3D Model Acquisition from Extended Image Sequences

Paul Beardsley, Phil Torr and Andrew Zisserman
Dept of Engineering Science, University of Oxford, Oxford OX1 3PJ

Abstract. A method for matching image primitives through a sequence is described, for the purpose of acquiring 3D geometric models. The method includes a novel robust estimator of the trifocal tensor, based on a minimum number of token correspondences across an image triplet; and a novel tracking algorithm in which corners and line segments are matched over image triplets in an integrated framework. The matching techniques are both robust (detecting and discarding mismatches) and fully automatic.
The matched tokens are used to compute 3D structure, which is initialised as it appears and then recursively updated over time. The approach is uncalibrated - camera internal parameters and camera motion are not known or required.
Experimental results are provided for a variety of scenes, including outdoor scenes taken with a hand-held camcorder. Quantitative statistics are included to assess the matching performance, and renderings of the 3D structure enable a qualitative assessment of the results.

1 Introduction

The aim of this work is to recover 3D models from long uncalibrated monocular image sequences. These models will be used for graphics and virtual reality applications. The sequences are generated by circumnavigating the object of interest (e.g. a house) acquiring images with a camcorder. Neither the internal calibration of the camera nor the motion of the camera are known. In particular the motion is unlikely to be smooth. The focus of this paper is on the matching and tracking of image primitives which underpins the structure recovery.

We build on the work of a number of previous successful systems which have recovered structure and motion from tracked image primitives (tokens). Coarsely these systems can be divided into those that use sequential [1, 3, 4, 10, 19, 28], and those that use batch updates [13, 15, 23]. Here the matching and structure update are sequential. However, the basic unit is an image triplet, rather than the more usual image pair. It is in this area of tracking technology [26, 27] that we have made the most significant innovations. A finite displacement (several cms) between views prohibits the utilisation of simple nearest neighbour token matching strategy between consecutive images. In this work:

1. Corner and line segments are matched simultaneously over image triplets in an integrated framework, by employing the trifocal tensor [14, 21, 22].
2. A robust estimator for the trifocal tensor is developed, with the tensor instantiated over three views using a minimal set (six) of point matches and a RANSAC scheme.

The use of a robust scheme provides protection against mismatches [6, 24] and independently moving objects [25].

Two important advantages of the method described here are that the camera model covers a full perspective projection, not its affine approximation (weak or para-perspective) as in [23], and no knowledge of camera internal parameters or relative motion is required. However, a consequence is that the 3D structure recovered is up a projective transformation, rather than Euclidean [7, 12].

Results are presented for a variety of real scenes, with an assessment of matching performance (lifetime of tracked tokens, total number of matches), and examples of the recovered structure. All of the processing (matching, structure recovery etc), is automatic, involving no hand picked points.

Notation and multiple view geometry

The representations of multiple view geometry are based on [3, 7, 8, 12, 14, 16].

For a triplet of images the image of a 3D point $\mathbf{X}$ is $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{x}''$ in the first, second and third images respectively, and similarly the image of a line is $\mathbf{l}$, $\mathbf{l}'$ and $\mathbf{l}''$, where $\mathbf{x} = (x_1, x_2, x_3)^\mathsf{T}$ and $\mathbf{l} = (l_1, l_2, l_3)^\mathsf{T}$ are homogeneous three vectors.

Image pairs - bilinear relations Corresponding points in two images satisfy the epipolar constraint

$$\mathbf{x}'^\mathsf{T}\mathbf{F}\mathbf{x} = 0 \tag{1}$$

where $\mathbf{F}$ is the 3×3 fundamental matrix, with maximum rank 2. This is the bilinear relation in the homogeneous coordinates of the corresponding points in two images.

