Automatic Line Matching across Views

Cordelia Schmid and Andrew Zisserman
Department of Engineering Science, University of Oxford
Parks Road, Oxford, UK OX1 3PJ

Abstract

This paper presents a new method for matching individual line segments between images. The method uses both greylevel information and the multiple view geometric relations between the images. For image pairs epipolar geometry facilitates the computation of a cross-correlation based matching score for putative line correspondences. For image triplets cross-correlation matching scores are used in conjunction with line transfer based on the trifocal geometry. Algorithms are developed for both short and long range motion. In the case of long range motion the algorithm involves evaluating a one parameter family of plane induced homographies. The algorithms are robust to deficiencies in the line segment extraction and partial occlusion.

Experimental results are given for image pairs and triplets, for varying motions between views, and for different scene types. The three view algorithm eliminates all mismatches.

1. Introduction

The goal of this paper is the automatic matching of line segments between images of scenes mainly containing planar surfaces. A typical example is an urban scene. Line matching is often the first step in the reconstruction of such scenes.

Line matching is a difficult problem for several reasons. The first is due to the deficiencies in extracting lines [6] and their connectivity: although the orientation of a line segment can be recovered accurately, the end points are not reliable, and furthermore the topological connections between line segments are often lost during segmentation. Some segmentation algorithms are more successful than others [13] but the problem remains. The second reason is that there is no strong disambiguating geometric constraint available: In the case of points (corners), correspondences must satisfy the epipolar constraint. For infinite lines there is no geometric constraint, whilst for lines of finite length there is only a weak overlap constraint arising from applying the epipolar constraint to end points.

Existing approaches to line matching in the literature are of two types: those that match individual line segments; and those that match groups of line segments. Individual line segments are generally matched on their geometric attributes — orientation, length, extent of overlap [1, 12, 20]. Some such as [4, 5, 10] use a nearest line strategy which is better suited to image tracking where the images and extracted segments are similar.

The advantage of matching groups of line segments is that more geometric information is available for disambiguation, the disadvantage is the increased complexity. A number of methods have been developed around the idea of graph-matching [2, 7, 9, 19]. The graph captures relationships such as left of, right of, cycles, collinear with etc, as well as topological connectedness. Although such methods can cope with more significant camera motion, they often have a high complexity and again they are sensitive to error in the segmentation process. These methods are complementary to the approach in this paper which is for matching individual line segments.

The approach in this paper is built on two novel ideas. The first is to exploit the intensity neighbourhood of the line. The use of affinity measures based on cross-correlation of intensity neighbourhoods has been very successful in disambiguating corner matches [21]. However, there are two problems with applying correlation directly to line neighbourhoods: first, the point to point correspondence is unknown; and second, corresponding neighbourhoods may well have a very different shape and orientation, and this is also unknown. For example, suppose a square neighbourhood in one image back-projects to a planar facet on one side of the line. The image of this region in the second image is a quadrilateral, but its shape depends entirely on the relative positioning of the cameras and plane. Even a (significant) rotation or scaling will defeat naive cross-correlation based on square neighbourhoods of the same orientation. The second novel part of our approach solves these problems: The epipolar geometry between the images can be used to provide point to point correspondences along the line segments. Further, the epipolar geometry, together with the matched lines, restricts the possible homographies

(projective transformations) between the images to a one-parameter family, and this family can be used to solve for the neighbourhood mapping. The algorithm thus delivers a correlation score between line segments which can be used to discriminate between correct and false matches. The implementation is robust to the instabilities of the extraction process, and to partial occlusion.

1.1. Overview

The paper is organised as follows. Two algorithms are developed for automatic line matching. The first, described in section 2, is applicable to "short range motion". This is the image motion that arises in image sequences where simple nearest neighbour tracking would almost work. The second, described in section 3, is applicable to "long range motion". This is the image motion that arises between views from a stereo rig, where the baseline is significant (compared to the distance to the scene). There may be significant rotation of the line between the images, and, more importantly, planar surfaces may have significantly different foreshortenings in the two images. The performance of both algorithms are discussed and examples given using real image pairs.

Section 4 describes the extension of the algorithms when more than two views are available. With three views there is a strong geometric constraint available for line matching. The trifocal tensor [8, 16, 17] enables lines matched in two views to be transferred to a third, and this process can be used to verify two view matches. An alternative is to treat the three views symmetrically and match simultaneously over the three. Results are given for a triplet of aerial images which show that all mismatches can be eliminated for image triplets.

