Wide Baseline Stereo Matching

Philip Pritchett and Andrew Zisserman
Robotics Research Group, Department of Engineering Science
Oxford University, OX1 3PJ.

Abstract

The objective of this work is to enlarge the class of camera motions for which epipolar geometry and image correspondences can be computed automatically. This facilitates matching between quite disparate views — wide baseline stereo.

Two extensions are made to the current small baseline algorithms: first, and most importantly, a viewpoint invariant measure is developed for assessing the affinity of corner neighbourhoods over image pairs; second, algorithms are given for generating putative corner matches between image pairs using local homographies. Two novel infrastructure developments are also described: the automatic generation of local homographies, and the combination of possibly conflicting sets of matches prior to RANSAC estimation.

The wide baseline matching algorithm is demonstrated on a number of image pairs with varying relative motion, and for different scene types. All processing is automatic.

1 Introduction

It is now possible to automatically compute the epipolar geometry and point correspondences between two images of a sequence from image information alone [16, 21]. This simultaneous computation of camera geometry and feature matches has been extremely successful in both stereo and motion applications. For example, it has allowed the matching of primitives through sequences of 30+ images without requiring any knowledge of the 3D scene structure or camera internal calibration or camera motion [1, 20].

The computation is achieved by the combination of two technologies: first, reliable "generic" visual primitives — interest points/"corners" [7] — with a matching affinity score based on cross-correlation of neighbourhood intensities; second, robust statistical estimation algorithms that can tolerate mismatches in putative matched sets, such as RANSAC [4], and Least Median of Squares [21]. Once epipolar geometry is available the stereo matching literature of the '70's and '80's provides many methods and constraints for determining correspondences of other image features such as lines and curves.

However, the current generation of algorithms which compute matches and epipolar geometry fail unless the relative camera motion and change in internal parameters between the views is restricted. There are several connected reasons for this failure [11], but the dominant one is that cross-correlation on corner intensity neighbourhoods fails to provide a veridical affinity score for matching corners.

The methods presented in this paper extend the "envelope" of matching techniques to overcome these restrictions. A *wide baseline stereo* algorithm is developed which can cope with significant rotations, translations and changes in internal parameters between views. Figure 1 illustrates the class of motions for which the wide baseline algorithm is applicable.

There have been a number of recent photogrammetry applications where such wide baseline matching is required. These are generally aimed at reconstructions from multiple views, where the baseline is large to improve the accuracy of reconstruction or a small number of views cover all aspects of the object [5, 15, 20]. Currently some of the correspondences for these applications are established by hand.

Although extensions to small baseline algorithms have been proposed [12, 20, 21] they are not sufficient to cope with the perspective foreshortening effects which can occur in wide baseline images.

The key enabling idea in this paper is the use of *local* planar homographies (plane projective transformations). These are used in two quite distinct rôles. First, to define a viewpoint invariant affinity measure (section 2); and second, to restrict search when generating putative corner matches (section 4.1).

Homographies have been used for these rôles before, but only when the epipolar geometry is known [8]. In particular homographies are used in rectification, where the images are mapped to those of a parallel camera geometry. However, in these previous cases the

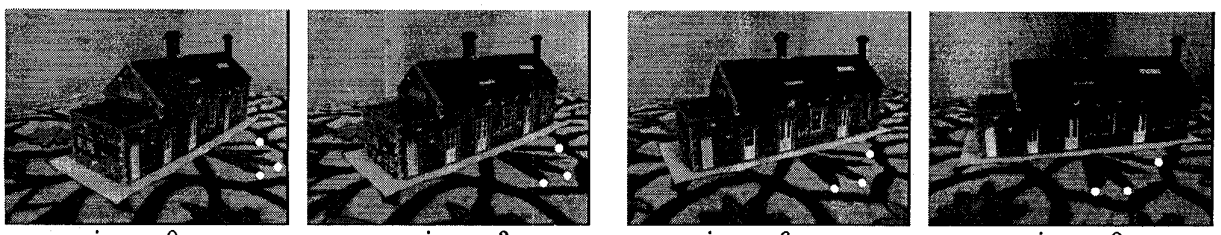


Figure 1: *A sequence consisting of small translations between views, with rotations to fixate on the house. There is a small baseline between any two consecutive images of the sequence, however between the first and last image there is a wide baseline. The affinity measure of the corners highlighted in white are given in figure 2.*