Image triplets - trilinear relations Corresponding points in three images, and corresponding lines in three images, satisfy trilinear relations which are encapsulated in the trifocal tensor, $\mathtt{T}$, a $3 \times 3 \times 3$ homogeneous tensor. Using the tensor a point can be transferred to a third image from correspondences in the first and second:

$$x''_l = x'_i \sum_{k=1}^{k=3} x_k \mathtt{T}_{kjl} - x'_j \sum_{k=1}^{k=3} x_k \mathtt{T}_{kil},$$

for all $i, j = 1 \ldots 3$. Similarly, a line can be transferred as

$$l_i = \sum_{j=1}^{j=3} \sum_{k=1}^{k=3} l'_j l''_k \mathtt{T}_{ijk}$$

i.e. the same tensor can be used to transfer both points and lines.

2 Matching strategies

Simple matching based only on similarity of image primitive attributes will inevitably produce mismatches. For image *pairs*, the fundamental matrix provides a constraint for identifying mismatches between image corners: corresponding corners are constrained to lie on (epipolar) lines. For *triplets* of images, the trifocal tensor provides a more powerful constraint for identifying mismatches for both points and lines: a primitive matched in two images defines the position of the corresponding primitive in the third image. It is a more powerful constraint because the position of a match is completely constrained, rather than just restricted to a line, and also because it applies to both corners and lines, rather than just corners. There is a natural symbiosis between a 2-image matching scheme and the robust computation of the fundamental matrix, and also between a 3-image matching scheme and the robust computation of the trifocal tensor.

In the following subsections we describe three robust matching schemes applicable to a camera moving through a scene that is largely static. No *a priori* information on camera internal parameters or motion is assumed, other than a threshold on the maximum disparity between images. The methodology for matching is essentially the same for all three schemes and involves three distinct stages. The stages are motivated in the description of the first matching scheme.

2.1 Matching corners between image pairs

1. **Seed correspondences by unguided matching**
   The aim is to obtain a small number of reliable seed correspondences. Given a corner at position (x, y) in the first image, the search for a match considers all corners within a region centred on (x, y) in the second image with a threshold on maximum disparity. The strength of candidate matches is measured by cross-correlation. The threshold for match acceptance is deliberately conservative at this stage to minimise incorrect matches.
2. **Robust computation of a geometric constraint**
   There is potentially a significant presence of mismatches amongst the seed matches. Correct matches will obey a geometric constraint, in this case the epipolar geometry. The aim then is to obtain a set of "inliers" consistent with the geometric constraint using a robust technique — RANSAC has proved the most successful [6, 9, 17]: A putative fundamental matrix (up to three solutions) is computed from a random set of seven corner correspondences (the minimum number required to compute a fundamental matrix). The support for this fundamental matrix is determined by the number of correspondences in the seed set within a threshold distance of their epipolar lines. This is repeated for many random sets, and the fundamental matrix with the largest support is accepted. The outcome is a set of corner correspondences consistent with the fundamental matrix, and a set of mismatches (outliers). The fundamental matrix is then re-estimated using all of its associated inliers to improve its accuracy.

3. **Guided matching**
   The aim here is to obtain additional matches consistent with the geometric constraint. The constraint provides a far more restrictive search region than that used for unguided matching. Consequently, a less severe threshold can be used on the matching attributes. In this case, matches are sought for unmatched corners searching only epipolar lines. This generates a larger set of consistent matches.

The final two steps are repeated until the number of matches stabilises.

Typically the number of corners in a 512 × 512 image of an indoor scene is about 300, the number of seed matches is about 100, and the final number of matches is about 200. Using corners computed to sub-pixel accuracy, the typical distance of a point from its epipolar line is ~0.2-0.4 pixels.

2.2 Matching points and lines between image triplets

The same three steps are used over image triplets, with the geometric constraint provided by the trifocal tensor.

1. **Seed correspondences by unguided matching**
   For lines, seed correspondences over the three images are obtained by matching on a number of attributes (see [2]). For corners, seed correspondences between images 1 & 2, and 2 & 3 are obtained using the fundamental matrix as described above.
2. **Robust computation of a geometric constraint**
   A full description of this method is given in section 3. Briefly, a putative trifocal tensor (up to three solutions) is computed from a random set of six seed point correspondences. The putative tensor is evaluated by measuring its support in the seed set, utilising both corners and lines. The tensor with the greatest support is chosen, and re-estimated using its consistent point and line correspondences. Inconsistent matches are assumed to be mismatches and are marked as unmatched.
3. **Guided matching**
   Corner and line matching is resumed, but now with a far more restrictive search area — for a putative match between a pair of tokens, only a region about the predicted position in the third image need be searched. This generates a larger set of consistent matches.

