in the image have been localised, robust matching requires an invariant description of the image region. One approach is to use information from the local image region itself. For example, the second moment matrix can be used to recover affine deformation parameters [2]. Degeneracies can cause problems for this approach. For example, if the local image region were circularly symmetric, it would be impossible to extract a rotation parameter.

An alternative approach is to use groups of interest points to recover the 2D transformation parameters. There are a number of reasons for adopting this approach. Firstly, improvements in the repeatability of interest points mean that the probability of finding a group of repeated interest points is sufficiently large. Secondly, the transformation computation is guaranteed to be non-degenerate.

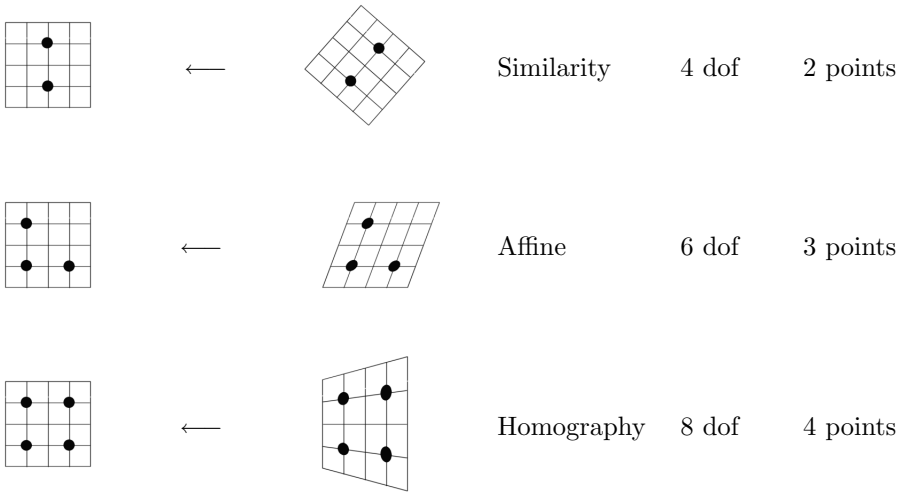


Figure 2: 2D transformation invariant features based on interest point groups. Groups of interest points which are nearest neighbours are formed, and used to calibrate the 2D transformation to a canoncial frame. The feature descriptor is the resampling of the image in the canonical frame.

Figure 3: Extraction of affine invariant features from a pair of images. Groups of 3 interest points are used to calibrate the affine transformation to a canonical frame. The image region is then resampled in the canonical frame to form the feature descriptor.

Figure 4: Finding consistent sets of feature matches. To find inliers to a homography we use a Hough transform followed by RANSAC. In this example, the test image was 640×512 pixels. The number of initial matches to consider was 5469, which was reduced to 279 by broad-bin Hough transform clustering, and further to 104 matches in the final solution using RANSAC.

Thirdly, and most importantly, since the interest points are very accurately localised, the 2D transformation estimate is also accurate.

We propose a family of 2D transformation invariant features based on groups of interest points as follows:

- Find groups of $2 \leq n \leq 4$ interest points which are nearest neighbours in scale-space
- Compute the $2 \times n$ parameter 2D transformation to a canonical frame
- Form the feature descriptor by resampling the region local to the interest points in the canonical frame

This is shown in figure 2. Aliasing is avoided by sampling from an appropriate level of the already constructed image pyramid. Partial illumination invariance is acheived by normalising in each of the R, G, B channels.

3.1 Feature Matching

Features are efficiently matched using a k-d tree. A k-d tree is an axis-aligned binary space partition, which recursively partitions the feature space at the mean in the dimension with the highest variance. We use 8×8 pixels for the canonical description, each with 3 components corresponding to normalised R, G, B values. This results in 192 element feature vectors.

4 Finding Consistent Sets of Feature Matches

Our procedure for finding consistent sets of feature matches consists of three parts, each of which refines the transformation estimate whilst rejecting outliers. First, we use a Hough transform to find a cluster of features in 2D transformation space. Next, we use RANSAC to improve the 2D transformation estimate. Finally we compute the fundamental matrix and use the epipolar geometry to reject additional outliers.