1.2. Implementation details

Line segments are extracted by applying a local implementation of the Canny edge detector with hysteresis. Edgels are then linked into chains, jumping up to a one pixel gap. Tangent discontinuities in the chain are located using a worm, and line segments are then fitted between the discontinuities using orthogonal regression. A very tight threshold is used for the line fitting so that curves are not piecewise linear approximated.

The geometric relations between the images required *a priori* here are the fundamental matrix for image pairs, and the trifocal tensor for image triplets. These relations are either calculated indirectly from known camera projection matrices for each view, or directly, and automatically, from point (corner) correspondences [3, 18, 21].

2 Short range motion

In the case of short range motion, line segments can be compared using (uncorrected) correlation. The basic idea is to treat each segment as a list of points to which neighbourhood correlation is applied as a measure of similarity. Only the point to point correspondence is required. In the absence of any other knowledge corresponding points could be obtained by searching along each line segment with a winner takes all matching strategy, similar to that used for matching corners on epipolar lines. However, knowing the epipolar geometry determines the point correspondences, as will now be described. Also, the epipolar geometry reduces the overall search complexity because it restricts which line segments need to be considered for matching.

2.1. F-guided matching

Corresponding image points, represented as homogeneous 3-vectors $\mathbf{x}$ and $\mathbf{x}'$, satisfy the epipolar constraint $\mathbf{x}'^{\top}\mathbf{F}\mathbf{x} = 0$. $\mathbf{F}$ is the fundamental matrix, which is a 3×3 matrix of rank 2. The epipolar line corresponding to $\mathbf{x}$ is $\mathbf{l}'^{e} = \mathbf{F}\mathbf{x}$, and the epipolar line corresponding to $\mathbf{x}'$ is $\mathbf{l}^{e} = \mathbf{F}^{\top}\mathbf{x}'$. Note, lines are also represented by homogeneous 3-vectors, and $'$ in all cases indicates the second image.

Suppose two image lines, $\mathbf{l}$ and $\mathbf{l}'$ correspond (i.e. have the same pre-image in 3-space) then the epipolar geometry generates a point-wise correspondence between the lines. A point $\mathbf{x}$ on $\mathbf{l}$ corresponds to the point $\mathbf{x}'$ which is the intersection of $\mathbf{l}'$ and the epipolar line $\mathbf{l}'^{e}$ of $\mathbf{x}$: The point $\mathbf{x}' = \mathbf{l}' \times \mathbf{l}'^{e} = \mathbf{l}' \times (\mathbf{F}\mathbf{x})$. This construction is valid provided $\mathbf{l}$ is not an epipolar line.

The matching score for a pair of line segments $\mathbf{l}$ and $\mathbf{l}'$ is computed as the average of the individual correlation scores for the points (pixels) of the line. The only points included in this average are those that are common to both line segments, i.e. the correlation is not extended past the ends of the measured segments.

For each segment in one image, a matching score is computed for all segments of the second image (within the search space, see below). The pair of segments with the best score is retained as the correct match, i.e. a winner takes all scheme.

The epipolar geometry can also be used to reduce the search space. The two end-points of a segment generate two epipolar lines in the other image. These two lines define a region, called the epipolar "beam", which necessarily intersects or contains the corresponding segment [20]. Given a segment in one image, this reduces the complexity of the search for corresponding segments.

2.2 Experimental results

Results are given for aerial images of an urban scene and for images of a toy house. At present the ground-truth matches are assessed by hand. A correlation window of 15×15 is used, and only lines of length 15 pixels or more are considered.

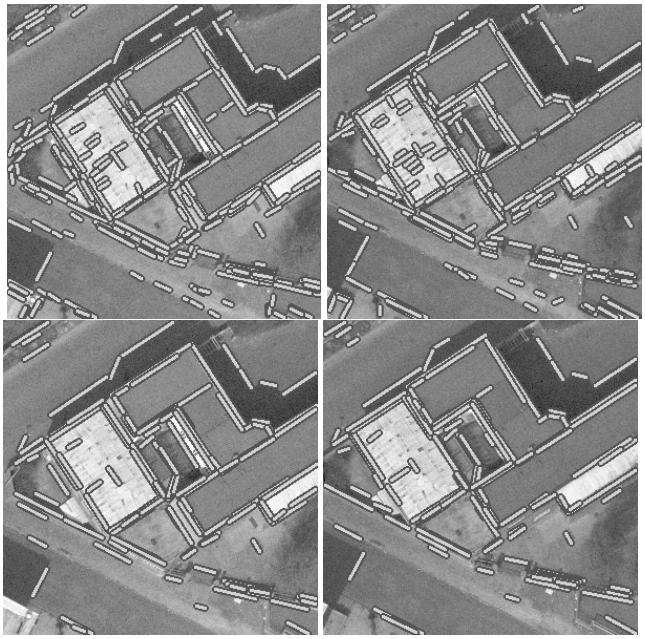

Figure 1. Upper pair: extracted lines segments superimposed on the left/right images; Lower pair: matched segments using the short range motion algorithm. 97.5% of the 122 matches shown are correct.