Here, both points and lines contribute to the estimate of the geometric constraint, and in turn the one constraint is used to search for both corner and line correspondences. In this manner matched corners provide support for line segment matching and vice-versa.

Typically the number of seed matches over a triplet is about 100 corners, and 10-20 lines. The final number of matches is about 150 and 10-50 respectively. Using corners computed to sub-pixel accuracy, the typical distance of a corner/line from its transferred position is ~1.0 pixel.

2.3 Matching between image primitives and 3D structure

The previous two matching schemes were for image to image matching. Once an estimate of 3D structure is available however (at any stage in the image sequence after the initialisation phase is completed) then it is possible to use the 3D structure to aid the matching. This augmented scheme is carried out whenever a new image arrives, to obtain matches between the last image of the sequence and the new image. The result provides both token matches in the images, and also a correspondence between existing 3D structure and tokens in the new image.

1. **Seed correspondences by unguided matching**
   As in the matching of corners between image pairs, section 2.1.
2. **Robust computation of a geometric constraint**

   a. Robust computation of the fundamental matrix
   As in the matching of corners between image pairs, section 2.1.
   b. Robust computation of the camera matrix
   The set of matches obtained above provide a correspondence between the existing 3D point structure and the new image corners. RANSAC is used to compute the 3×4 camera matrix $\mathtt{P}$, which projects the 3D points onto their correspondences in the new image. A putative projection matrix is computed from a random sample of six correspondences. Support for this matrix is given by those correspondences whose projection lies within a threshold distance of the associated image primitive. The projection matrix with greatest support is re-estimated using all consistent point and line matches. Inconsistent matches are assumed to be mismatches and are marked as unmatched.
3. **Guided matching**
   Corner matching is resumed. $\mathtt{P}$ is used to project any unmatched 3D points and lines onto the new image, and a match is searched for around the projected position. This generates a larger set of consistent matches.

Typically, we find that the majority of matches are obtained in the initial matching stage when the fundamental matrix is used. However, the use of the camera matrix computation can add 5-10 matches in a total of 200. The r.m.s. error between projected 3D structure and actual image tokens in the new image is $\sim$0.2-0.5 pixels.

2.4 Implementation details

Two types of image primitives are used - corners and line segments - extracted independently in each image. Corners are detected to sub-pixel accuracy using the Harris corner detector [11]. Lines are detected by the standard procedure of: Canny edge detection [5]; edge linking; segmentation of the chain at high curvature points; and finally, straight line fitting to the resulting chain segments. The straight line fitting is by orthogonal regression, with a tight tolerance to ensure that only actual line segments are extracted, i.e. that curves are not piecewise linear approximated. Further implementation details are given in [2].

2.5 Comparison of pairwise and triplet based matching

We compare two schemes for matching between images:

1. **Method 1: Pairwise based** Corners are matched between image pairs 1 & 2 and 2 & 3 as in section 2.1; 3D point structure is instantiated from the matches between 1 & 2; based on the 2 & 3 matches, the matrix P which projects the 3D structure to image 3 is computed as in section 2.3.
2. **Method 2: Triplet based** This is the matching scheme described in section 2.2. In this scheme both corners and lines are matched.

Experiment I: Number of matches/mismatches for image triplets We assess and compare the two matching schemes by two measures — the number of matches, and the number of mismatches.

Figure 1 shows three consecutive images of a model house, processed by the 2-image and 3-image schemes, with matched corners superimposed on the images. For the 2-image scheme, only those points which survived over all three images are shown, to enable a proper comparison with the 3-image approach. There is little difference in the distribution of the matches found. Furthermore, there are no mismatches under either scheme. Figure 2 shows the same information for an outdoor scene. In this case, there are a few mismatches under both schemes.

