Viewpoint Invariant Texture Matching and Wide Baseline Stereo

Frederik Schaffalitzky and Andrew Zisserman

Department of Engineering Science
University of Oxford
Oxford, OX1 3PJ, UK

Abstract

We describe and demonstrate a texture region descriptor which is invariant to affine geometric and photometric transformations, and insensitive to the shape of the texture region. It is applicable to texture patches which are locally planar and have stationary statistics. The novelty of the descriptor is that it is based on statistics aggregated over the region, resulting in richer and more stable descriptors than those computed at a point.

Two texture matching applications of this descriptor are demonstrated: (1) it is used to automatically identify regions of the same type of texture, but with varying surface pose, within a single image; (2) it is used to support wide baseline stereo, i.e. to enable the automatic computation of the epipolar geometry between two images acquired from quite separated viewpoints.

Results are presented on several sets of real images.

1 Introduction

The objective of this paper is a texture region descriptor which is invariant to affine geometric and photometric transformations and is insensitive to the segmentation of the textured region. The underlying assumptions are that the textured regions are locally planar and stationary (the statistical distribution is spatially homogeneous). A descriptor with this degree of invariance is sufficient to identify a textured region irrespective of viewpoint and lighting in a perspective image. Such a descriptor enables the identification of the same type of texture within an image, such as brick texture on two sides of a house. It also enables texture regions to be matched between images acquired from different viewpoints, such as two views of the front of a house. In the first case – *intra*-image matching – the matched texture regions correspond to different surface patches, however in the second – *inter*-image matching – the matched texture regions correspond to the same surface patch. We demonstrate both these cases in this paper.

The target application of this texture descriptor is to enable the automatic computation of epipolar geometry between two views despite significant changes in viewpoint – the wide baseline stereo problem [18]. This area is reviewed below. However, the texture descriptor is not limited to this application, and provides a step towards viewpoint invariant texture segmentation [1] and classification. The method of achieving the required invariance is described and demonstrated in section 2, with the use of the descriptor for viewpoint invariant texture classification illustrated in section 2.3.

The wide baseline application is described in section 3.

1.1 Wide baseline stereo background

For a significant variety of scene types the epipolar geometry can be computed automatically from two uncalibrated images provided the motion between the views is limited [22, 26]. The computation methods are based on robust statistics and proceed in three steps: interest points are detected independently in each image; putative point matches are then computed between the images based on a measure of proximity and intensity neighbourhood similarity; then the fundamental matrix (representing the epipolar geometry) and a subset of these matches consistent with the epipolar geometry are determined simultaneously, using a robust estimation algorithm such as RANSAC or LMS.

The second step is essentially matching of interest points by tracking, and it is this that fails in the wide baseline case. For example, if there is a significant difference in surface foreshortening between the two images, then cross-correlation on the *image* intensity neighbourhood of corresponding interest points will not correctly measure *surface* similarity. This is because corresponding surface points in the neighbourhood are not being compared by an image based cross-correlation. Similarly, if there is significant rotation of the image about the camera principal axis (cyclo-rotation) then again a correlation measure for determining similarity of surface neighbourhoods is defeated, and furthermore the proximity search for matches (looking in the neighbourhood of the point's previous position) fails.

It is this point matching problem that is addressed in the literature on widebase line stereo. Pritchett and Zisserman [18] approached the problem by computing local planar homographies in order to correct the cross-correlation

measure and proximity matching. The homographies were obtained both by matching feature groupings and by a pyramid search. Other authors have sought to obtain a descriptor to *label* or characterise the interest points based on their intensity neighbourhood. Then, provided this descriptor is invariant to viewpoint, matching can proceed by identifying points with the same label. In order to achieve this viewpoint invariance it is sufficient, to a good approximation, that the descriptors are invariant to affine geometric *and* photometric transformations.