homography is defined only by the camera motion and internal parameters (i.e. the epipolar geometry) and is global for each image. In contrast in the wide baseline algorithm there are many local homographies for an image pair, and these are defined by 3D scene structure as well as epipolar geometry. The single global homography for each image is then a special case.

The wide baseline algorithm has a number of stages. First, homographies between the images are generated (section 3). These homographies are used to generate sets of putative matches (section 4). A fundamental matrix (representing the epipolar geometry) and a consistent set of matches are then obtained via RANSAC (section 5). Results on several image pairs are given at the end of the paper. All the stages are processed automatically.

2 A viewpoint invariant affinity measure

In small baseline algorithms the affinity between corners is determined by cross-correlating the corner intensity neighbourhoods. This affinity measure is used to assess corner matches over the two views.

Photometrically, cross-correlation is invariant to a linear (affine) transformation of image intensities — a constant addition and common scaling. Geometrically cross-correlation is not even invariant to a simple rotation of the intensity neighbourhood. Two examples will illustrate the need for a greater geometric invariance in the affinity measure:

Ex. A Suppose that the motion consists of a small lateral translation followed by a large rotation about the principal axis (cyclo-rotation). Since cross-correlation is not invariant to rotations, it does not provide a useful affinity measure in this simple motion case. This situation can be remedied by a global rotation of the image about the principal axis before cross-correlating, or by using rotationally invariant cross-correlation [12].

Ex. B Suppose that the (non-cyclo) rotation of the camera about the optical centre is significant, or the translation of the camera is significant. Consider the image of a planar surface. A square corner neighbourhood in one image back-projects to a planar facet in the world, and the image of this facet in the second image is a quadrilateral, its shape depending entirely on the relative positioning of the cameras and plane. There may be severe projective distortion of this quadrilateral due to differing perspective foreshortenings of the plane in each image. This case cannot be remedied by a global rotation of the image about the principal axis before cross-correlating. However, it can be remedied by a planar homography of the image, but since this homography depends on the relative positioning of the cameras and plane, it must be different for different planes.

It is evident then that in both examples warping by a homography is sufficient to make cross-correlation geometrically invariant. Consequently, a viewpoint invariant affinity measure can be defined by the intensity cross-correlation of pixels at image points $\mathbf{x}$ in a square neighbourhood of the first image with pixels at image points $\mathbf{x}'$ in the second image, where $\mathbf{x}$ is mapped to $\mathbf{x}'$ by the homography.

Although the motivation has been in terms of planes, as will be demonstrated in the sequel, the affinity measure performs well for smooth surfaces since locally the homography defined by the tangent plane is sufficient. Also, the class of homography needed to produce an invariant measure depends upon the cameras' internal parameters, motion, and the relative position of the surface. It is not always necessary to use the most general projective homography (8 degrees of freedom (dof), defined by four point correspondences). Affine (6 dof, 3 point correspondences) and similarity (4 dof, 2 point correspondences) often suffice.

The constancy of the viewpoint invariant affinity

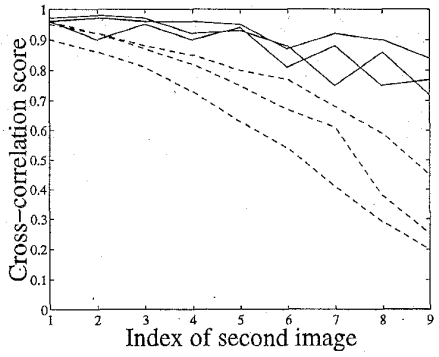


Figure 2: *Comparison of standard cross-correlation (dashed line) with the viewpoint invariant measure (solid line) for corresponding corner neighbourhoods. The three corners are those of figure 1.*