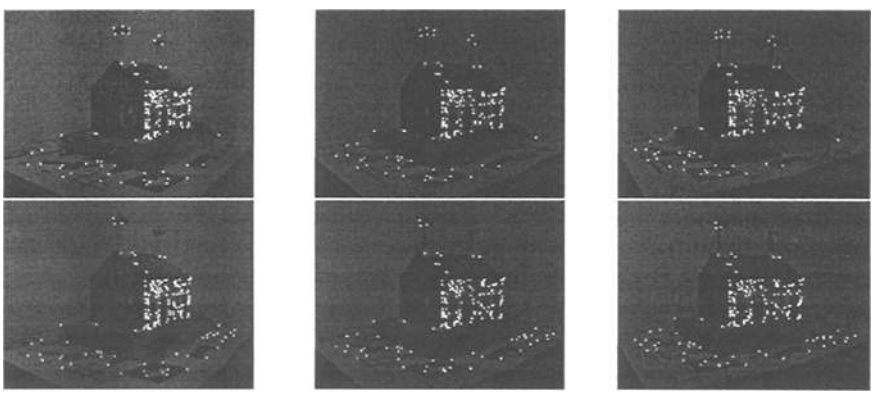

Fig. 1. Three images from a sequence of a model house. Camera motion is a lateral displacement of about 3-4cm between frames. The image size is 760x550 pixels, and about 400 corners are detected in each image. **Upper 3 images:** In the 2-image scheme, about 200 matches are obtained between each pair. The r.m.s. perpendicular distance of points from epipolar lines is about 0.4 pixels. About 160 matches survive across all three frames. R.m.s. error between projected 3D points and corresponding image corners is about 0.5 pixels. **Lower 3 images:** In the 3-image scheme, about 180 matches are obtained across all three images. The r.m.s. error between transferred points (using the trifocal tensor) and actual points in the third image is about 1.0 pixel. R.m.s. projected error is again about 0.5 pixels.

The 2-image matching scheme gives rise to some image matches which exist only between image 1-2, or image 2-3. This can be because a number of the proposed matches (actually mismatches) accidentally agree with the epipolar

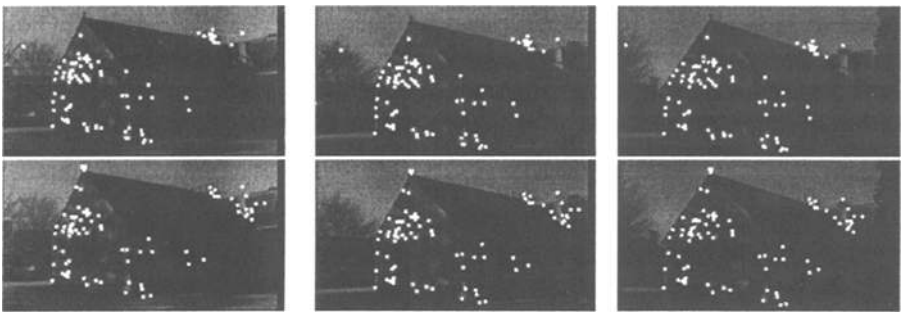

Fig. 2. Three images from a sequence of a chapel, acquired by a hand-held camcorder. Camera motion is lateral and a few centimetres between frames. Image size and corner count as in Figure 1. **Upper 3 images:** In the 2-image scheme, about 150-200 matches are obtained between each pair, with r.m.s. distance of points to epipolar lines about 0.4 pixels. About 80 matches survive across all three frames. **Lower 3 images:** In the 3-image scheme, the number of tokens matched across all three images is again about 80. The r.m.s. error between transferred points (using the trifocal tensor) and actual points in the third image is about 1.5 pixels. R.m.s. projected error is about 0.5 pixels.

geometry. (Other reasons for matches existing only between image 1-2 or image 2-3 are that the corner detector does not continue to find the point, or the physical point moves out of view.) A mismatch which accidentally agrees with the epipolar geometry generates a meaningless 3D point, which cannot project to a potential match in the third image, so the corner is not matched across the triplet. Figure 3 shows the full set of matches for images 1-2 of the outdoor scene (a superset of the matches in Figure 2, all consistent with the estimated fundamental matrix), and the mismatches present in this set. Epipolar mismatches of this type occur particularly in an area of texture which gives many corners of similar appearance along the epipolar line (such as trees and bushes in outdoor scenes).

