Feature Based Methods for Structure and Motion Estimation

P. H. S. Torr[1] and A. Zisserman[2]

[1] Microsoft Research Ltd, 1 Guildhall St
Cambridge CB2 3NH, UK
philtorr@microsoft.com
[2] Department of Engineering Science, University of Oxford
Oxford, OX1 3PJ, UK
az@robots.ox.ac.uk

1 Introduction

This report is a brief overview of the use of "feature based" methods in structure and motion computation. A companion paper by Irani and Anandan [16] reviews "direct" methods.

Direct methods solve two problems simultaneously: the motion of the camera and the correspondence of every pixel. They effect a global minimization using all the pixels in the image, the starting point of which is (generally) the image brightness constraint (as explained in the companion paper).

By contrast we advocate a feature based approach. This involves a strategy of concentrating computation on areas of the image where it is possible to get good correspondence, and from these an initial estimate of camera geometry is made. This geometry is then used to guide correspondence in regions of the image where there is less information. Our thesis is as follows:

Structure and motion recovery should proceed by first extracting features, and then using these features to compute the image matching relations. It should not proceed by simultaneously estimating motion and dense pixel correspondences.

The "image matching relations" referred to here arise from the camera motion alone, not from the scene structure. These relations are the part of the motion that can be computed directly from image correspondences. For example, if the camera translates between two views then the image matching relation is the epipolar geometry of the view-pair.

The rest of this paper demonstrates that there are cogent theoretical and practical reasons for advocating this thesis when attempting to recover structure *and* motion from images. We illustrate the use of feature based methods on two examples. First, in section 2, we describe in detail a feature based algorithm for registering multiple frames to compute a mosaic. The frames are obtained by a camera rotating about its centre, and the algorithm estimates the point-to-point homography map relating the views. Second, section 3 discusses the more general case of structure and motion computation from images obtained by a camera rotating *and* translating. Here it is shown how feature matching methods form the basis for a dense 3D reconstruction of the scene (where depth is obtained for

B. Triggs, A. Zisserman, R. Szeliski (Eds.): Vision Algorithms'99, LNCS 1883, pp. 278–294, 2000.

every pixel). Section 4 explores the strengths and weaknesses of feature based and direct methods, and summarises the reasons why feature based methods perform so well.

2 Mosaic Computation

Within this section the feature based approach to mosaicing is described. Given a sequence of images acquired by a camera rotating about its centre, the objective is to fuse together the set of images to produce a single panoramic mosaic image of the scene. For this particular camera motion, corresponding image points (i.e. projections of the same scene point) are related by a point-to-point planar homography map which depends only on the camera rotation and internal calibration, and does not depend on the scene structure (depth of the scene points). This map also applies if the camera "pans and zooms" (changes focal length whilst rotating).

A planar homography (also known as a plane projective transformation, or collineation) is specified by eight independent parameters. The homography is represented as a 3×3 matrix that transforms homogeneous image coordinates as:

$$\mathbf{x}' = \mathbf{H}\mathbf{x}.$$

We first describe the computation of a homography between an image pair, and then show how this computation is extended to a set of three or more images.

2.1 Image Pairs

The automatic feature based algorithm for computing a homography between two images is summarized in table 1, with an example given in figure 1.

The point features used (developed by Harris [12]) are known as interest points or "corners". However, as can be seen from figure 1 (c) & (d), the term corner is misleading as these point features do not just occur at classical corners (intersection of lines). Thus we prefer the term interest point. Typically there can be hundreds or thousands of interest points detected in an image.

It is worth noting two things about the algorithm. First, interest points are not matched purely using geometry – i.e. only using a point's position. Instead, the intensity neighbourhood of the interest point is also used to rank possible matches by computing a normalized cross correlation between the point's neighbourhood and the neighbourhood of a possible match. Second, robust estimation methods are an essential part of the algorithm: more than 40% of the putative matches between the interest points (obtained by the best cross correlation score and proximity) are incorrect. It is the RANSAC algorithm that identifies the correct correspondences.

Given the inlying interest point correspondences: $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}, i = 1 \ldots n$, the final estimate of the homography is obtained by minimizing the following cost function,

