Navigation using Affine Structure from Motion

P.A. Beardsley, A. Zisserman and D.W. Murray

Robotics Group, Dept of Eng Science, University of Oxford, Oxford OX1 3PJ, UK.
tel: +44-865 273154 fax: +44-865 273908 email: [pab,az,dwm]@robots.ox.ac.uk

Abstract. A structure from motion algorithm is described which recovers structure and camera position, modulo a projective ambiguity. Camera calibration is not required, and camera parameters such as focal length can be altered freely during motion. Unlike recent schemes which compute projective or affine structure using a batch process, the structure is updated sequentially over an image sequence. A specialisation of the algorithm to recover structure modulo an affine transformation is described. We demonstrate how the affine coordinate frame can be periodically updated to prevent drift over time.
Structure is recovered from image corners detected and matched automatically and reliably in image sequences. Results are shown for reference objects and indoor environments. Finally, the affine structure is used to construct free space maps enabling navigation through unstructured environments and avoidance of obstacles. The path planning involves only affine constructions. Examples are provided for real image sequences.

1 Introduction

The recovery of structure from motion (SFM) is a sufficiently mature field for several working systems to have been applied to the navigation of mobile vehicles [3, 4, 14]. Such systems recover 3D Euclidean structure and require a calibrated camera. More recently, researchers have investigated SFM when the camera is uncalibrated [9, 13] recovering structure modulo a projective ambiguity. These methods work in *batch* mode, determining structure and camera projection matrices from a complete sequence of images. In contrast, we devise and apply an algorithm which recovers projective structure by *sequential update*. The structure is obtained using matched corners in images from a camera moving in a static scene, and the work could fairly be described as a projective counterpart to the Euclidean system 'DROID' of Harris *et al* [3, 4]. Unlike DROID, camera calibration is not required. However, when partial or approximate calibration is available it is exploited to render "Quasi-Euclidean" structure i.e. structure within a small "projective skew" of the actual Euclidean structure.

The projective SFM scheme is specialised to produce affine structure by use of a result of Moons *et al* [10]. Extra invariants are available when the recovered structure is affine including ratios of lengths on parallel segments, ratios of areas on parallel planes, ratios of volumes, and centroids. The affine structure is applied to the task of path planning in an unstructured environment. A basic mechanism in classical path planning is to find a mid-point locus between obstacles. This is an affine, not Euclidean, construct - thus many of the techniques from path

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

planning can be utilized when only affine rather than Euclidean structure is available. To demonstrate this, we navigate a camera held by a robot arm to a specified target, where the direct path to the target is occluded by unmodelled objects. The target is reached by incrementally determining free space regions from the affine structure as the robot moves, and path planning through these regions.

2 Camera models and projective representations

We introduce the camera models and notation used in the rest of the paper. These results are based mainly on [2, 7, 11]. Perspective projection from 3D projective space $\mathcal{P}^3$ to the image plane $\mathcal{P}^2$ is modelled by a 3×4 matrix $\mathtt{P}$

$$\mathbf{x} = \mathtt{P}\mathbf{X} \tag{1}$$

where $\mathbf{x} = (x, y, 1)^\top$ and $\mathbf{X} = (X, Y, Z, 1)^\top$ are homogeneous vectors. With homogeneous quantities, $=$ indicates equality up to a non-zero scale factor. $\mathtt{P}$ can be partitioned as

$$\mathtt{P} = (\mathtt{M}| - \mathtt{M}\mathbf{t}) \tag{2}$$

where $\mathbf{t}$ is the centre of projection, since the centre projects as $\mathtt{P}(\mathbf{t}^\top, 1)^\top = \mathbf{0}$. This partitioning is valid provided the left 3×3 matrix $\mathtt{M}$ is not singular (i.e. the optical centre is not on the plane at infinity). In a Euclidean frame, $\mathtt{P}$ can be further decomposed as

$$\mathtt{P} = \mathtt{K}(\mathtt{R}| - \mathtt{R}\mathbf{t}) \tag{3}$$

where $\mathtt{R}$ and $\mathbf{t}$ are the rotation and translation of the camera in the Euclidean frame. $\mathtt{K}$ is a 3×3 matrix encoding the camera intrinsic parameters

$$\mathtt{K} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

where α_x and α_y are the focal length measured in pixels along the x and y directions respectively, and (x_0, y_0) is the principal point.