Such descriptors have been developed systematically over the past five years, starting from invariance to a restricted class of transformations and progressing towards full affine invariance. Schmid and Mohr [20] achieved image rotational invariance using a set of differential operators acting on the interest point's intensity neighbourhood, and employed several scales to achieve overall similarity invariance. Gouet *et al.* [8] extended this to colour images. Lindeberg and Gårding [14] described an iterative procedure to determine an affine invariant neighbourhood, and also the idea that extrema of derivatives over scale may be used to determine a characteristic scale. Both ideas have been influential [7, 15]. In particular, Schmid and Mikolajczyk [17] used a characteristic scale to obtain reliable interest point matching over an impressive range of scale and rotations. Baumberg [2] employed the iterative procedure and demonstrated a local descriptor with affine geometric and photometric invariance, using multiple scales in the interest point detection. Descriptors with affine invariance have also been developed for point pairs [21], and for regions by Tuytelaars and Van Gool based on corners and edges [23], or local intensity extrema [24].

In all the above cases it is necessary to compute the descriptor over the same surface region – the region is determined from a single image, but it must be determined independent of viewpoint. The imaged region must therefore adapt its shape with viewpoint, and this is a large part of the ingenuity involved in constructing these invariants.

This is where the work of this paper differs most significantly from those above. We are aiming for descriptors with the same class of invariance, but additionally the descriptors should not depend on extracting a viewpoint invariant surface patch. Instead the descriptor measures statistical properties of the texture and these can be measured over any neighbourhood provided the texture is stationary.

2 Affine invariant texture descriptors

This section describes the theory and computation of texture descriptors which are invariant to the following class of transformations: an affine transformation of the geometry $\mathbf{x} \mapsto \mathtt{A}\mathbf{x}+\mathbf{b}$ where $\mathbf{x}, \mathbf{b}$ are 2-vectors and $\mathtt{A}$ is a 2×2 matrix; and a photometric affine map on the intensity $I \mapsto sI + t$; and are also insensitive to the shape of the region used in their computation.

The assumption is that the statistical distribution of the surface texture is spatially stationary (i.e. is translationally invariant), so that the statistics of the distribution can be gathered from regions of any shape and (reasonable) size. A key idea is that measures on this distribution may be used to achieve affine normalization – both geometric and photometric. Further measures are then used to describe the texture in the affine normalized frame. This is similar to the canonical frame method of Lamdan *et al.* [12] in the case of affine transformations of planar geometric configurations.

Rather than achieve affine normalization in one step it is more convenient to first achieve *partial* affine normalization, in that any affinely distorted region is mapped to the same region up to rotation and isotropic scaling. The remaining transformations can then be absorbed by measuring scale and rotation invariant texture descriptors in this normalized frame. The computed descriptors will then be invariant to a full affine transformation of the original image.

The normalization (up to scale and rotation) proceeds by affinely mapping the texture such that a statistical measure is isotropic (i.e. has no distinguished directions). The descriptor used here is the second moment matrix, for which isotropy means that it is a scalar multiple of an identity matrix. If the texture on the surface is isotropic, then the normalizing transformation will determine a fronto-parallel view of the surface – in a similar manner to recovering surface orientation in shape-from-texture using isotropy [3, 25]. However, in general the texture need not be isotropic, and the normalized texture is simply a frame that can be reached from any affine transformation of the texture in the sense that all such frames are related by a similarity transformation.

Our approach to affine invariant descriptors is inspired by previous work by several authors [1, 2, 14]. Where we differ is that the normalization and descriptors are not determined independently at each point (using image derivatives computed on the neighbourhood of the point). Instead, since the texture is stationary by assumption, the normalization and descriptors are computed from statistical measures aggregated over a region. This enables richer and more stable descriptors to be obtained and, as will be demonstrated, the descriptors are insensitive to the particular patch of texture used.

In the following we detail the computation of the normalization and then assess its stability against changes in viewpoint, lighting, and region support. We then detail the computation of the texture descriptor and demonstrate that it satisfies two essential requirements: stability – it is indeed invariant to the specified class of transformations; and discriminability – the value of the descriptor is not the same for all textures.

2.1 Computing the normalized frame

The *second moment matrix* of an image region $\Omega \subseteq \mathbb{R}^2$ is the covariance matrix of the gradient of I over the region:

$$\mathtt{M}_I = \int_\Omega \nabla I \otimes \nabla I \, \frac{dx\,dy}{|\Omega|} = \int_\Omega \begin{pmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{pmatrix} \frac{dx\,dy}{|\Omega|}$$

where $|\Omega|$ denotes the area of Ω and the subscripts denote partial derivatives with respect to the image coordinates. The second moment matrix has been used in numerous works on texture. It is a measure of the local variation in the image gradient and is the basis for the Harris corner detector [10].