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \tag{1}$$

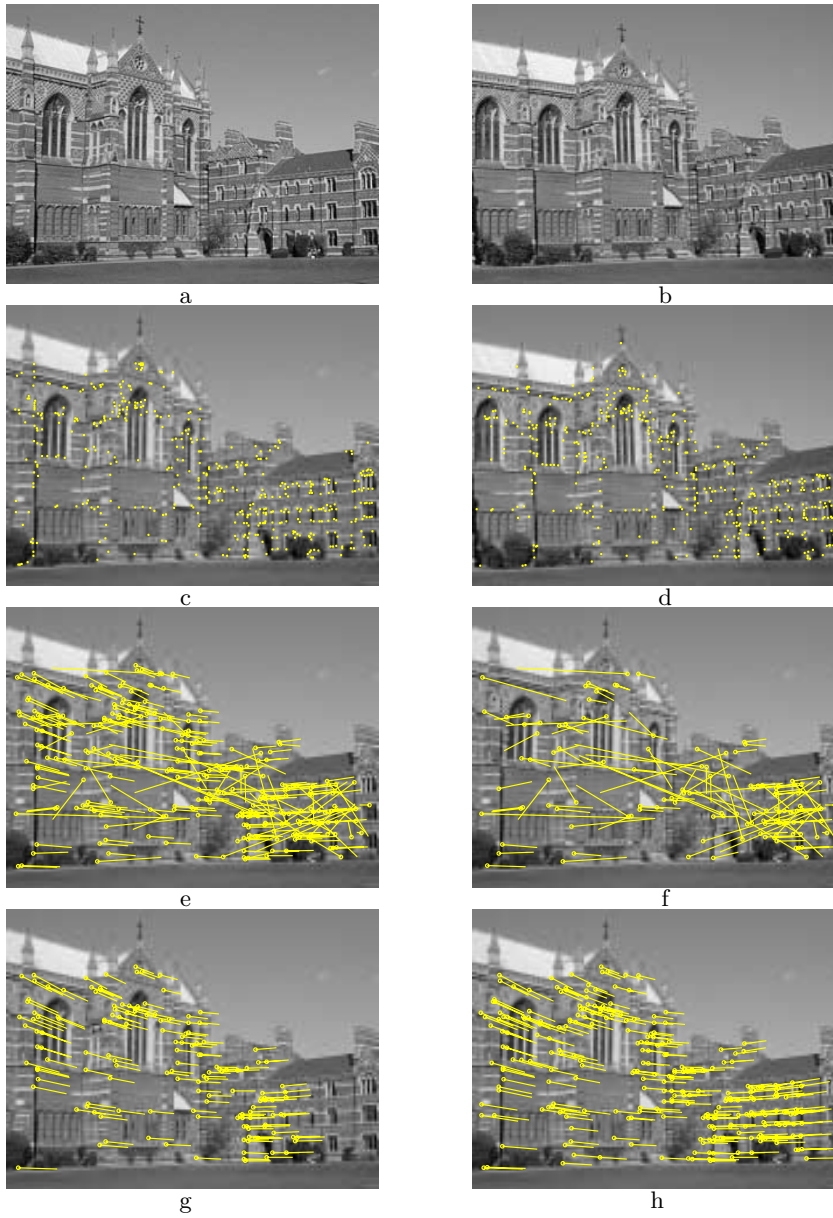


Fig. 1. Automatic computation of a homography between two images using the algorithm of table 1. *(a) (b) Images of Keble College, Oxford. The motion between views is a rotation about the camera centre so the images are exactly related by a homography. The images are* 640×480 *pixels. (c) (d) Detected point features superimposed on the images. There are approximately 500 features on each image. The following results are superimposed on the left image: (e) 268 putative matches shown by the line linking matched points, note the clear mismatches; (f) RANSAC outliers — 117 of the putative matches; (g) RANSAC inliers — 151 correspondences consistent with the estimated* $\mathbf{H}$*; (h) final set of 262 correspondences after guided matching and MLE. The estimated* $\mathbf{H}$ *is accurate to subpixel resolution.*

Table 1. *The main steps in the algorithm to automatically estimate a homography between two images using RANSAC and features. Further details are given in [14].*

Objective Compute the 2D homography between two images.

Algorithm

1. **Features:** Compute interest point features in each image to sub pixel accuracy (e.g. Harris corners [12]).
2. **Putative correspondences:** Compute a set of interest point matches based on proximity and similarity of their intensity neighbourhood.
3. **RANSAC robust estimation:** Repeat for N samples
   (a) Select a random sample of 4 correspondences and compute the homography $\mathbf{H}$.
   (b) Calculate a geometric image distance error for each putative correspondence.
   (c) Compute the number of inliers consistent with $\mathbf{H}$ by the number of correspondences for which the distance error is less than a threshold.
   Choose the $\mathbf{H}$ with the largest number of inliers.
4. **Optimal estimation:** re-estimate $\mathbf{H}$ from all correspondences classified as inliers, by minimizing the maximum likelihood cost function (1) using a suitable numerical minimizer (e.g. the Levenberg-Marquardt algorithm [24]).
5. **Guided matching:** Further interest point correspondences are now determined using the estimated $\mathbf{H}$ to define a search region about the transferred point position.

The last two steps can be iterated until the number of correspondences is stable.

where $d(\mathbf{x}, \mathbf{y})$ is the geometric distance between the image points $\mathbf{x}$ and $\mathbf{y}$. The cost is minimized over the homography $\widehat{\mathbf{H}}$ and corrected points $\{\hat{\mathbf{x}}_i\}$ such that $\hat{\mathbf{x}}'_i = \widehat{\mathbf{H}}\hat{\mathbf{x}}_i$. This gives the maximum likelihood estimate of the homography under the assumption of Gaussian measurement noise in the position of the image points. A fuller discussion of the estimation algorithm is given in [14] with variations and improvements (the use of MLESAC rather than RANSAC) given in [34].