measure is demonstrated in figure 2, where it is compared to standard cross-correlation (i.e. an identity homography) as the baseline is extended. Both affinity measures use a 7×7 pixel neighbourhood defined in the zeroth image. For the standard cross-correlation a 7×7 neighbourhood of the same orientation is used in the other image. For the viewpoint invariant measure the corresponding point is computed via a homography, and the intensity at sub-pixel resolution determined by bi-linear interpolation. The (affine) homography is computed in this example from three corner correspondences on the same plane.

The viewpoint invariant affinity measure is far more stable than standard cross-correlation — deteriorating to only 0.7, compared to 0.2 for standard cross-correlation.

3 Automatic generation of homographies

This section describes methods for obtaining homographies between the images. Two contrasting approaches are given. The first based on feature group correspondences, the second on pyramid techniques. These homographies are used in section 4 to generate sets of putative corner matches.

3.1 Generating homographies from feature groups

To obtain corresponding feature groups we have at our disposal the large repository of ideas developed in the stereo matching and object recognition literature. The feature focus [2] approach is particularly suitable — using whatever distinctive features a particular image sequence provides us with.

Here we illustrate the idea by using correspondences of closed regions bound by four line segments. This follows the approach of Venkateswar and Chellappa [19], but does not involve maintaining relationships between the closed regions. Four line correspondences are sufficient to compute a projective homography. Three would be sufficient for an affine homography.

Line four cycles are grouped by searching for parallelograms amongst line segments obtained by edge detection, linking, and line fitting. The topology of the edges is preserved during line fitting and this simplifies the search. The grouping constraint only requires approximate parallelism and length equality for opposite sides. Often only 3-cycles are obtained and the remaining line is inferred to close the region. Figure 3 shows an example.

Once parallelograms have been grouped they are matched pairwise between the images. This has a quadratic complexity, but generally the number of four cycles is small, so the cost is not too high. Putative parallelogram matches are verified by computing a projective homography (from the four line correspondences) and cross-correlating the projectively warped region enclosed by the four-cycle.

3.2 Generating homographies from image pyramids

Here Gaussian pyramid techniques [3, 6] are used. A similarity transformation is estimated by optimising intensity cross-correlation of the entire image over the four parameters at the highest level of the pyramid. Moving down the pyramid, local affine transformations are computed at each level in a similar manner to the estimation of the similarity. The inherited transformation from the level above provides the starting point for the new transformation estimation.

4 Generating corner matches

In this section it is shown how a homography can be used to generate a large set of putative corner matches. Section 4.1 introduces the basic component common to all the generation strategies — that a transformation simplifies the correspondence problem by predicing the position for the match. Then several strategies are outlined which attempt to grow an initial set of matches, generated from a local homography. This is an important stage in the overall wide baseline algorithm as its success will maximise

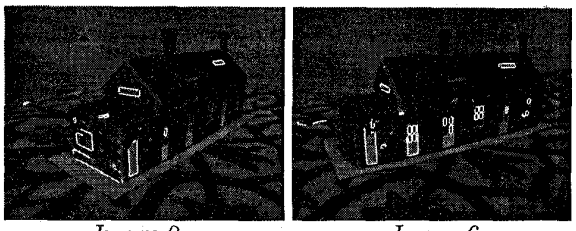


Image 0 Image 6

Figure 3: *Automatic extraction of parallelogram line cycles. The 3 and 4 cycles are shown before parallelogram completion. There are 15 groupings in image 0, and 26 in image 6.*

the number of putative correspondences that can be generated given an initial local homography.

These sets are used as the correspondences in a robust estimation of the fundamental matrix, as described in section 5.

4.1 Matching via a homography

In addition to providing an invariant affinity measure for individual corners, a homography can also greatly simplify matching for sets of corners by reducing the search complexity.