Figure 1 (upper) shows the line segments extracted on the aerial images using the method described in the introduction. 248 and 236 segments are obtained for the left and right images, respectively. The short range motion algorithm produces the matches displayed in the lower figure. 97.5% of the 122 matches obtained are correct.

For the toy house example, the number of segments extracted is 120 and 135 for the left and right images. Matches obtained by the short range algorithm are shown in figure 2. 94.5% of the 73 matches are correct.

As an example of the matching scores, the correct match between lines a in figure 2 has a correlation score of 0.92, compared to the incorrect match of a to b with a score of -0.67, i.e. a significant difference. Note also that segments b are matched correctly despite their different segmented lengths in the two images.

On average the epipolar beam reduces the search complexity to about a third, i.e. only a third of the line segments need be considered. The beam constraint is also used in the long range motion algorithm described in the following section.

The examples demonstrate that very good results are obtained in the case of short range motion. Nevertheless, the method as implemented will fail when the correlation measure is no longer sufficiently discriminating to distinguish correct from false matches. Imagine that the motion consists of a small lateral motion (as above) followed by a large rotation about the optical centre (cyclo-rotation). Such an image motion will defeat the cross-correlation affinity measure because it is not invariant to rotations. There are (at least) two solutions to this: first, the use of a rotationally invariant correlation measure — such a measure has been developed by [15] in the case of corner matching; second, using the orientation of the epipolar lines to determine an in plane rotation to compensate for the cyclo-rotation. We have not investigated these solutions yet. However, neither solution can overcome the failure of cross-correlation when there are significant foreshortening effects between the two images, and this is the subject of the long range motion algorithm described next.

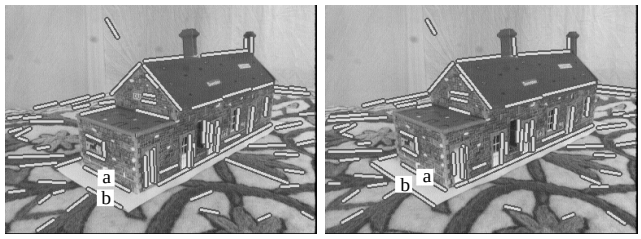

Figure 2. Matched line segments using the short range motion algorithm. 94.5% of the 73 segments shown are matched correctly.

3 Long range motion

In the case of large deformation between images, correlation will fail if the correlation patch is not corrected. This correction is achieved by a projective warping using a homography computed from the fundamental matrix. The geometric basis for this is now introduced.

The correspondence of lines between images determines a one parameter family of homographies which map the line in one image to the line in the other image, and which are also consistent with the epipolar geometry i.e. are homographies that could have arisen from images of planes in 3-space: Given the fundamental matrix between two views, 3D structure can be determined from image correspondences up to a projective ambiguity of 3-space. The correspondence of two image lines determines a line in 3-space, and a line in 3-space lies on a one parameter family (a pencil) of planes, see figure 3. This pencil of planes induces a pencil of homographies between the two images which map the corresponding lines to each other.

The assumptions underpinning the use of homographies here is that the scene is approximated locally by a plane or junctions of planes. This approximation is generally valid and in the case of images of rooms or aerial images the homography is often exact.

3.1 Computing the planar homography

Luong and Vieville [11] show that the homography (planar projective transformation) between two images induced

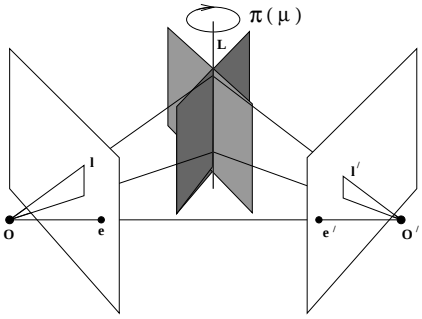

Figure 3. Image lines $\mathbf{l}$ and $\mathbf{l}'$ determine a line $\mathbf{L}$ in 3-space. The line $\mathbf{L}$ is contained in a one parameter family of planes $\boldsymbol{\pi}(\mu)$. This family of planes induces a one parameter family of homographies between the images.