2.2 From Image Ppairs to a Mosaic

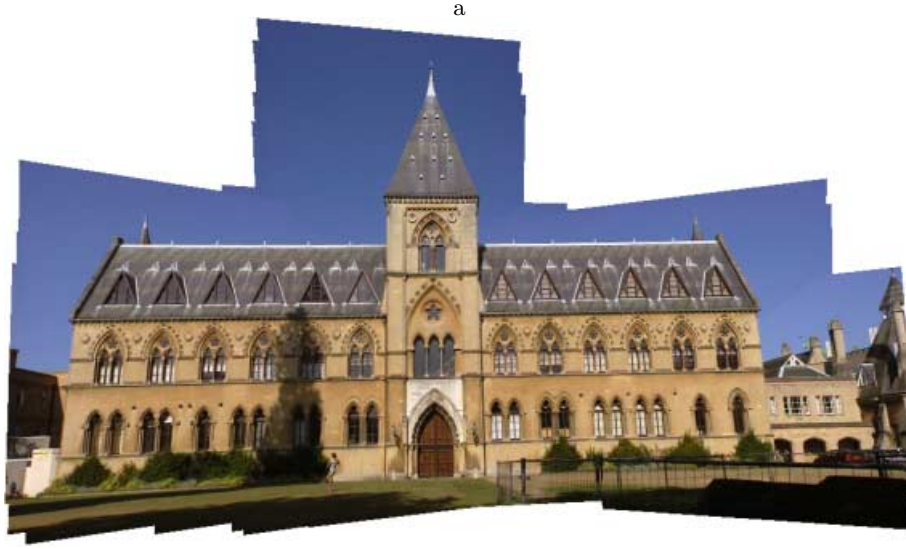
The two frame homography estimation algorithm can readily be extended to constructing a mosaic for a sequence as follows:

1. Compute interest point features in each frame.
2. Compute homographies and correspondences between frames using these point features.
3. Compute a maximum likelihood estimate of the homographies and points over all frames.
4. Use the estimated homographies to map all frames onto one of the input frames to form the mosaic.

In computing the maximum likelihood estimate the homographies are parametrized to be consistent across frames. So, for example, the homography between the first and third frame is obtained exactly from the composition of homographies between the first and second, and second and third as $\mathbf{H}_{13} = \mathbf{H}_{23}\,\mathbf{H}_{12}$. This is achieved by computing all homographies with respect to a single set of corrected points $\hat{\mathbf{x}}_i$. Details are given in [7].

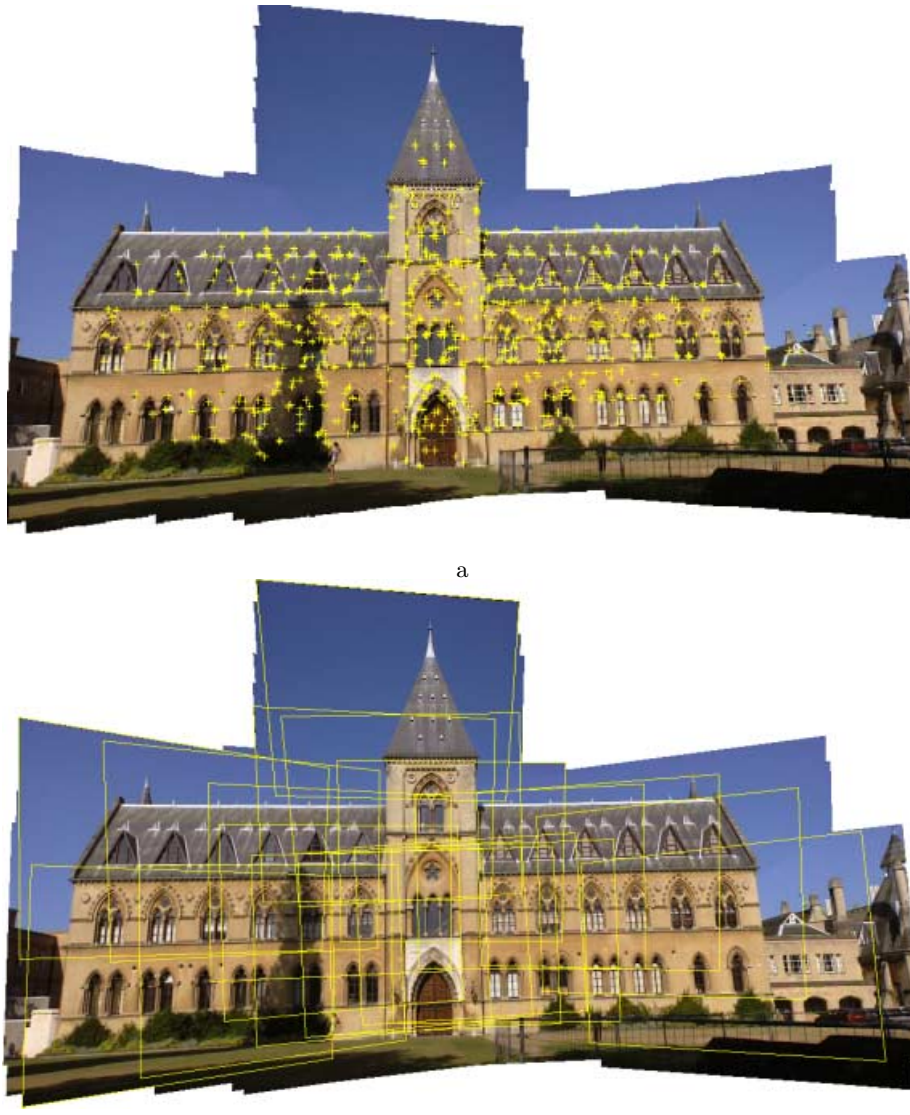
The application of this algorithm to a 100 frame image sequence is illustrated in figures 2 and 3. The result is a seamless mosaic obtained to subpixel accuracy.

a



b

Fig. 2. Automatic panoramic mosaic construction. (a) Every 10th frame of a 100 frame sequence acquired by a hand held cam-corder approximately rotating about its lens centre. Note, each frame has a quite limited field of view, and there is no common overlap between all frames. (b) The computed mosaic which is seamless, with frames aligned to subpixel accuracy. The computation method is described in [7].