Suppose the homography exactly represented the mapping between image points, then corner points mapped from one image to the other would coincide exactly with their correspondences. Generally, the homography is only an approximation of the mapping, (e.g. where the surface is not planar) so that mapped corners do not coincide exactly. If the homography is computed from four corresponding points, for example, this disparity generally increases with distance from the basis used. The acceptance of a putative corner match then depends on both the viewpoint invariant affinity measure, and a search window disparity which varies according to distance from the basis.

4.2 Growing matches

The aim here is to obtain large sets of (putatively) matched corners starting from a supplied local homography — despite the fact that generally the homography does not apply to the entire image.

The first step is to obtain an initial set of corner correspondences consistent with the homography using the disparity matching of section 4.1. The aim is then to "grow" these correspondences, whilst also providing a homography for both the viewpoint invariant affinity measure and matching. Since generating an initial local homography is in general a more costly process than utilising the homography, it is in our best interests to obtain as many matches as possible from a given homography.

Several growing strategies have been developed each varying in its goal and assumptions. Each strategy has a basic transformation type — an affine or projective transformation, with the methods differing according to whether only matches consistent with a single transformation are sought or whether many sets of matches each consistent with a local transformation are found.

The first method — single transformation — can be thought of as attempting to find all matches that are consistent with a single planar homography. If the initial local homography arises from a dominant plane in the scene, then this method will find the matches on that plane.

The second method — varying transformation — attempts to find new homographies (affine or projective) from the initial homography, and use these new homographies to compute new sets of matches, which will be used in the same manner. This method allows a set of matches to be built up over a curved surface, and also allows matching to "jump" over plane boundaries.

Two of these algorithms will now be described and illustrated.

4.3 Plane following using a single projective homography

Initial matches are first obtained as described above using the supplied local homography. A RANSAC algorithm is used to select a subset of these matches within a fixed region which are consistent with a projective homography. The size of the region is then increased and the procedure repeated, using the RANSAC determined projective homography instead of the supplied homography. The size increase of the region is not homogeneous, but is governed by the proportion of new matches found. These two steps are then iterated, and the algorithm terminates when the number of new matches is below threshold.

An example is shown in figure 4. Coplanar points are clearly obtained. The success of the algorithm is indicated in table 1. Each supplied homography generates a set of over 100 correspondences, of which over 60% are consistent with the final estimated epipolar geometry.

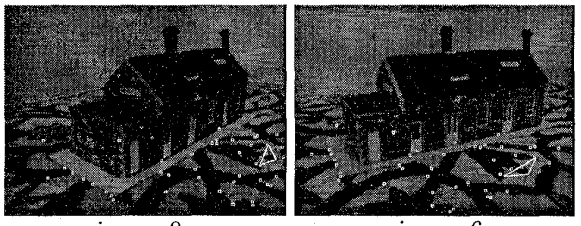


image 0 image 6

Figure 4: *Matches generated from a supplied affine homography (indicated by the white basis triangle) using a plane following algorithm. This is basis 3 from table 1.*

basis	#initial matches	#inliers to F	#grown matches	#inliers to F
1	27	14	134	78
2	69	39	162	84
3	54	41	133	69

Table 1: *Number of matches generated by the plane following algorithm for three affine homographies.*

4.4 Surface following using affine transformations

Initial matches are first obtained as described above using the supplied local homography. The selection window is then subdivided into four, and for each subdivision matches consistent with an affine homography are determined using RANSAC. Where sufficient support for an affine homography exists, three consistent matches are chosen as a basis and a new region centred upon the centroid of the basis is matched. The effect is that new matches are "grown" on the periphery of each selection window. Novel matches are stored. The algorithm terminates when no new affine bases can be found.

An example is shown in figure 5. Even though the initial (affine) homography is for the front of the house, correspondences are obtained on many other planes. Table 2 indicates the success of the algorithm in "inflating" the number of matches: the set of matches consistent with the estimated epipolar geometry is increased to over 100 for each set of initial matches.