Under an affine geometric transformation with linear part $\mathtt{A}$, the second moment matrices of the original image I and the transformed image J are related by

$$\mathtt{M}_I = \mathtt{A}^\top \mathtt{M}_J \mathtt{A}$$

What is required for normalization is an affine transformation with linear part $\mathtt{A}$ having the property that $\mathtt{A}^\top \mathtt{A} = \mathtt{M}_I$.

Under an affine photometric transformation $I \mapsto sI + t$, the constant bias t is eliminated by the gradient operator and the scale factor s shows up as a scale factor s^2 in the moment matrix. The computed normalizer $\mathtt{A}$ will therefore be subject to scaling by s and so a problem of scale selection remains. This will be addressed later.

Implementation details. A matrix $\mathtt{A}$ satisfying $\mathtt{A}^\top \mathtt{A} = \mathtt{M}_I$ is not uniquely determined because pre-multiplying it by any 2×2 rotation matrix will yield another solution. However, this is the only ambiguity. A unique solution can be obtained by requiring that $\mathtt{A}$ be symmetric (and positive definite), in which case an algebraic square root of $\mathtt{M}_I$ is given by replacing the eigenvalues of $\mathtt{M}$ by their square roots (since $\mathtt{M}_I$ is positive definite, this is always possible).

However, image gradients computed in the original image and skew-normalized frame will differ, because the kernel of the derivative operator will cover different pixels as its effective shape changes. This must be taken into account in order that the skew-normalized frame is viewpoint invariant, and is achieved by the following iterative procedure:

- **Initialization:** using circularly symmetric derivative operators on the original image compute the square root of the moment matrix and use it to form a skew-normalized version of the original image patch.
- **Iterate:**
  1. re-estimate the gradients and the moment matrix in the transformed image using circularly symmetric operators;
  2. concatenate the new skew normalization to the previous one;
  3. generate a new skew normalized image.

The termination criterion is that no point in the region move by more than 0.5 pixels in two consecutive updates. Typically, convergence is reached within 10 iterations. Image derivatives are computed using derivatives of gaussians with $\sigma = 0.7$, and integrals over image regions are implemented as sums over discrete pixels in the normalized frame.

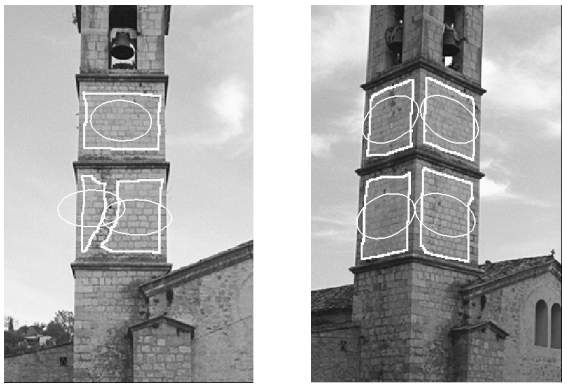


Figure 1: Two images of the same textured surfaces acquired from different points of view. The affine distortion between the imaged sides of the tower is evident, as is the difference in brightness. The second moment matrix is computed independently in each of the marked regions (indicated by their boundaries), and is represented as an ellipse centred on the region's centroid. Note the computed ellipse appears to be attached to the surface: it transforms in the same manner (covariantly) as the surface between views. Furthermore for different regions on the same side of the tower the ellipse is virtually identical (up to size). This demonstrates the viewpoint invariance of the computation, and also the insensitivity to the shape of the support regions.

This approach is guided by the ideas of Lindeberg [14] and the method of Baumberg [2] although the latter author had the added complication of having to find the correct support region for the whole integral whereas in our application this is fixed (it is provided by the given region segmentation) and only the shape of the filter kernels adapt.

The descriptors are made invariant to affine photometric scaling by robustly normalizing the range of the intensity signal over the region before computing the descriptor. Other important implementation details include: avoiding undersampling the original texture; robust computation of gradients; and removal of spatial illumination variation over the texture region (i.e. $I \mapsto s(\mathbf{x})I + t$). Further details are given in [19].