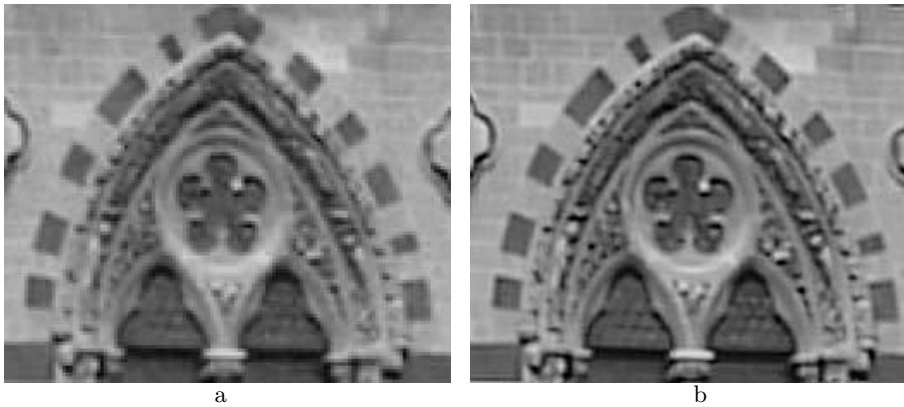


a

b

Fig. 3. Details of the mosaic construction of figure 2. (a) 1000 of the 2500 points used in the maximum likelihood estimation, note the density of points across the mosaic. (b) Every 5th frame (indicated by its outline), note again the lack of frame overlap. A super-resolution detail of this mosaic is shown in figure 4.

The mosaic can then form the basis for a number of applications such as video summary [17], motion removal [19], auto-calibration [13], and super-resolution. For example, figure 4 shows a super-resolution detail of the computed mosaic. The method used [8] is based on MAP estimation, which gives a slight improvement over the generally excellent Irani and Peleg algorithm [18].



a b

Fig. 4. Super-resolution detail of the mosaic of figure 2. The super-resolution image is built from a set 20 images obtained from partially overlapping frames. The original frames are jpeg compressed. (a) One of the set of images used as input for the super-resolution computation. The image is 130×110 pixels, and has the highest resolution of the 20 used. (b) The computed double resolution image (260×220 pixels). Note, the reduction in aliasing (e.g. on the dark bricks surrounding the Gothic arch) and the improvement in sharpness of the edges of the brick drapery. Details of the method are given in [8].

To summarize: in this case the "image matching relation" is a homography which is computed from point feature correspondences. Once the homography is estimated the correspondence of every pixel is determined.

3 Structure and Motion Computation

This section gives an example of (metric) reconstruction of the scene and cameras directly from an image sequence. This involves computing the cameras up to a scaled Euclidean transformation of 3-space (auto-calibration) and a dense model of the scene. The method proceeds in two overall steps:

1. Compute cameras for all frames of the sequence.
2. Compute a dense scene reconstruction with correspondence aided by the multiple view geometry.

Unlike in the mosaicing example, in this case the camera centre is moving and consequently the map between corresponding pixels depends on the depth

of the scene points, i.e. for general scene structure there is not a simple map (such as a homography) global to the image.

3.1 Computing Cameras for an Image Sequence

The method follows a similar path to that of computing a mosaic.

1. **Features:** Compute interest point features in each image.
2. **Multiple view point correspondences:** Compute two-view interest point correspondences and simultaneously the fundamental matrix $\mathbf{F}$ between pairs of frames, e.g. using robust estimation on minimal sets of 7 points, as described in [32]; Compute three-view interest point correspondences and simultaneously the trifocal tensor between image triplets, e.g. using robust estimation on minimal sets of 6 points, as described in [33]; Weave together these 2-view and 3-view reconstructions to get an initial estimation of 3D points and cameras for all frames [10]. This initial reconstruction provides the basis for bundle adjustment.
3. **Optimal estimation:** Compute the maximum likelihood estimate of the 3D points and cameras by minimizing reprojection error over all points. This is bundle adjustment and determines a projective reconstruction. The cost function is the sum of squared distances between the measured image points $\mathbf{x}_j^i$ and the projections of the estimated 3D points using the estimated cameras:
$$\min_{\hat{\mathbf{P}}^i, \hat{\mathbf{X}}_j} \sum_{ij} d(\hat{\mathbf{P}}^i \hat{\mathbf{X}}_j, \mathbf{x}_j^i)^2 \tag{2}$$
where $\hat{\mathbf{P}}^i$ is the estimated camera matrix for the ith view, $\hat{\mathbf{X}}_j$ is the jth estimated 3D point, and $d(\mathbf{x}, \mathbf{y})$ is the geometric image distance between the homogeneous points $\mathbf{x}$ and $\mathbf{y}$.
4. **Auto-calibration:** Remove the projective ambiguity in the reconstruction using constraints on the cameras such as constant aspect ratio, see e.g. [23].

Further details of automatic computation of cameras for a sequence are given in [1,3,4,10,14,22,30,31].