5 Computation of the Fundamental Matrix

The previous section showed that a plentiful supply of correspondences can be generated from a few local homographies. It then remains to compute a fundamental matrix from these correspondences. As there may be a significant proportion of mismatches remaining in the match sets, the robust estimation algorithm RANSAC is used to obtain a fundamental matrix together with a set of consistent matches.

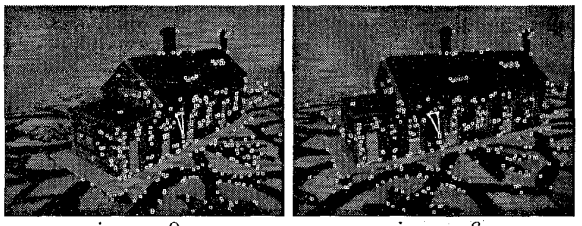


image 0 image 6

Figure 5: *Matches generated from a supplied affine homography (indicated by the white basis triangle) using a surface following algorithm. This is basis 4 from table 2.*

basis	#initial matches	#inliers to F	#grown matches	#inliers to F
1	27	14	386	134
2	69	39	343	125
3	54	41	309	123
4	60	33	416	173

Table 2: *Number of matches generated by the surface following algorithm for four affine homographies.*

The issue now is how to combine the sets of matches provided by the various initial homographies and growing algorithms. One possibility is simply to take the union of all the matches. This has the possible disadvantage that a particular corner might be allocated more that one correspondences because it is in several of the sets. However, RANSAC might well be able to cope with this since a multiple correspondence is simply a different type of mismatch. Another possibility, which is the one adopted here, is to enforce single correspondence. Where there is conflict the match with higher viewpoint invariant affinity measure is selected. Once this set of potential matches has been generated, the fundamental matrix is computed using RANSAC based on seven correspondences.

In a typical example, figure 6, there are 3 initial homographies generating 343, 309 and 416 matches each. These three sets combine to form a new set of 446 putative matches of which 132 are consistent with the RANSAC estimate of the epipolar geometry.

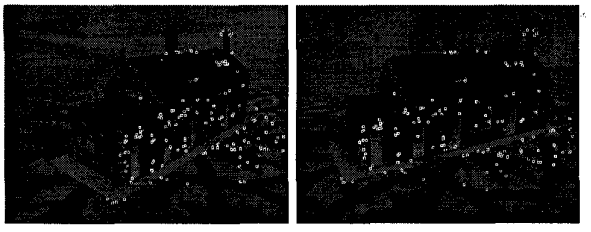


Figure 6: *Images 0 and 6 of the house. There are 132 inlying matches of a total of 446 (inliers and outliers).*

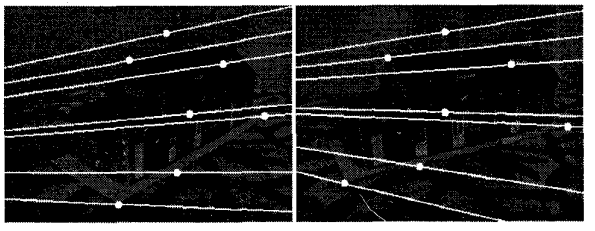


Figure 7: *Corresponding epipolar lines and points using the computed fundamental matrix.*

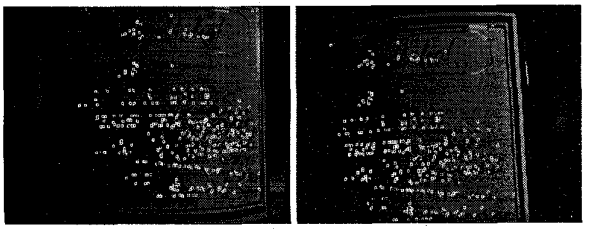


Figure 8: *Images of a bottle with 366 inlying matches of a total of 541 (inliers and outliers). Note, the matched corners clearly indicate the common regions of the two images.*