Fig. 3. For the sequence in Figure 2 under the 2-image matching scheme, the left image shows the full set of matches obtained between the first two images. The right image shows those matches which are in actuality outliers. They are congregated mainly on the ambiguous texture area created by the trees in the background, and are accepted by the 2-image matching because they accidentally agree with the epipolar geometry. Only one of these outliers survives when a third image is processed, however - see Figure 2.

Figure 4 indicates how the robust trifocal tensor computation in the 3-image matching scheme enables direct identification of mismatches in the set of seed point and line matches.

Fig. 4. Results for matching a triplet of images using the trifocal tensor, superimposed onto the last image. Top **left:** point correspondences over the three images, with small squares indicating the previous positions. **centre:** matches consistent with the computed trifocal tensor, **right:** outlying matches (mismatched points). At bottom, the same information for line matches. In all, there are 101 seed point correspondences over the three images, 76 are indicated as inliers, and 25 as outliers. There are 15 seed line correspondences, 11 are indicated as inliers, and 4 as outliers.

In summary, the experimental results do not suggest that the 3-image matching scheme produces a marked improvement over the 2-image scheme. There are still good reasons for favouring the 3-image approach, however. Firstly, it is computationally much more elegant and efficient to use the trifocal tensor to match over three images and eliminate mismatches, rather than creating 3D structure from the first two images and then projecting it to the third. Secondly, mismatches in the 2-image scheme which accidentally agree with the epipolar geometry are only detected after processing has moved to the third image. By this stage, it is cumbersome to return to the first two images, remove the mismatch and attempt to rematch the points. Furthermore, the mismatches may have adversely affected the computation of the fundamental matrix, leading to missed matches. In contrast, the 3-image scheme offers the possibility of detecting suspect pairwise matches immediately by using the trifocal tensor. Finally, the 3-image approach and the trifocal tensor allow the integrated use of points and lines, unlike the 2-image approach where corners drive the processing.

Experiment II: Track Lifetime The image matches are extended over a sequence as follows. For the 2-image scheme, the method for images 2 & 3 described above is extended naturally to images n & $n+1$ using the structure computed from n images. For the 3-image scheme, matches between images images $n-1$, n & $n+1$ are generated using the robust estimation of the trifocal tensor for this triplet.

Figure 5 shows two comparisons of the 2-image and 3-image schemes, in terms of overall matching statistics along a sequence. As expected, the total number of matches is generally higher for the 2-image scheme because it includes transient corners which only appear for 2 images, as well as a small number of mismatches which accidentally agree with the epipolar geometry. The 2-image scheme also has more long surviving matches; this is because the 3-image scheme applies a stronger geometric constraint, and so is less tolerant of localisation errors in the features.