Evaluation of the normalization. Figures 1 and 2 show examples of computed moment matrices for various image patches. The ellipse drawn on the figures is the locus of displacements $\mathbf{h}$ from the centroid satisfying $\mathbf{h}^\top \mathtt{M}_I \mathbf{h} =$ constant. The figures demonstrate that the computed ellipse is co-variant with viewpoint, and is insensitive to position, shape, and scale of the support patch. This is a qualitative assessment, a quantitative assessment is given below for the affine invariant texture descriptor (i.e. including both normalization and texture description in the normalized frame).

Figure 3 shows the skew-normalized versions of the two images from figure 1, using the moment matrices computed

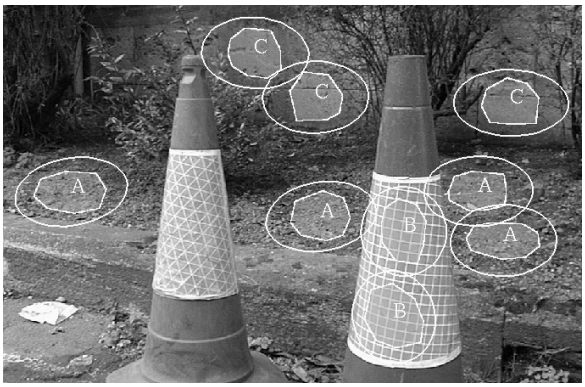


Figure 2: Computed second moment matrices (ellipses) for different support regions (highlighted) from various planar textures. The ellipses vary very little within each texture, despite the change in position, shape and size of the support region.



Figure 3: Normalized images obtained by applying affine transformations to the images of figure 1 so that the second moment matrix is isotropic. The two textures are now related by an image rotation and scaling (although a small amount of residual perspective distortion does remain).

over the regions shown. The scale of the normalizing transforms is chosen so as to preserve area (*i.e.* determinant 1) for the purpose of illustration. Note that although there is considerable variation in lighting and viewpoint between the original images, the rectified images are rotated (and slightly scaled) versions of each other.

2.2 Measurements in the normalized frame

Having achieved an affine skew-normalization, the problem remains to extract descriptors invariant to the spatial scale and rotation of the frame. We first address rotational invariance.

There is a considerable literature on computing rotationally invariant texture descriptors (e.g. see [5, 9, 16] for citations). A common approach is to compute the response of a bank of oriented filters applied to each pixel. The descriptor then consists of measures on the distribution of these responses aggregated over the texture patch. These descriptors include piecewise representation as cluster centres, statistical measures (mean, covariance ...), and histogram binning [11] amongst many others. We follow this standard approach, using a filter bank based on Zernike moments.

The descriptor we employ is to use the responses of a rotationally invariant bank of local operators to label each pixel according to which filter gives the strongest response. The descriptor, then, is the histogram of pixel labels found. This type of texture descriptor is similar to the texton approach of [16].

Concerning scale invariance: it is found empirically that the histogram does have a certain degree of invariance to scale as pixels labels are not extremely sensitive to scale (the property of being extremal amongst all responses is quasi-invariant). However, to truly handle invariance to scale, the histogram is computed at five different scales of the filter bank support, and the five (L^1-normalized) histograms constitute the region descriptor. Further details are given in [19].

We investigated selecting a characteristic scale for the region, following the approach of Lindeberg [13], by choosing the scale at which the scale-normalized Laplacian operator attains a local maximum. While this works very well for interest points (see for example [17] for excellent results), in a textured region the method will often fire uniformly at all possible scales as the point of application is moved about, and consequently is not helpful here.

In summary: the descriptor consists of the set of histograms (normalized to unit L^1-norm).

Evaluation of the affine invariant descriptors. In the tests that follow we are interested in determining texture regions with the closest descriptors. A descriptor consists of a set of histograms. The "distance" between two histograms is determined using a χ^2 measure, and the "distance" between two descriptors is computed by finding the minimum distance between histograms at corresponding scales (including a common shift in scale as necessary).