4.1 Hough Transform Clustering

A useful property of our invariant-feature approach is that each match provides us with the parameters of a 2D transformation. This can be used to efficiently reject outliers whose 2D transform estimates are inconsistent. We do this by finding a cluster (peak) in transformation space. This is known as a generalised Hough transform.

For similarity transforms, we parameterise the 2D transformation space by translations (t_1, t_2), log scale $(\log s)$ and rotation (θ). We discretise with bin sizes that are $1/8$ of the image size for translation, one octave for scale, and $\pi/8$ radians for rotation.

4.2 RANSAC Transformation Estimation

We refine the 2D transformation estimate using RANSAC. RANSAC has the advantage that it is largely insensitive to outliers, but it will fail if the fraction of outliers is too great. This is why we use Hough transform clustering as a first step. See figure 4.

If the scene is 3-dimensional, we first select inliers which are loosely consistent with a 2D transformation using the above methods, using a large error tolerance. This will hopefully find a dominant plane in the image, with the error tolerance allowing for parallax due to displacement from the plane. Then, given a set of points with relatively few outliers, we compute the fundamental matrix. This is used to find a final set of feature matches which is consistent with the epipolar geometry (figure 5).

5 Results

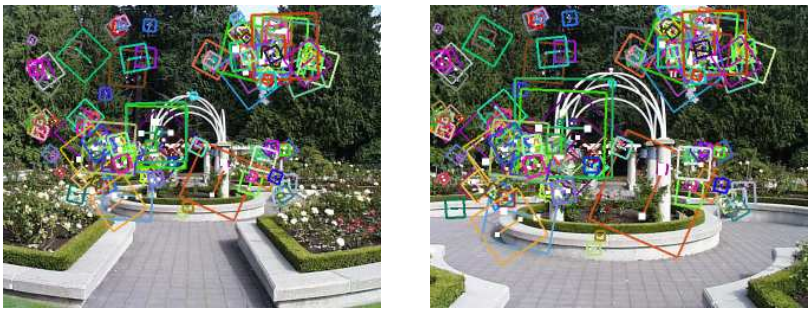
We have applied our invariant features to various recognition and registration problems including object recognition, panorama stitching (rotation estimation) and 3D matching (fundamental matrix estimation). Figure 1 shows successful matching despite a large change in viewpoint. Figure 5 shows the epipolar geometry computed from invariant feature matches. It can be seen that this epipolar geometry is consistent with the images, which are related by camera translation along the optical axis. Figure 6 shows results for object recognition. In this example, we have solved for a homography between the object in the two views. Note the large scale changes between the objects in the images in this case.

5.1 Repeatibility

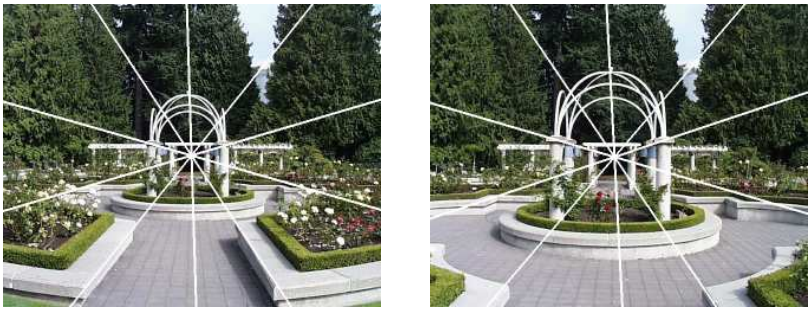
We have computed the repeatibility of our interest points using a test set of images from the Annapurna region in Nepal (see figure 8). Repeatibility is defined for two images related by a homography as the fraction of interest points which are consistent with that homography, up to some tolerance (see [11]).

Figure 9 demonstrates the effect of sub-pixel / sub-scale accuracy of interest point location on repeatability. This correction gives a clear improvement in the accuracy of interest point location. It is particularly important for high levels

(a) UBC Rose Garden



(b) Correct feature matches



(c) Epipolar geometry

Figure 5: A pair of images from the UBC Rose Garden. Similarity invariant features formed from groups of 2 interest points are extracted and matched. Outliers are rejected by requiring (loose) consistency with a global similarity transformation. Then, the fundamental matrix is computed and the epipolar constraint used to select a final set of consistent matches. Note that the epipolar geometry is consistent with a camera translation along the optical axis.