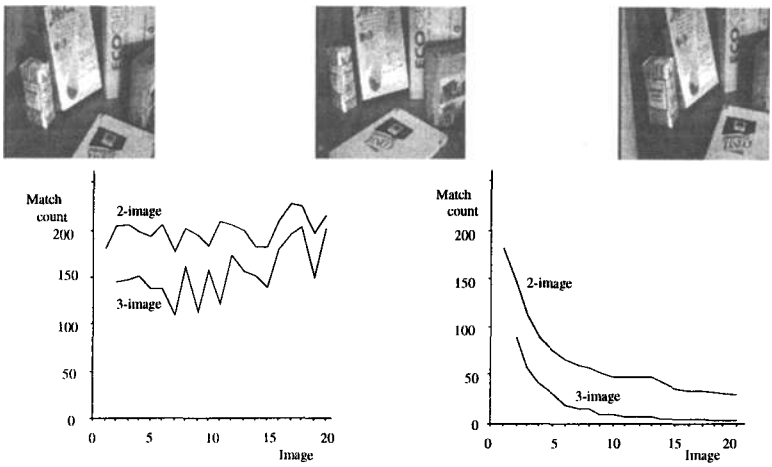

Fig. 5. Three images from a sequence of twenty, taken as the camera moves while fixating a set of objects. The left graph shows the total number of matches at each stage, as an indicator of overall performance. The right graph shows the number of matches which have been tracked continuously from the start of the sequence at each point.

3 Robust computation of the trifocal tensor

In this section we describe the robust computation of the trifocal tensor from a set of putative corner and line correspondences over three images, obtained as described in the previous section. The trifocal tensor has 27 elements, but only their ratio is significant, leaving 26 that must be specified. Each triplet of point correspondences provides four independent linear equations for the elements of the tensor, and each triplet of line correspondences provides two linear equations. Therefore provided that $2n_l + 4n_p \geq 26$ (where n_l is the number of lines, and n_p is the number of points), the tensor can be determined uniquely (up to scale) using a linear algorithm. Consequently, the tensor can be computed linearly from a minimum of 7 points or 13 lines or a combination of the two. However, the tensor has only 18 independent degrees of freedom, which means that it can be computed from 6 point correspondences, though not uniquely — there are 1 or 3 solutions, according to the number of real roots of an associated cubic.

For random sampling methods, such as Least Median Squares (LMS) [20] or RANSAC it is extremely important that the minimum number of correspondences are used, so as to reduce the probability of a mismatch being included in the random sample of correspondences, furthermore a six point solution will be exact, having only 18 degrees of freedom in the coefficients. This is why the novel six point solution used here is important. It is not a problem that there are three solutions, since the correct solution of the three is identified when measuring support for the solution from the full set of putative matches. The method for finding the trifocal tensor from six points uses the theory of Quan [18] for computing an invariant of six points from 3 views, and is described in [2].

3.1 Comparison of six and seven point robust schemes

Ideally every possible subsample of the full set of putative matches would be considered, but this is usually computationally infeasible, so m the number of samples, is chosen sufficiently high to give a probability Υ in excess of 95% that a good subsample is selected. The expression for this probability Υ is [20]

$$\Upsilon = 1 - (1 - (1 - \epsilon)^p)^m, \tag{2}$$

where ϵ is the fraction of contaminated data, and p the number of tokens in each sample. Table 1 gives some sample values of the number m of subsamples required to ensure $\Upsilon \geq 0.95$ for given p and ϵ. It can be seen that the smaller the data set needed to instantiate a model, the less samples are required for a given level of confidence. If the fraction of data that is contaminated is unknown, as is usual, an educated worst case estimate of the level of contamination must be made in order to determine the number of samples to be taken. It can be seen that as the proportion of outliers increases many more samples need to be taken for the seven point algorithm then for the six point method.

Features	Fraction of Contaminated Data, ϵ						
p	5%	10 %	20 %	25 %	30 %	40 %	50 %
6	3	4	10	16	24	63	191
7	3	5	13	21	35	106	382

Table 1. *The number m of subsamples required to ensure $\Upsilon \geq 0.95$ for given p and ϵ, where Υ is the probability that all the data points selected in one subsample are non-outliers.*

The six point algorithm gives better results than the seven point algorithm described in [25], when tested on both real data and synthetic data (where the ground truth is known). This is for two reasons, the first being that that the six point algorithm requires fewer correspondences, and so has less chance of including an outlier as evinced by Table 1; the second, and perhaps more important, is that the six point algorithm exactly encodes the constraints on the parameters of the trifocal tensor. The seven point algorithm on the other hand has too many degrees of freedoms, 27, when there should only be 18. This means that the tensor is over parameterised and a least squares solution will usually give a result that violates the constraints on its coefficients leading to a solution that is implicitly incorrect.