by a world plane $\boldsymbol{\pi}$ is given by

$$\mathbf{H}(\mathbf{a}) = [\mathbf{e}']_\times \mathbf{F} + \mathbf{e}'\mathbf{a}^\top \tag{1}$$

such that the images of points on $\boldsymbol{\pi}$ are related by $\mathbf{x}' = \mathbf{H}\mathbf{x}$; $=$ indicates equality up to scale; $\mathbf{H}$ is the 3×3 homogeneous matrix representing the homography; $\mathbf{e}'$ is the epipole in the second view ($\mathbf{F}^\top\mathbf{e}' = \mathbf{0}$); $[\mathbf{e}']_\times$ is the skew 3×3 matrix representing the vector product (i.e. $[\mathbf{e}']_\times\mathbf{x} = \mathbf{e}' \times \mathbf{x}$); and $\mathbf{a}$ is a 3-vector which parameterises the 3-parameter family of planes in 3-space.

If the 3×4 camera projection matrix for the first view has the canonical form $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$, i.e. the world coordinate frame is aligned with the first camera, then the world plane is represented by the four vector $\boldsymbol{\pi}^\top = (\mathbf{a}^\top, 1)$.

Given the correspondence of image lines, $\mathbf{l}, \mathbf{l}'$, the homographies induced by planes containing the line in 3-space are reduced from a 3-parameter to a one-parameter family. Under a homography a line transforms as $\mathbf{l} = \mathbf{H}^\top\mathbf{l}'$. Imposing this relation on the homographies of equation (1) we obtain, after a short calculation [14],

$$\mathbf{H}(\mu) = [\mathbf{l}']_\times \mathbf{F} + \mu\mathbf{e}'\mathbf{l}^\top \tag{2}$$

where μ is the single scalar parameter of the pencil.

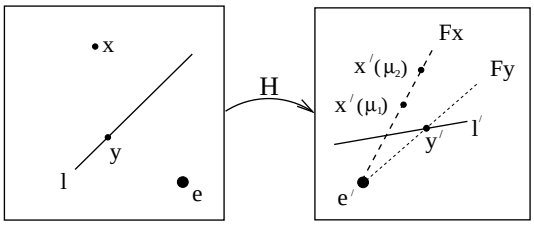

Figure 4. As μ varies, a point $\mathbf{x}$, which is not on the line $\mathbf{l}$, is mapped by $\mathbf{H}(\mu)$ to a point $\mathbf{x}'(\mu)$ which moves along the epipolar line ($\mathbf{F}\mathbf{x}$) corresponding to $\mathbf{x}$. However, the point $\mathbf{y}$, which lies on $\mathbf{l}$, is mapped to a fixed point $\mathbf{y}'$ for all values of μ.

The behaviour of $\mathbf{H}(\mu)$ as μ varies is illustrated in figure 4. Since $\mathbf{H}(\mu)$ is compatible with the epipolar geometry, all members of the family map the epipoles to each other, i.e. $\mathbf{e}' = \mathbf{H}(\mu)\mathbf{e}$) and corresponding epipolar lines are mapped to each other i.e. if $\mathbf{x}$ and $\mathbf{x}'$ are corresponding points — not necessarily lying on the plane $\boldsymbol{\pi}(\mu)$ — and $\mathbf{l}'^e, \mathbf{l}^e$ their corresponding epipolar lines, then $\mathbf{l}^e = \mathbf{H}^\top(\mu)\mathbf{l}'^e$.

Points on $\mathbf{l}$ and $\mathbf{l}'$ are mapped to a point on the line in the other image under the one dimensional homography induced by the epipolar lines, i.e. $\mathbf{y}' = \mathbf{H}(\mu)\mathbf{y} = \mathbf{l}' \times (\mathbf{F}\mathbf{y})$. Points which are not on $\mathbf{l}$ and $\mathbf{l}'$ are mapped to points on their corresponding epipolar lines as illustrated in figure 4.

Once the value of the parameter μ is known, a matching score can be computed using pixel based cross-correlation with the point correspondences provided by $\mathbf{H}(\mu)$.

3.2 H-correlation score

Given a putative line match the aim is to compute a cross-correlation score using $\mathbf{H}(\mu)$. A single score is obtained by computing the value μ^*, with corresponding homography $\mathbf{H}^*$, for which the cross-correlation is highest over all μ. In the following we first describe the computation of the $\mathbf{H}$-correlation score, and then the estimation of μ^*.