For two cameras with $\mathbf{x}_1 = \mathtt{P}_1\mathbf{X}$ and $\mathbf{x}_2 = \mathtt{P}_2\mathbf{X}$, corresponding points in the two images satisfy the epipolar constraint

$$\mathbf{x}_2^\top \mathtt{F}\mathbf{x}_1 = 0 \tag{5}$$

where $\mathtt{F}$ is the 3×3 *fundamental matrix*, with maximum rank 2. The epipolar line in image 2 corresponding to $\mathbf{x}_1$ is $\mathbf{l}_2 = \mathtt{F}\mathbf{x}_1$, and similarly in image 1 corresponding to $\mathbf{x}_2$ is $\mathbf{l}_1 = \mathtt{F}^\top\mathbf{x}_2$, where $\mathbf{l}_i$ are homogeneous vectors. Partitioning $\mathtt{P}_1$ and $\mathtt{P}_2$ as in equation (2) facilitates a number of equivalent representations of $\mathtt{F}$

$$\mathtt{F} = \mathtt{M}_2^{-\top}[\mathbf{t}_1 - \mathbf{t}_2]_\times \mathtt{M}_1^{-1} = [\mathtt{M}_2(\mathbf{t}_1 - \mathbf{t}_2)]_\times \mathtt{M}_2\mathtt{M}_1^{-1} = \mathtt{M}_2^{-\top}\mathtt{M}_1^\top[\mathtt{M}_1(\mathbf{t}_1 - \mathbf{t}_2)]_\times \tag{6}$$

where the notation $[\mathbf{v}]_\times$ is the vector product represented as a matrix

$$[\mathbf{v}]_\times = \begin{pmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{pmatrix}$$

Under a 3D projective transformation, $\mathbf{X}' = \mathtt{H}\mathbf{X}$, where $\mathtt{H}$ is a non-singular 4×4 matrix, a camera matrix $\mathtt{P}$ is transformed to $\mathtt{P}' = \mathtt{P}\mathtt{H}^{-1}$ since

$$\mathbf{x}_i = \mathtt{P}'\mathbf{X}' = \mathtt{P}\mathtt{H}^{-1}\mathtt{H}\mathbf{X} = \mathtt{P}\mathbf{X} \tag{7}$$

In the following, a canonical camera matrix $\mathtt{P}_1 = (\mathtt{I}|\mathbf{0})$ will be used, where $\mathtt{I}$ is the 3×3 identity. This can always be achieved by setting $\mathtt{H}^{-1}$ in equation (7) to be

$$\mathtt{H}^{-1} = \begin{pmatrix} \mathtt{M}_1^{-1} & \mathbf{t}_1 \\ \mathbf{0}^\top & 1 \end{pmatrix}$$

3 Projective Structure From Motion

Corner correspondences between images are used to recover the position of 3D points and the optical centre up to a projectivity of $\mathcal{P}^3$ i.e. if $\mathbf{X}_E = (X, Y, Z, 1)^\top$ is a homogeneous vector representing the Euclidean position (X, Y, Z) of a point then the recovered position is $\mathbf{X} = \mathtt{H}\mathbf{X}_E$ where $\mathtt{H}$ is a non-singular 4×4 matrix which is unknown but the same for all points.

3.1 Frame initialisation

Previous methods for projective reconstruction from two or more images [2, 9] have selected a five point basis from the 3D points. The problem with this procedure is that if one of the points is poorly localised in an image, the accuracy of the entire reconstruction degrades. Instead, we follow more closely the approach of Hartley [6] and utilise *all* point matches to determine the projective frame, by specifying the perspective projection matrices $\mathtt{P}_1$ and $\mathtt{P}_2$ for two images.