The six point algorithm is also considerably faster. In the case of the seven point algorithm the eigenvector of a 27×27 matrix must be found, which is slower than the solution of a cubic. Furthermore far fewer six point samples need to be taken to get a given degree of confidence in the result.

4 Structure Recovery

Camera matrices are generated at the start of a sequence, using either the fundamental matrix in a 2-image scheme, or the trifocal tensor in a 3-image scheme, and matched corners and line segments are then used to instantiate estimates of 3D point and line structure. An update process is subsequently employed for each new image added to the sequence. Matching between the last image and the new image provides a correspondence between existing 3D structure and the new image primitives, enabling computation of the camera matrix P for the new image. Once P is known, existing estimates of 3D structure are updated from the new observations using an Extended Kalman Filter. Then, P is recomputed using a non-linear computation which minimises the squared distance between the projection of the updated 3D structure and the actual image observations on the image plane for the new image. New points and line segments in 3D are initialised whenever new structure becomes visible in the course of the sequence.

4.1 Results

This section contains experimental results for the estimated 3D structure. The structure is "Quasi-Euclidean" (a form which is close to being true Euclidean) which can be obtained given approximate knowledge of the camera internal parameters as described in [3].

Figures 6 and 7 show results for the sequence of a model house, and the outdoor scene of a chapel. Point and line structure is shown. The recovered structure is best illustrated by using Delaunay triangulation to obtain image connectivity of the structure, and then mapping image intensity onto the triangular facets in 3D. Lines significantly improve the rendering since they often mark boundaries of object planes.

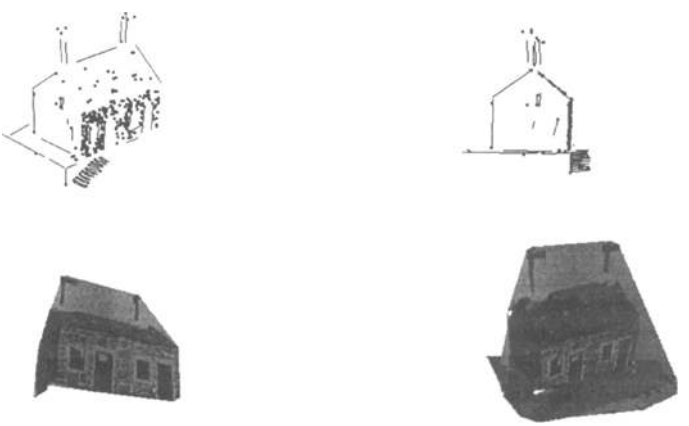

Fig. 6. At top is the point and line structure recovered for the model house of figure 1. Top-right shows the front wall and roof viewed edge on. The bottom images are obtained by rendering image intensity onto the 3D structure and viewing it from novel viewpoints (viewpoints which were never seen in the original sequence).

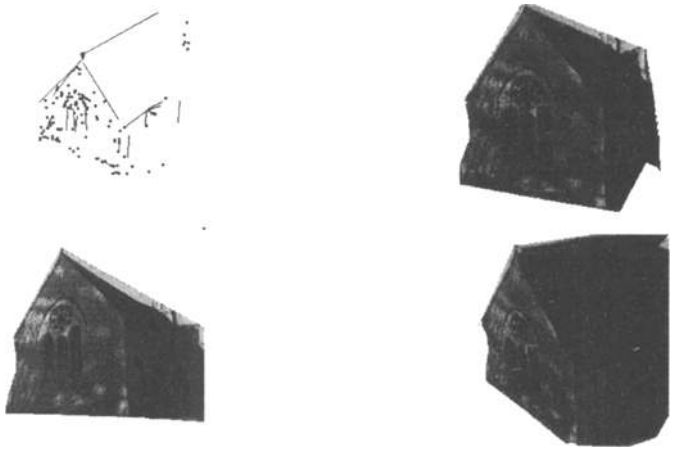

Fig. 7. Results for the outdoor scene of a chapel of Figure 2. Details are as for the previous figure. The dappling effect on the front of the chapel is sunshine through trees.