3.2 Computing a Dense Reconstruction

Given the cameras, the multi-view geometry is used to help solve for dense correspondences. There is a large body of literature concerning methods for obtaining surface depths given the camera geometry: a classical stereo algorithm may be used, for example an area based algorithm such as [9,20]; or space carving, e.g.[21,28]; or surface primitives may be fitted directly, e.g. piecewise planar models [2] or piecewise generalized cylinders [15]; or optical flow may be used, constrained by epipolar geometry [36].

Figure 5 shows an example of automatic camera recovery from five images, followed by automatic dense stereo reconstruction using an area based algorithm. The method is described in [23].

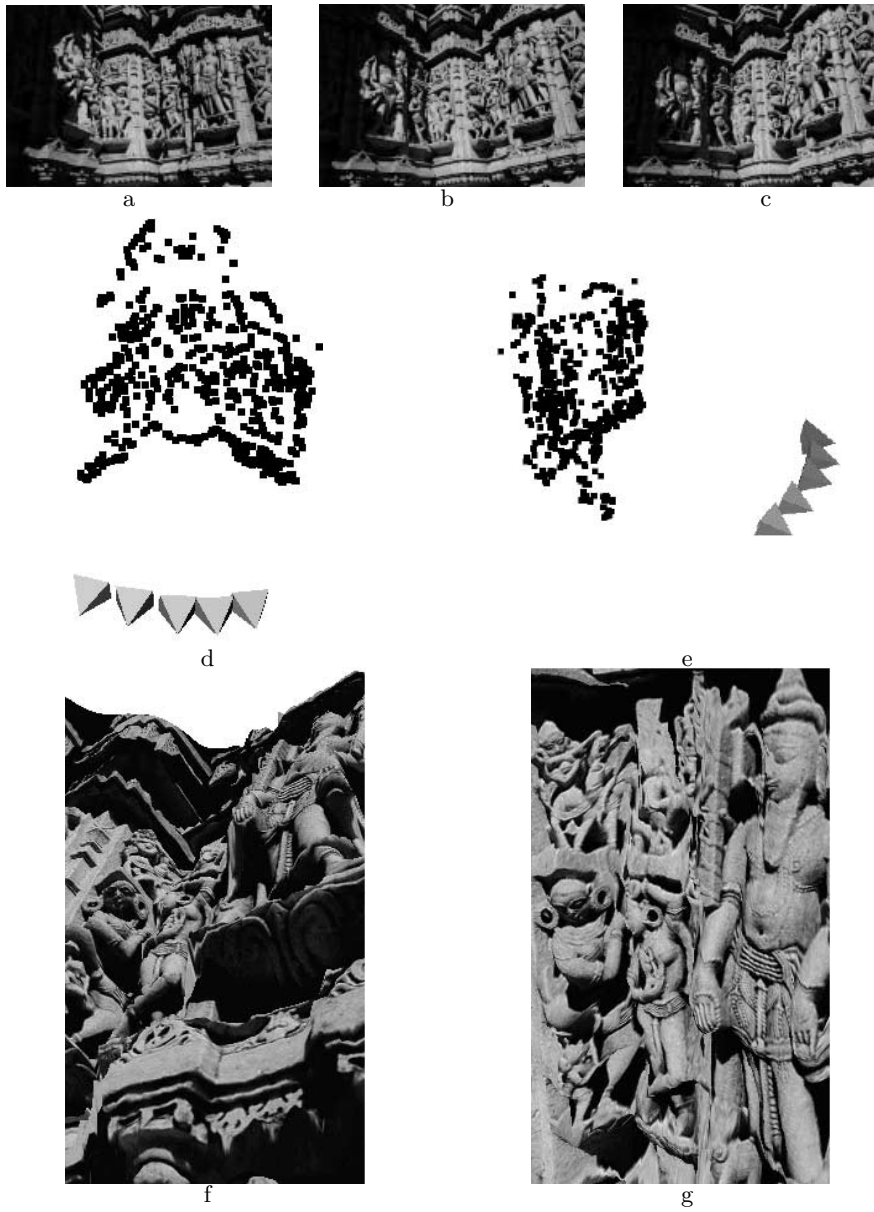


Fig. 5. Automatic generation of a texture mapped 3D model from an image sequence. The input is a sequence of images acquired by a hand held camera. The output is a 3D VRML model of the cameras and scene geometry. (a)–(c) three of the five input images. (d) and (e) two views of a metric reconstruction computed from interest points matched over the five images. The cameras are represented by pyramids with apex at the computed camera centre. (f) and (g) two views of the texture mapped 3D model computed from the original images and reconstructed cameras using an area based stereo algorithm. Figures courtesy of Marc Pollefeys, Reinhard Koch, and Luc Van Gool [23].

To summarize: the key point is that features are a convenient intermediate step from input images to dense 3D reconstruction. In this case the "image matching relations" are epipolar geometry, trifocal geometry, etc that may be computed from image interest point correspondences. This is equivalent to recovering the cameras up to a common projective transformation of 3-space. After computing the cameras, the features need not be used at all in the subsequent dense scene reconstruction.

4 Comparison of the Feature Based and Direct Methods

Within this section the two methods are contrasted. We highlight three aspects of the structure and motion recovery problem: invariance, optimal estimation, and the computational efficiency of the methods. Then list the current state of the art.

4.1 The Importance of Invariance

Features have a wide range of *photometric* invariance. For example, although thus far we have only discussed interest points, lines may be detected in an image as an intensity discontinuity (an "edge"). The invariance arises because a discontinuity is still detectable even under large changes in illumination conditions between two images. Features also have a wide range of *geometric* invariance – lines are invariant to projective transformations (a line is mapped to a line), and consequently line features may be detected under any projective transformation of the image.