The stability and discriminability of the descriptor were tested on synthetic data generated from the Brodatz [4] collection of textures. For a given Brodatz texture, three sub-regions were chosen; these should be classified as the same texture (if our descriptor is insensitive to segmentation). Three synthetic affine warps were then applied to each of the three regions; the resulting warped texture regions should also be classified as the same texture (if our descriptor is affine invariant). From a selection of eight Brodatz textures we thus obtain $8 \times 3 \times 3 = 72$ texture samples for which we know the ground truth classification decisions and the only parameter left to tune is the threshold on the similarity measure. The ROC curve for this classifier is shown in figure 4. The results demonstrate the effectiveness of this classifier, assuming a suitable threshold has been found.

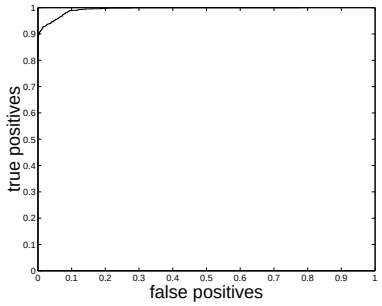


Figure 4: The ROC curve for texture matching using our texture descriptor, applied to a set of 72 regions of warped Broadatz textures. The parameter used to trace the ROC curve is a threshold on our similarity measure.

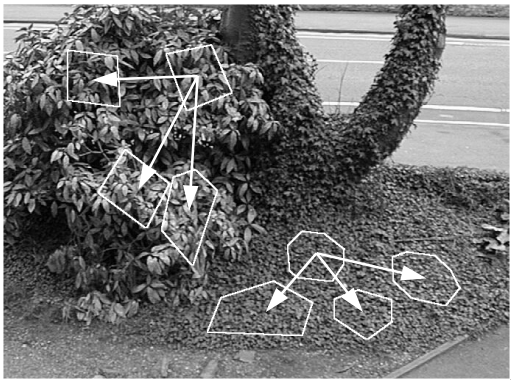


Figure 5: Two different textures and four regions manually selected within each texture. Invariant descriptors were computed for each region and, using our similarity measure, each region's three most similar other regions determined. In all eight cases, the three regions from the same texture were chosen.

2.3 Intra-image texture matching

In many cases, similar texture occurs multiple times in the same image in different orientations. Regions of textures of the same type can be determined using the texture descriptor. This is demonstrated in figure 5 which shows different hand selected parts of the same texture matched in a single image.

Intra-image texture matching can be automated by first segmenting the image into texture patches, and then matching these patches using the invariant descriptor. The segmentation algorithm is image-based so will place boundaries between regions of the same texture type (but different orientation) but the view-point invariant descriptor is able to detect their similarity and group them together. We illustrate this for the right image of figure 6.

The image is segmented into texture regions using a local implementation of the Malik *et al.* texton segmentation algorithm [16] based on the normalized cut criterion. We deliberately over-segment so that textures are not grouped



Figure 6: A wide baseline pair of images.



Figure 7: The segmentation of the images in figure 6 into regions of homogeneous texture using the method of [16]. The over-segmentation is intentional.

over weak boundaries. This is illustrated in figure 7 which shows the regions found. The sky has been extensively over-segmented, but we reject regions of (locally) uniform intensity as non-texture so this has no effect on the later stages of the algorithm anyway. One of the final groupings of texture patches is illustrated in figure 8.

3 Wide baseline stereo matching

This section describes an application of the texture descriptors developed in section 2 to inter-image matching. It is shown that the difficulties of wide baseline matching are ameliorated by the use of these descriptors.

3.1 Image matching using texture descriptors

The approach adopted here is to first segment each image into homogeneous texture regions (as in figure 7); and then attempt to match these regions between images using the invariant descriptor for each region. Since several regions will have the same texture, each region in one image will have a set of possible matches in the other. This ambiguity is resolved by computing and matching features, since sets of features (such as interest points around a distinguishing mark) are unique to a particular region. The result is a



Figure 8: Similar textures at different orientation can be detected within a single image. For one region, the 10 most similar other regions in the image are shown by arrows. The original image is shown on the left.