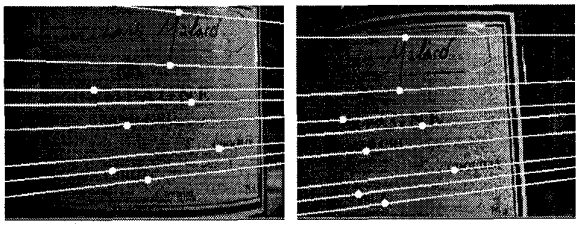


Figure 9: *Example epipolar lines and correspondences using the computed fundamental matrix.*

Figure 10: *Images of a chapel with 195 inlying matches of a total of 315 (inliers and outliers). Images courtesy of Fraunhofer IGD.*

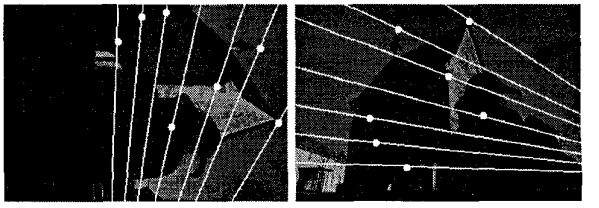
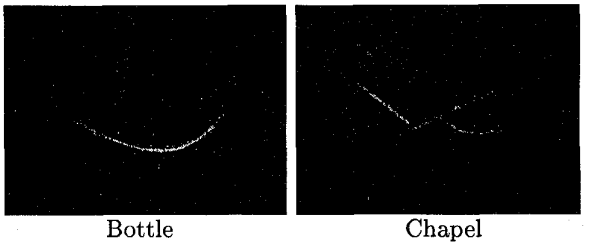


Figure 11: *Example epipolar lines and correspondences using the computed fundamental matrix.*



Bottle Chapel

Figure 12: *Plan views of 3D projective reconstruction of the bottle and chapel using inlying matches and the computed epipolar geometry.*

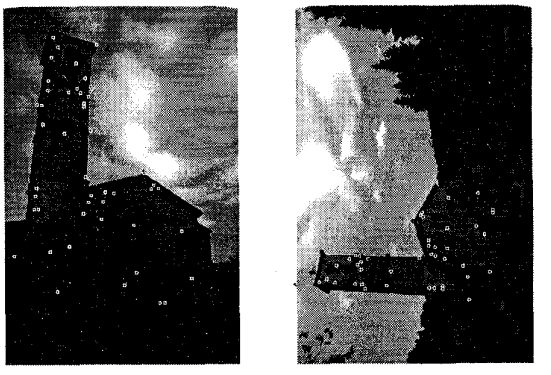


Figure 13: *Images of the Valbonne church with 50 inlying matches of a total of 94 (inliers and outliers). Images courtesy of INRIA, Sophia Antipolis.*

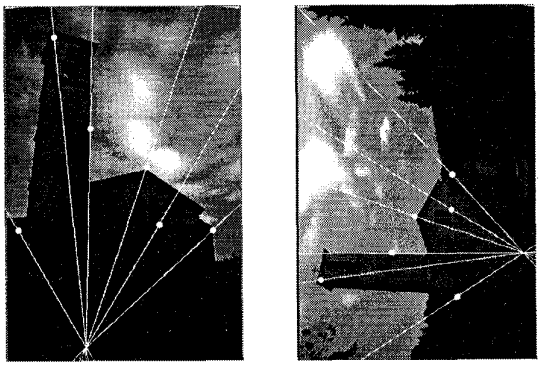


Figure 14: *Example epipolar lines and correspondences using the computed fundamental matrix.*

6 Algorithm summary and examples

1. Automatically generate a set of local homographies. For example, by matching feature groups or using a pyramid search (Section 3).

2. Use each homography to generate a set of corner matches (Section 4).

3. Sets of matches generated from each local homography are combined, and a fundamental matrix and consistent matches estimated using RANSAC (Section 5). A MLE of the fundamental matrix is then obtained from the inliers via a non-linear minimization.