Figure 6: Object recognition using invariant features from interest point groups. The white outlines show the recognised pose of each object according to a homography. For 3D objects, this model is appropriate if the depth variation is small compared to the camera depth.

of the pyramid, where sample spacings correspond to large distances in the base image. In addition to increasing the accuracy of transformation computations, accurate interest point localisation also enables more accurate feature descriptors, which improves matching.

6 Conclusions

In this paper we have introduced a family of features based on groups of scale-invariant interest points. The geometrical and illumination invariance of these features makes them particularly applicable for solving difficult correspondence problems. We have shown the importance of sub-pixel / sub-scale localisation of interest points, which critically improves the accuracy of descriptors. To reject outliers we use Hough transform clustering followed by RANSAC to select a set of feature matches that are loosely consistent with a global 2D transformation. We then compute the fundamental matrix, and use the epipolar constraint to reject remaining outliers. These techniques enable practical recognition and registration tasks to be performed in a few seconds of computation using 1GHz Pentium III machines.

Future work will look at more efficient parameterisation of feature descriptors, and alternative methods for computing local canonical frames.

Figure 7: Cylindrical panorama of the Annapurna sequence of images. This was computed by estimating the (3 dof) rotations between images from feature matches.

1 2 3 4

Figure 8: Images from the Annapurna sequence. There are 15 images in total. The images are related by rotation about the camera centre.

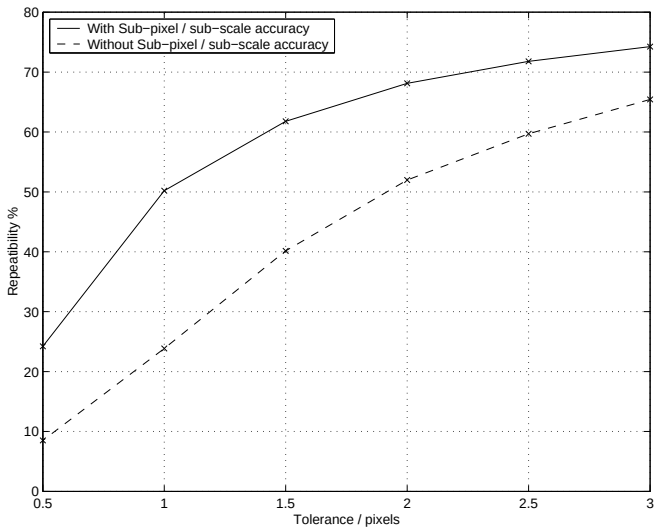


Figure 9: Repeatibility of interest points with and without sub-pixel / sub-scale accuracy.

References

[1] D. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[2] A. Baumberg. Reliable Feature Matching Across Widely Separated Views. In *Proceedings of the Interational Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2000.

[3] P. Beardsley, P. Torr, and A. Zisserman. 3D Model Acquisition from Extended Image Sequences. In *Proceedings of 4th European Conference on Computer Vision (ECCV'96)*, volume II, pages 683–695, Cambridge, April 1996. Springer-Verlag.

[4] P. Burt and E. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 9(4):532–540, 1983.

[5] B. Funt, K. Barnard, and L. Martin. Is Machine Colour Constancy Good Enough? In *Proceedings of 5th European Conference on Computer Vision (ECCV'98)*, pages 445–459, 1998.

[6] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[7] D. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, September 1999.

[8] K. Mikolajczyk and C. Schmid. Indexing Based on Scale Invariant Interest Points. In *Proceedings of the International Conference on Computer Vision*, pages 525–531, 2001.

[9] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. Canonical frames for planar object recognition. In *Proceedings of the European Conference on Computer Vision*, pages 757–772, 1992.

[10] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.

[11] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. In *Proceedings of the International Conference on Computer Vision*, pages 230–235, Bombay, 1998.

[12] T. Tuytelaars and L. Van Gool. Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions. In *Proceedings of the 11th British Machine Vision Conference*, pages 412–422, Bristol, UK, 2000.

[13] Zhengyou Zhang, Rachid Deriche, Olivier Faugeras, and Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry,. *Artificial Intelligence, December 1995*, 78:87–119, 1995.