Acknowledgements
Thanks to Phil McLauchlan for the images in Figure 5. Financial support for this work was provided by EU ACTS Project VANGUARD.

References

1. N. Ayache and F. Lustman. Fast and reliable passive trinocular stereovision. *Proc. International Conference on Computer Vision*, 1987.
2. P.A. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. Technical report, Dept of Eng Science, University of Oxford, 1996.
3. P.A. Beardsley, A.P. Zisserman, and D.W. Murray. Navigation using affine structure and motion. In *Proc. 3rd European Conference on Computer Vision*, pages 85–96. Springer-Verlag, 1994.
4. P.A. Beardsley, A.P. Zisserman, and D.W. Murray. Sequential update of projective and affine structure from motion. To appear in IJCV, 1996.
5. J.F. Canny. A computational approach to edge detection. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 8:769–798, 1986.
6. R. Deriche, Z. Zhang, Q.T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *Proc. 3rd European Conference on Computer Vision*, pages 567–576. Springer-Verlag, 1994.
7. O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. 2nd European Conference on Computer Vision*, pages 563–578. Springer-Verlag, 1992.
8. O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In E. Grimson, editor, *Proc. 5th International Conference on Computer Vision*, pages 951–956, Cambridge, MA, June 1995.
9. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, vol. 24:381–95, 1981.

10. C.G. Harris and J.M. Pike. 3D positional integration from image sequences. In *Third Alvey Vision Conference*, pages 233–236, 1987.
11. C.G. Harris and M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pages 147–151, 1988.
12. R. Hartley. Invariants of points seen in multiple images. GE internal report, to appear in PAMI, GE CRD, Schenectady, NY 12301, USA, 1992.
13. R.I. Hartley. Euclidean reconstruction from uncalibrated views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of invariance in computer vision*, pages 237–256. Springer-Verlag, 1994.
14. R.I. Hartley. A linear method for reconstruction from points and lines. In E. Grimson, editor, *Proc. 5th International Conference on Computer Vision*, pages 882–887, Cambridge, MA, June 1995.
15. R. Mohr, F. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. *Proc. Conference Computer Vision and Pattern Recognition*, pages 543–548, 1993.
16. J.L. Mundy and A.P. Zisserman. *Geometric invariance in computer vision.* MIT Press, 1992.
17. L. Quan. Invariants of 6 points from 3 uncalibrated images. In J. O. Eckland, editor, *Proc. 3rd European Conference on Computer Vision*, pages 459–469. Springer-Verlag, 1994.
18. L. Robert, M.Buffa, and M.Hebert. Weakly-calibrated stereo perception for robot navigation. In E. Grimson, editor, *Proc. 5th International Conference on Computer Vision*, pages 46–51, Cambridge, MA, June 1995.
19. P.J. Rousseeuw. *Robust Regression and Outlier Detection.* Wiley, New York, 1987.
20. A. Shashua. Trilinearity in visual recognition by alignment. In *Proc. 3rd European Conference on Computer Vision*, pages 479–484. Springer-Verlag, 1994.
21. M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, pages 171–183, 1990.
22. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorisation method. *International Journal of Computer Vision*, pages 137–154, 1992.
23. P.H.S. Torr. *Motion segmentation and outlier detection.* PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.
24. P.H.S. Torr, A. Zisserman, and D.W. Murray. Motion clustering using the trilinear constraint over three views. In R. Mohr and C. Wu, editors, *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision*, pages 118–125. Springer–Verlag, 1995.
25. P.H.S.Torr and D.W.Murray. Outlier detection and motion segmentation. In *SPIE 93*, 1993.
26. X.Hu and N.Ahuja. Matching point features with ordered geometric, rigidity and disparity constraints. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 16(10):1041–1048, 1994.
27. C. Zeller and O. Faugeras. Projective, affine and metric measurements from video sequences. In *Proceedings of the International Symposium on Optical Science, Engineering and Instrumentation*, 1995.
28. Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis.* Springer-Verlag, 1992.