Figures 6 to 14 show examples of the wide baseline computation. The figures show the automatically matched points, example epipolar lines, and reconstructed 3D scene points.

7 Conclusions and future work

We have demonstrated the automatic estimation of epipolar geometry and consistent matches for view pairs separated by a wide baseline. The epipolar geometry produced by the wide baseline algorithm is then the starting point for generating additional feature matches using guided matching.

Other strategies for obtaining homographies are currently being investigated including joint photometric/geometric invariants [18], and local graph matching [10]. Also, we are investigating the automatic selection of which strategy to apply.

Wide baseline algorithms are also being developed for other multiple view relations which currently are only estimated automatically for small baselines. Of particular importance is the automatic computation of the trifocal geometry (represented by the trifocal tensor [9, 13, 14]) for three views [17].

Acknowledgements

We are grateful for financial support from Sharp Research Labs of Europe and EU ACTS Project VANGUARD. We would like to thank Andrew Fitzgibbon, Phil Torr, and Paul Beardsley for discussions and assistance. Images were provided by RobotVis INRIA Sophia Antipolis, and Fraunhofer IGD.

References

[1] P.A. Beardsley, P.H.S. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. *ECCV 1996*, pages 683-695, Springer-Verlag, 1996.

[2] R.C. Bolles and R.A. Cain. Recognizing and locating partially visible objects: the local-feature-focus method. *International Journal of Robotics Research*, 1(3), 1982.

[3] P.J. Burt. Fast filter transforms for image processing. *CGIP*, 16:20–51, 1981.

[4] 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.

[5] N. Georgis, M. Petrou, and J.V. Kittler. Obtaining correspondences from 2D perspective views with wide angular separation of non-coplanar points. *Proc. of the European-Chinese Workshop on Computer Vision*, 376–379, 1995.

[6] F. Glazer, G. Reynolds, and P. Anandan. Scene matching by hierarchical correlation. *CVPR83*, 432–441, 1983.

[7] C.G. Harris and M. Stephens. A combined corner and edge detector. *Fourth Alvey Vision Conference*, 147–151, 1988.

[8] R. Hartley and R. Gupta. Computing Matched-epipolar Projections. *CVPR 1993*, pages 549 – 555, 1993.

[9] R. Hartley. A linear method for reconstruction from lines and points. In *ICCV 1995*, pages 882–887, 1995.

[10] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *PAMI*, 11(11):1168–1180, 1989.

[11] P.C. Pritchett and A. Zisserman. Wide baseline stereo matching. Technical report, Department of Engineering Science, University of Oxford, 1997.

[12] C. Schmid and R. Mohr. Matching by local invariants. Research report 2644, INRIA Rhône-Alpes, Grenoble, France, 1995.

[13] A. Shashua. Trilinearity in visual recognition by alignment. In *ECCV 1994*, pages 479–484, Springer-Verlag, 1994.

[14] M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *IJCV*, pages 171–183, 1990.

[15] C. Taylor, P. Debevec, and J. Malik. Reconstructing polyhedral models of architectural scenes from photographs. In *ECCV 1996*. Springer-Verlag, 1996.

[16] P.H.S. Torr and D.W. Murray. Outlier detection and motion segmentation. In *SPIE 93*, 1993.

[17] P.H.S. Torr and A. Zisserman Robust parameterisation and computation of the trifocal tensor. *BMVC*, 1996.

[18] L. Van Gool and T. Moons and D. Ungureanu. Affine / photometric invariants for planar intensity patterns. In *ECCV 1996*, pages 642–651, Springer-Verlag, 1996.

[19] V. Venkateswar and R.Chellappa. Hierarchical stereo and motion correspondence using feature groupings. *IJCV*, pages 245–269, 1995.

[20] C. Zeller. *Projective, Affine and Euclidean Calibration in Compute Vision and the Application of Three Dimensional Perception.* PhD thesis, RobotVis Group, INRIA Sophia-Antipolis, 1996.

[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. *Artificial Intelligence*, 78:87–119, 1995.