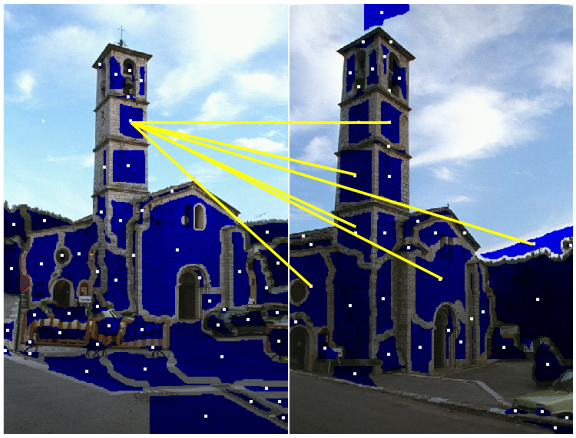


Figure 9: A region (left) and the regions (right) with which it is putatively matched. The white dots are the centroids of the regions and are for identification only; the translational part of the affine transformation between the regions is computed from image point correspondences.

set of interest point matches which then forms the basis for computing epipolar geometry.

In more detail, suppose there is a putative match between a region in one image and a region in another. This match arises because the texture descriptors of the two regions are similar. We want to verify or reject this match. Note that even if the match is correct the two regions may not correspond to precisely the same surface patch due to imperfect segmentation, but there will be some overlap. The putative region match partially provides the affine transformation between the regions. This transformation is obtained by composing the affine normalizations associated with each region, and is determined up to a rotation and translation (i.e. a planar Euclidean transformation) between the images. If the rotation were known, then interest points computed in each region could be matched by using the transformation to essentially reduce the correspondence problem to that of tracking, where only translation remains. Since only a rotation needs to be estimated, we try a number of rotations and choose the one with the maximum number of feature correspondences that have normalized cross correlation above a threshold. In this manner feature correspondences are used to verify or reject the putative region matches.

Finally, the most powerful advantage of having the affine transformation is that further matches can be generated very efficiently for verified regions, in fact with the ease of the short baseline case with horizontal epipolar lines.

Algorithm summary

1. **Basic segmentation:** Segment each image into regions with different textures.
2. **Texture labelling:** Compute invariant descriptors for each texture patch.
3. **Putative inter-image texture matching:** Guided by the similarity measure for invariant descriptors, establish a set of putative inter-image texture region matches.
4. **Verification:** Verify the inter-image texture matches using interest points within each region and the affine transformations provided by the texture descriptors.
5. **Point match generation:** Grow further point matches for each of the verified region matches.
6. **Robust estimation of the fundamental matrix:** Using the matched interest points use a standard method to robustly estimate $\mathbf{F}$ and a consistent set of point correspondences.

3.2 Example and implementation details

The method will be illustrated on the image pair of figure 6. In this case there are 53 and 43 texture regions (as shown in figure 7), and 1600 interest points computed in each image. The putative region matches (e.g. see figure 9), computed from the texture descriptor ranking, are verified/rejected on their associated interest point matches. Point matches are assessed by locally affine registering their intensity neighbourhoods and then minimizing an SSD error that corrects for lighting changes. A local neighbourhood match is accepted if the RMS fitting error is at most 0.1 (intensity is scaled to the range $[-1, +1]$) and the scale factor for lighting correction is between 0.5 and 2. Of all the interest point matches from the region match verification step, 120 point matches passed this interest point verification step. The 11 verified region matches are shown in figure 10.

In order to improve the effectiveness of the RANSAC epipolar geometry estimator, we "drown" the outliers that persist by growing more point matches (with a 5×5 search window) using correlation guided by the local SSD registrations. This yielded 1800 point matches of which 1400 survived the robust fitting step. This shows that incor-

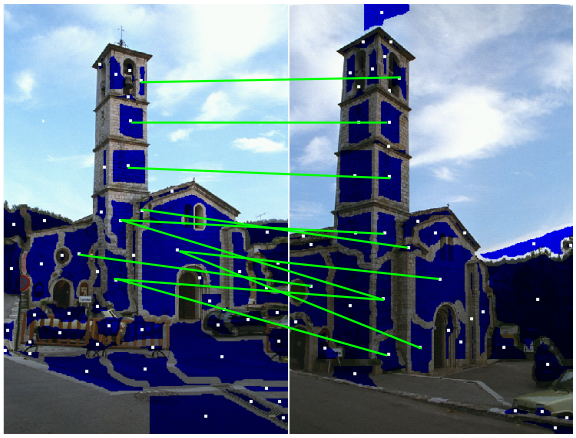


Figure 10: Verified 11 region matches. These are the putative region matches which have been verified using correlation on corner features. Note that regions are matched despite foreshortening size changes of 50% or more, and significant changes in segmentation between the two images. There are multiple matches for some regions due to this segmentation difference.