In the case of Harris interest points, the feature is detected at the local minima of an autocorrelation function. This minima is also invariant to a wide range of photometric and geometric image transformations, as has been demonstrated by Schmid *et al* [27]. Consequently, the detected interest point across a set of images corresponds to the same 3D point.

This photometric and geometric invariance is perhaps the primary motivation for adopting a feature based approach.

In direct approaches it is necessary to provide a photometric map between corresponding pixels, for example the map might be that the image intensities are constant (image brightness constraint), or that corresponding pixels are related by a monotonic function. If this map is incorrect, then erroneous correspondences between pixels will result. In contrast in feature based methods a photometric map is used only to *guide* interest point correspondence.

As an example, consider normalized cross correlation on neighbourhoods. This measure is used in the algorithms included in this paper to *rank* matches. Normalized cross-correlation is invariant to a local affine transformation of intensities (scaling plus offset). If, in a particular imaging situation, the cross-correlation measure is not invariant to the actual photometric map between the images, then the ranking of the matches may be erroneous. However, the position of the interest points is (largely) invariant to this photometric map. Thus with

the immunity to mismatches provided by robust estimation, the transformation estimated from the interest points will still be correct. It is for this reason that the estimated camera geometry is largely unaffected by errors in the model of (or invariance to) the photometric map.

If normalized cross-correlation is used in a direct method, and is invariant to the actual photometric map, then in principle the correct pixel correspondence will be obtained. However, if it is not invariant to the actual photometric map, then direct methods may systematically degrade but feature based methods will not.

As a consequence, feature based methods are able to cope with severe viewing and photometric distortion, and this has enabled wide base line matching algorithms to be developed. An example is given in figure 6.

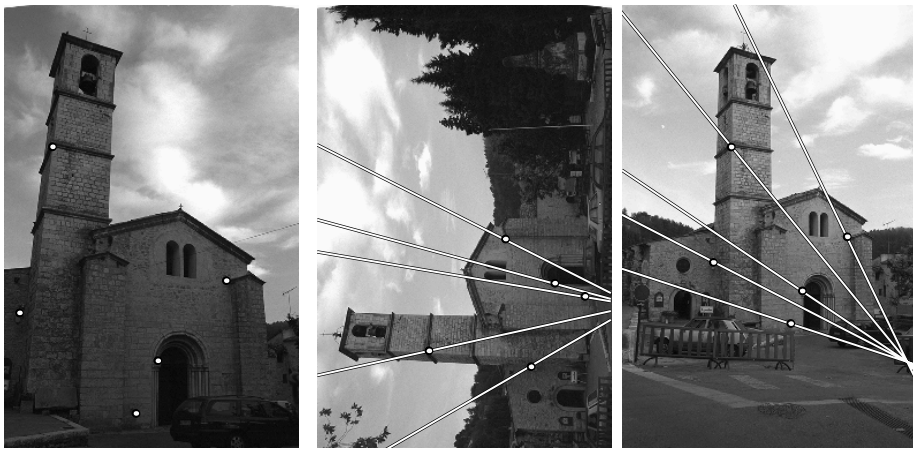


Fig. 6. Wide baseline matching. *Three images acquired from very different viewpoints with a hand-held camera. The trifocal tensor is estimated automatically from interest point matches and a global homography affinity score. Five of the matched points are shown together with their corresponding epipolar lines in the second and third images. The epipolar geometry is determined from the estimated trifocal tensor. Original images courtesy of RobotVis INRIA Sophia Antipolis. The wide baseline method is described in [25].*

4.2 Optimal Estimation

A significant advantage of the feature based approach is that it readily lends itself to a bundle adjustment method over a long sequence, and this provides a maximum likelihood estimate of the estimated quantities (homographies in the mosaic example, cameras in the structure and motion example). This reveals a key difference between the feature based and direct methods: in feature

based approaches the errors are uncorrelated between features so that statistical independence is a valid assumption in estimation.

Consider the "least squares" cost functions that are typically used (e.g. (2)). For this to be a valid maximum likelihood estimation two criteria must be satisfied: first, each of the squares to be summed must be the log likelihood of that error, and second each must be conditionally independent of any of the other errors. In the case of feature based methods the sum of squared error that is minimized is the distance between the backprojected reconstructed 3D point and its measured correspondence in each image. There is evidence that these errors are independent and distributed with zero mean in a Gaussian manner [35]. The same cannot be said when using the brightness constraint equation to estimate global motion models [5,11]. Because the quantities involved are image derivatives obtained by smoothing the image there is a large amount of conditional dependence between the errors. It is not clear what effect this violation of the conditions for maximum likelihood estimation might be but it is possible that the results produced may be biased.

To summarize: for direct methods it is not straightforward to write down a practical likelihood function for all pixels. Modelling of noise and statistics is much more complicated, and simple assumptions of independence invalid. Thus attempting a global minimization treating all the errors as if they were uncorrelated will lead to a biased result.

4.3 Computational Efficiency and Convergence

Consider computing the fundamental matrix from two views. Interest point correspondences yield highly accurate camera locations at little computational cost. If instead every pixel in the image is used to calculate the epipolar geometry the computational cost rises dramatically. Furthermore the result could not have been improved on as only pixels where the correspondence is well established are used (the point features). Use of every pixels means introducing much noisy data, as correspondence simply cannot be determined in homogeneous regions of the image either from the brightness constraint or from cross correlation. Thus the introduction of such pixels could potentially introduce more outliers, which in turn may cause incorrect convergence of the minimization. To determine correspondence in these regions requires additional constraints such as local smoothness.