As straight line segments often occur at the junction of planar facets, the homographies are in general different for the two sides of the line. It is therefore necessary to process the two sides separately. The maximum of the two correlation scores is used as the matching score for the line.

The cross-correlation is evaluated for a rectangular strip on one side of the line segment by using $\mathbf{H}(\mu)$ to associate corresponding points. The length of the strip is the common overlap of the lines determined by the epipolar geometry. The area of the strip must be sufficient to include neighbouring texture, otherwise the cross-correlation will not be discriminating. In the implementation a strip of width 14 pixels was found to be sufficient. The computation is carried out to sub-pixel resolution using bilinear interpolation.

The homography $\mathbf{H}^*$ which maximizes the cross-correlation must then be estimated. This involves estimating μ^*. However μ is a projective parameter and is not directly measurable or meaningful in the image. Instead of μ the homography is parametrized by the mapping of a single point (which in turn determines μ). The corner of the rectangular correlation strip is ideal for this purpose. The set of possible correspondences in the other image $\mathbf{x}'(\mu)$ lie on the epipolar line of $\mathbf{x}$ (cf. figure 4). The value μ^* is obtained by searching the epipolar line for the $\mathbf{x}'(\mu^*)$ which maximizes the cross-correlation. Consider the point $\mathbf{x}'$ which has the same distance to the line $\mathbf{l}'$ as $\mathbf{x}$ has to $\mathbf{l}$. This corresponds to a scale/foreshortening factor of one between the images. To restrict the search we limit the possible scale factors to the range 1/3 to 3. This range defines an interval on the epipolar line in the second image. The correlation score is then evaluated at 10 equi-spaced points on this interval, and

the best score determines μ^*.

3.3 Experimental results

Figure 5 shows experimental results using the long range motion algorithm. For this pair of aerial images there is a significant rotation between the images. 93% of the 55 matches obtained are correct. In the second example, there

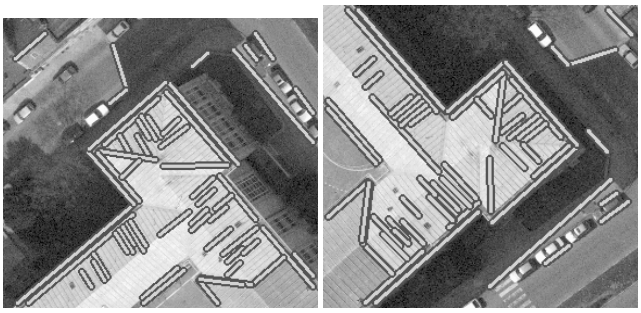

Figure 5. Matched line segments using the long range motion algorithm. There is a significant rotation between the images. 93% of the 55 matches shown are correct.

are significant foreshortening effects between the two images. Figure 6 displays the 53 matches obtained. 77% of them are correctly matched.

Figure 6. Matched line segments using the long range motion algorithm. There is significantly different foreshortening between the planes of the house in the two images. 77% of the 53 lines shown are correctly matched.

4 Three view matching

The trifocal geometry provides a geometric constraint for corresponding lines over three views. Lines $\mathbf{l}_1$, $\mathbf{l}_2$ and $\mathbf{l}_3$ in three views are corresponding if they are the image of the same line in 3-space. Given the trifocal tensor and corresponding lines in two images, the corresponding line in the third image is determined. This constraint provides a method for verifying the two image matches determined by the short or long range algorithms. This "two plus one" method proceeds as follows: The putative two view matches predict a line in the third image (via the trifocal tensor). There are then two stages of verification. First, a geometric verification, a line segment should be detected at the predicted position in the third image. Second, an intensity neighbourhood verification, the pairwise correlation algorithms should support this match. In practice these two stages of verification eliminate all mismatches.

An alternative matching procedure is to compute matches simultaneously over three views and then verify. This procedure is outlined in the following section. It is demonstrated in [14] that this procedure generates more correct matches than the two plus one approach, without adding any mismatches.