Figure 11: The final 1400 interest point matches after a robust fit and non-linear optimization of the epipolar geometry. A match is indicated by the line linking the corner to its position in the other view.

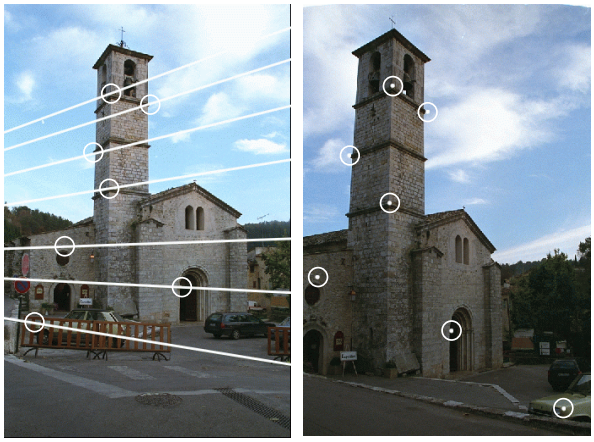


Figure 12: Epipolar geometry computed from the point matches found.

rect matches are far less likely to produce more incorrect matches by guided correlation than correct matches are to produce correct matches. The point matches are shown in figure 11.

The computed epipolar geometry is illustrated in figure 12. The quality is excellent, and all steps of the algorithm are automatic. The two images used are part of a set of 15 and we have run the algorithm on all 105 pairs from this set with more than 65% giving at least 100 interest point matches, and more than 80% giving at least 20 matches.

An example on a different type of scene is given in figure 13. Again the entire process from images to computed epipolar geometry is automatic. These two scene types (brick building facades and textured rocks) demonstrates that the invariant descriptor has sufficient stability and discriminability for both semi-regular and stochastic textures.

4 Conclusions and Extensions

We have described and demonstrated a class of statistical descriptors which are affine invariant and insensitive to region segmentation. We are currently investigating classes of textures for which the affine normalization procedure is guaranteed to converge, and classes for which the method is inapplicable. For example, if the surface texture is highly anisotropic (such as regularly spaced stripes) then the affine normalization stage will fail as there is not a reasonable transformation to an isotropic frame. We are also investigating the equivalence class of textures under the histogram descriptor. Details of this work are given in [19].

There are many variations possible on the presented two stage descriptor: for example other statistical measures (than isotropy) could be used to determine a normalized frame; and there is much scope for fuller texture descriptions in the frame. Chetverikov [6] describes an affine invariant descriptor applicable to textures that are at least weakly regular. It is based on the auto-correlation function, and has been demonstrated predominantly for Brodatz textures. This offers a very different alternative to our approach, as a texture descriptor is obtained directly without involving an intermediate normalization.

The intra-image application of the descriptor of section 2.3 suggests that texture segmentation could be made viewpoint invariant. The segmentation method we have used, based on textons and normalized cuts [16], is invariant to Euclidean image transformations but not to affine transformations. However, the second moment ellipse may be used to indicate a change in surface orientation for the same texture (e.g. bricks on adjacent walls of a building),

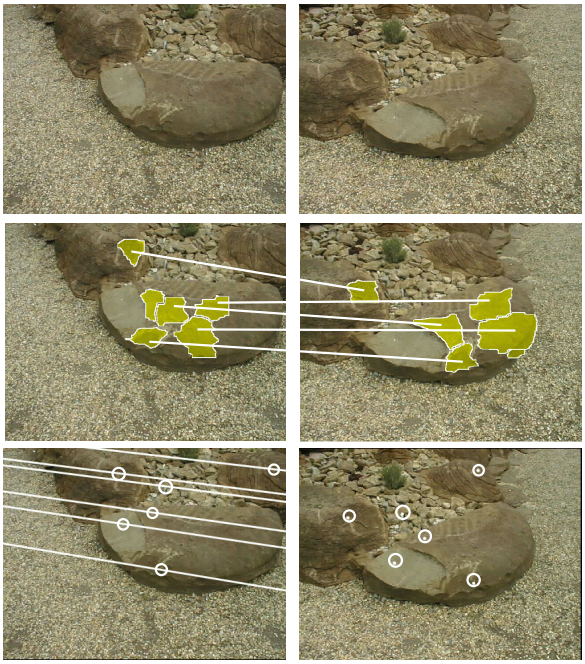


Figure 13: Another example of epipolar geometric obtained using the method described in this paper. Top: original images. Middle: some of the matched regions. Bottom: final epipolar geometry computed from 221 interest point matches.