To summarize: features can be thought of as a computational device to leap frog us to a solution using just the good (less noisy) data first, and then incorporating the bad (more noisy) data once we are near a global minima.

4.4 Scope and Applications

Finally we list some of the current achievements of feature based structure from motion schemes and ask how direct methods compare with this. A list of this sort will of course date, but it is indicative of the implementation ease and computational success of the two approaches.

Automatic estimation of the fundamental matrix and trifocal tensor. Point features facilitate automatic estimation of the fundamental matrix and trifocal tensor. There is a wide choice of algorithms available for interest points. These algorithms are based on robust statistics – this means that they are robust to effects such as occlusion and small independent motions in an otherwise rigid scene.

The fundamental matrix cannot be estimated from normal flow alone. The trifocal tensor can [29], but results comparable to the feature based algorithms have yet to be demonstrated. Direct methods can include robustness to minor occlusion and small independent motions. Although pyramid methods can be deployed to cope with larger disparities direct methods have met with only limited success with wide base line cases.

Application to image sequences. Features have enabled automatic computation of cameras for extended video sequences over a very wide range of camera motions and scenes. This includes auto-calibration of the camera. An example is shown in figure 7 of camera computation with auto-calibration for hundreds of frames.

In contrast direct methods have generally been restricted to scenes amenable to a "plane plus parallax" approach, i.e. where the scene is dominated by a plane so that homographies may be computed between images.

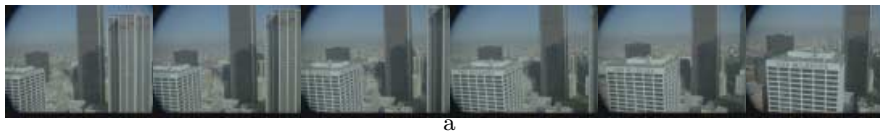
Features other than points. Although this paper has concentrated on interest point features, other features such as lines and curves may also be used to compute multi-view relations. For example, figure 8 shows an example of a homography computed automatically from an imaged planar outline between two views.

5 Conclusions

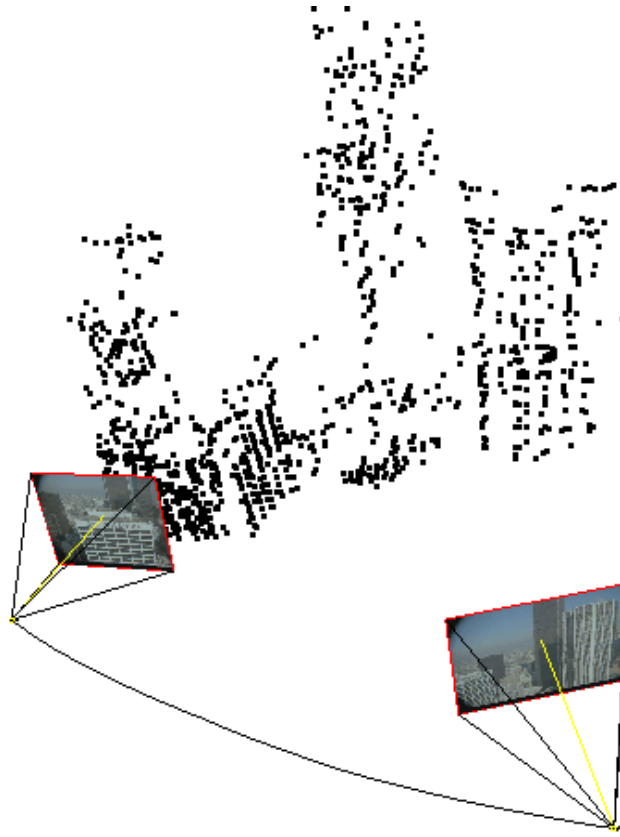
It is often said (by the unlearned) that feature based methods only furnish a sparse representation of the scene. **This is missing the point**, feature based methods are a way of initializing camera geometry/image matching relations so that a dense reconstruction method can follow.

The extraction of features – *the seeds of perception* [6] – is an intermediate step, a computational artifice that culls the useless data and affords the use of powerful statistical techniques such as RANSAC and bundle adjustment.

The purpose of this paper has not been to argue against the use of direct methods where appropriate (for instance in the mosaicing problem under small image deformations). Rather it has been to suggest that for more general structure and motion problems, the currently most successful way to proceed is via the extraction of photometrically invariant features. The benefit being that just a few high information features can be used to find the correct ball park of the solution. Once this is found more information may be introduced, and a "direct" method can be used to improve the result.



a

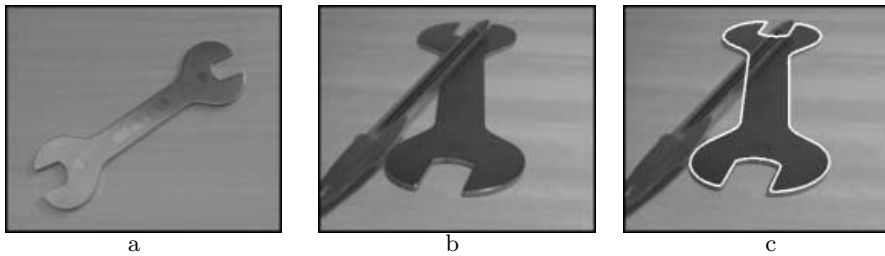


b

Fig. 7. Reconstruction from extended image sequences. (a) Six frames from 120 frames of a helicopter shot. (b) Automatically computed cameras and 3D points. The cameras are shown for just the start and end frames for clarity, with the path between them indicated by the black curve. The computation method is described in [10].