4.1 Matching over three images simultaneously

All line triplet combinations are considered (subject to the epipolar beam constraint). There are then three stages of verification. First it is verified that the (infinite) lines satisfy the trifocal constraint. If this geometric constraint is satisfied, then the trifocal tensor (or equivalently, the three pairwise fundamental matrices) determines the common parts of the three segments. The second stage examines this common segment. If there is no common part then the match is rejected. Otherwise, we have now verified that the match is geometrically correct for the finite segments. The third stage is to compute line correlation scores as described in sections 2 and 3 between line pairs $\mathbf{l}_1/\mathbf{l}_2$ and $\mathbf{l}_2/\mathbf{l}_3$. If these two score are sufficiently large then the triplet is a potential match. Cases where there are multiple matches possible for a particular line segment are resolved by a winner takes all scheme.

It might be thought that the three view method would have a significantly higher complexity, but this cost can be largely avoided by an initial correlation based pre-filter on view pairs to remove ridiculous matches and thus restrict putative matches for each line to a small number. Details are given in [14].

The reason that the three view matching method includes more matches than the two plus one method is that in the latter method a winner takes all scheme is applied after only two view matching. This earlier application of winner takes all may eliminate some correct matches which could have been verified in the third view.

4.2 Experimental Results

The result for three aerial images using matching over three views with short range motion correlation is given in figure 7. All of the 89 matches obtained are correct. Figure 8 shows the 3D reconstruction obtained using these matches. The position of the line in 3D is determined using a bundle adjustment which minimises the re-projection error over the three images.

5 Conclusion and Extensions

We have demonstrated two algorithmic approaches which significantly improve on the state of the art for line matching across two or more images. The algorithms cover the cases of both short and long range motion. The long

Figure 7. Matching across three views using both line transfer and a short range matching correlation score. All of the 89 matches obtained are correct.

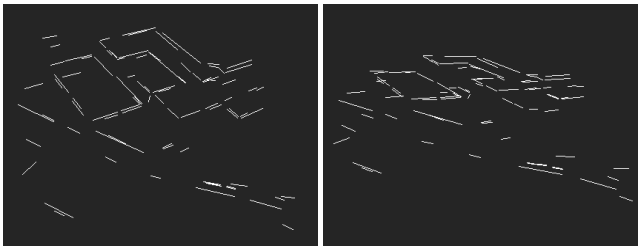

Figure 8. Two views of the 3D reconstruction of the line matches from figure 7.

range algorithm will work equally well in the short range case, of course, but is more expensive. Although we have not investigated the choice, it is likely that the process that generates the fundamental and trifocal tensors will have sufficient information to choose which of the short or long range algorithms is appropriate.

Finally, we mention three extensions which we are currently investigating. The first is the use of vanishing points to reduce matching complexity. The second is to use μ as a line grouping constraint — since coplanar lines lie on planes with the same value of μ. The third extension is to apply the same ideas (intensity neighbourhoods and local homographies) to curve matching, both for plane curves and space curves.

Acknowledgements

We are very grateful to Andrew Fitzgibbon for both discussions and software. Financial support for this work was provided by EU Esprit Project IMPACT.

References

[1] N. Ayache. *Stereovision and Sensor Fusion.* MIT-Press, 1990.

[2] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *IJCV*, 1987.

[3] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. *ECCV*, 1996.

[4] J. Crowley and P. Stelmazyk. Measurement and integration of 3D structures by tracking edges lines. *ECCV*, 1990.

[5] R. Deriche and O. Faugeras. Tracking line segments. *ECCV*, 1990.

[6] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint.* MIT-Press, 1993.

[7] P. Gros. Matching and clustering: Two steps towards object modelling in computer vision. *IJRR*, 1995.

[8] R. Hartley. A linear method for reconstruction from lines and points. *ICCV*, 1995.

[9] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *PAMI*, 1989.

[10] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff distance. *PAMI*, 1993.

[11] Q. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. Technical report, University of California, Berkeley, 1993.

[12] G. Médioni and R. Nevatia. Segment-based stereo matching. *CVGIP*, 1985.

[13] C. Rothwell, J. Mundy, and B. Hoffman. Representing objects using topology. *Object Representation in Computer Vision II*, LNCS 1144, Springer, 1996.

[14] C. Schmid and A. Zisserman. Automatic Line matching across views. Technical report, University of Oxford, 1997.

[15] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. *CVPR*, 1996.

[16] A. Shashua. Trilinearity in visual recognition by alignment. *ECCV*, 1994.

[17] M. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *IJCV*, 1990.

[18] P. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *BMVC*, 1996.

[19] V. Venkateswar and R.Chellappa. Hierarchical stereo and motion correspondence using feature groupings. *IJCV*, 1995.

[20] Z. Zhang. Token tracking in a cluttered scene. *IVC*, 12(2):110–120, 1994.

[21] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI Journal*, 1995.