and to compensate for orientation effects between regions. The texton and normalized cut framework could thus be extended to include this information. Some progress in viewpoint invariant segmentation has already been demonstrated by [1, 6].

The inter-image application of section 3 adds another semi-local affine invariant descriptor to the raft of those already available [2, 7, 21, 23, 24] for wide baseline matching. The texture descriptor is especially suited in its sole use to scenes which contain multiple textures. Of course, for unknown scene types, all available descriptors should be applied and used in concorde.

Acknowledgements

We are grateful to RobotVis INRIA Sophia-Antipolis for providing the Valbonne images and for financial support from EC Project VIBES.

References

[1] C. Ballester and M. González. Affine invariant texture segmentation and shape from texture by variational methods. *J. Math. Imaging and Vision*, 9:141–171, 1998.

[2] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. CVPR*, pages 774–781, 2000.

[3] A. Blake and C. Marinos. Shape from texture: estimation, isotropy and moments. *Artificial Intelligence*, 1990.

[4] P. Brodatz. *Textures: A Photographic Album for Artists & Designers.* Dover, New York, 1966.

[5] M. J. Chantler, G. McGunnigle, and J. Wu. Surface rotation invariant texture classification using photometric stereo and surface magnitude spectra. In *Proc. BMVC.*, pages 486–495, 2000.

[6] D. Chetverikov. Pattern regularity as a visual key. *Image and Vision Computing*, 18:975–985, 2000.

[7] S. Gilles. *Robust Description and Matching of Images*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1998.

[8] V. Gouet, P. Montesinos, and D. Pel. A fast matching method for color uncalibrated images using differential invariants. In *Proc. BMVC.*, pages 367–376, 1998.

[9] G. M. Haley and B. S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE PAMI*, 8(2):255–269, 1999.

[10] C. J. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Alvey Vision Conf.*, pages 147–151, 1988.

[11] S. Konishi and A. L. Yuille. Statistical cues for domain specific image segmentation with performance analysis. In *Proc. CVPR*, 2000.

[12] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson. Object recognition by affine invariant matching. In *Proc. CVPR*, pages 335–344, 1988.

[13] T. Lindeberg. Principles for automatic scale selection. In B. Jahne, editor, *Handbook of Computer Vision and Applications*, chapter 11, pages 239–275. Academic Press, 1999.

[14] T. Lindeberg and J. Gårding. Shape-adapted smoothing in estimation of 3-d depth cues from affine distortions of local 2-d brightness structure. In *Proc. ECCV*, pages 389–400, May 1994.

[15] D. Lowe. Object recognition from local scale-invariant features. In *Proc. ICCV*, pages 1150–1157, Sep 1999.

[16] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue combination in image segmentation. In *Proc. ICCV*, pages 918–925, Kerkyra, Greece, Sep 1999.

[17] C. Mikolajczyk and C. Schmid. Indexing based on scale invariant points. In *Proc. ICCV*, 2001.

[18] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. ICCV*, pages 754–760, Jan 1998.

[19] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture description and matching. Technical report, Dept. of Engineering Science, University of Oxford, 2001.

[20] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE PAMI*, 19(5):530–534, May 1997.

[21] D. Tell and S. Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In *Proc. ECCV*. Springer-Verlag, Jun 2000.

[22] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, 1997.

[23] T. Tuytelaars and L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *Int. Conf. on Visual Information Systems*, pages 493–500, 1999.

[24] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. BMVC.*, pages 412–425, 2000.

[25] A. P. Witkin. Recovering surface shape and orientation from texture. *Artificial Intelligence*, 1981.

[26] Z. Zhang, R. Deriche, O. D. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.