Acknowledgements

The mosaicing and super-resolution results given here were produced by David Capel, and the fundamental matrix and cameras for an image sequence by Andrew Fitzgibbon. We are very grateful to both of them.



a b c

Fig. 8. Computing homographies using curve features. *The homography between the plane of the spanner in (a) and (b) is computed automatically from segments of the outline curve. This curve is obtained using a Canny-like edge detector. Note the severe perspective distortion in (b). The mapped outline is shown in (c). The computation method is robust to partial occlusion and involves identifying projectively covariant points on the curve such as bi-tangents and inflections. Details are given in [26].*

References

1. S. Avidan and A. Shashua. Threading fundamental matrices. In *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*, pages 124–140, 1998.
2. C. Baillard and A. Zisserman. Automatic reconstruction of piecewise planar models from multiple views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 559–565, June 1999.
3. P. A. Beardsley, P. H. S. Torr, and A. Zisserman. 3D model aquisition from extended image sequences. In *Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, pages 683–695, 1996.
4. P. A. Beardsley, A. Zisserman, and D. W. Murray. Navigation using affine structure and motion. In *Proc. European Conference on Computer Vision*, LNCS 800/801, pages 85–96. Springer-Verlag, 1994.
5. J. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proc. European Conference on Computer Vision*, LNCS 588, pages 237–252. Springer-Verlag, 1992.
6. J. M. Brady. Seeds of perception. In *Proceedings of the 3rd Alvey Vision Conference*, pages 259–265, 1987.
7. D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara*, pages 885–891, June 1998.
8. D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Proc. International Conference on Pattern Recognition*, 2000.
9. I.J. Cox, S.L. Hingorani, and S.B. Rao. A maximum likelihood stereo algorithm. *Computer vision and image understanding*, 63(3):542–567, 1996.
10. A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. European Conference on Computer Vision*, pages 311–326. Springer-Verlag, June 1998.
11. K.J. Hanna and E. Okamoto. Combining stereo and motion analysis for direct estimation of scene structure. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 357–365, 1993.
12. C. J. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference, Manchester*, pages 147–151, 1988.

13. R. I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. European Conference on Computer Vision*, LNCS 800/801, pages 471–478. Springer-Verlag, 1994.
14. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521623049, 2000.
15. P. Havaldar and G. Medioni. Segmented shape descriptions from 3-view stereo. In *Proc. International Conference on Computer Vision*, pages 102–108, 1995.
16. M. Irani and P. Anandan. About direct methods. In *Vision Algorithms: Theory and Practice.* Springer-Verlag, 2000.
17. M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proc. 5th International Conference on Computer Vision, Boston*, pages 605–611, 1995.
18. M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4:324–335, 1993.
19. M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, 1994.
20. R. Koch. 3D surface reconstruction from stereoscopic image sequences. In *Proc. 5th International Conference on Computer Vision, Boston*, pages 109–114, 1995.
21. K. Kutulakos and S. Seitz. A theory of shape by space carving. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 307–314, 1999.
22. S. Laveau. *Géométrie d'un système de N caméras. Théorie, estimation et applications.* PhD thesis, INRIA, 1996.
23. M. Pollefeys, R. Koch, and L. Van Gool. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 90–96, 1998.
24. W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1988.
25. P. Pritchett and A. Zisserman. Matching and reconstruction from widely separated views. In R. Koch and L. Van Gool, editors, *3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506*, pages 78–92. Springer-Verlag, June 1998.
26. C. Rothwell, A. Zisserman, D. Forsyth, and J. Mundy. Planar object recognition using projective shape representation. *International Journal of Computer Vision*, 16(2), 1995.
27. C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Proc. International Conference on Computer Vision*, pages 230–235, 1998.
28. S.M. Seitz and C.R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*, pages 1067–1073, 1997.
29. G. Stein and A. Shashua. Model-based brightness constraints: on direct estimation of structure and motion. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 400–406, 1997.
30. P. Sturm. *Vision 3D non calibrée: Contributions à la reconstruction projective et étude des mouvements critiques pour l'auto calibrage.* PhD thesis, INRIA Rhône-Alpes, 1997.
31. P. H. S. Torr, A. W. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *International Journal of Computer Vision*, 32(1):27–44, August 1999.

32. P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.
33. P. H. S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997.
34. P. H. S. Torr and A. Zisserman. Robust computation and parameterization of multiple view relations. In *Proc. 6th International Conference on Computer Vision, Bombay, India*, pages 727–732, January 1998.
35. P. H. S. Torr, A. Zisserman, and S. Maybank. Robust detection of degenerate configurations for the fundamental matrix. *Computer Vision and Image Understanding*, 71(3):312–333, September 1998.
36. J. Weber and J. Malik. Rigid body segmentation and shape description from dense optical flow under weak perspective. In *Proc. International Conference on Computer Vision*, pages 251–256, 1995.