$\mathtt{P}_1$ is chosen to have the canonical form $\mathtt{P}_1 = (\mathtt{I}|\mathbf{0})$ (Section 2). The fundamental matrix $\mathtt{F}$ is obtained from corner matches (Section 3.3) and decomposed as the matrix product $\mathtt{F} = -[\mathtt{M}_2\mathbf{t}_2]_\times\mathtt{M}_2 = [\mathbf{s}]_\times\mathtt{M}_2$ (equation (6)). It can be shown that the most general form of $\mathtt{P}_2$ which is consistent with $\mathtt{P}_1$ and $\mathtt{F}$ is $\mathtt{P}_2 = [\mathtt{M}_2 + \mathbf{a}^\top\mathbf{s}|b\mathbf{s}]$ where $\mathbf{a}$ is a 3-vector and b a scalar i.e. there are 4 DOF in $\mathtt{P}_2$ (plus the usual homogeneous scale factor). In the absence of any other information, values for $\mathbf{a}$ and b are arbitrary. However, if partial or approximate intrinsic parameters or ego-motion are available then this information is utilised, firstly by using natural camera measurements i.e. an image corner at pixel position (u, v) is assigned homogeneous coordinates $\mathbf{x} = ((u - x_0)/\alpha_x, (v - y_0)/\alpha_y, 1)^\top$ where $\alpha_x, \alpha_y, (x_0, y_0)$ are defined in equation (4), and secondly by choosing $\mathtt{P}_2$ (setting the 4 DOF) so that $\mathtt{P}_2 = (\mathtt{R}|-\mathtt{R}\mathbf{t})$ where $\mathtt{R}$ and $\mathbf{t}$ are the approximate rotation and translation of the camera. This produces a "Quasi-Euclidean" frame i.e. a frame which is exactly Euclidean if the intrinsic parameters and ego-motion are exact, but subject to a projective skew otherwise.

3.2 3D point initialisation and update

The 3D coordinates of scene points and optical centres are computed in the projective coordinate frame as in [2]. The camera centre $\mathbf{C}_i = (\mathbf{t}_i^\top, 1)^\top$ associated with $\mathtt{P}_i = (\mathtt{M}_i| - \mathtt{M}_i\mathbf{t}_i)$ is determined from $\mathtt{P}_i\mathbf{C}_i = \mathbf{0}$. The 3D coordinates of a point $\mathbf{X}$ which projects to $\mathbf{x}_1$ in image 1 by $\mathtt{P}_1$ and $\mathbf{x}_2$ in image 2 by $\mathtt{P}_2$ are found, as in conventional stereo, by intersecting backprojected rays. The ray

for a point $\mathbf{x}_i$ in image i is given by $\alpha \mathbf{C}_i + \beta \mathbf{X}_i^\infty$, where $\mathbf{X}_i^\infty = \mathtt{M}_i^{-1}\mathbf{x}_i$ is the ray's intersection with the plane at infinity. In practice the reconstructed rays from the two cameras will be skew so $\mathbf{X}$ is determined as the midpoint of the perpendicular between the rays. Such an approach would not be justified in an arbitrary projective frame because "mid-point" and "perpendicularity" are not projective invariants. However, the estimate is good in a Quasi-Euclidean frame.

Update of 3D point coordinates involves two stages - *(1)* use of matches between corners in I_n (image n) which have associated 3D coordinates and corners in I_{n+1} to compute the perspective projection matrix for I_{n+1}, *(2)* update of 3D point coordinates with an Iterated Extended Kalman Filter (IEKF). A separate IEKF is run on each 3D point - the state vector is the point's non-homogeneous 3D coordinates, the observation vector is the point's non-homogeneous image coordinates, and the observation equation is the standard projection equation (1) expressed in non-homogeneous form. In the usual way, a covariance matrix is produced and updated in association with each state vector.

3.3 Implementation

Image corners are extracted to sub-pixel accuracy using [5]. Correspondence matching is carried out automatically, without any information about camera motion, as a three stage process:

Unguided matching: the aim is to obtain a small number of reliable seed matches, then to compute $\mathtt{F}$ and $\mathtt{P}$ which are used to guide further matching. Given a corner at position (x, y) in the first image, the search for a match is centred on (x, y) in the second image, and the strength of candidate matches is measured by cross-correlation. The threshold for match acceptance is deliberately conservative to minimise incorrect matches.

Use of epipolar geometry: the seed matches are used to compute $\mathtt{F}$ by an iterative linear process - at each iteration, matches are weighted according to their agreement with the current estimate of $\mathtt{F}$; at the final iteration, outlying matches are removed altogether and marked as unmatched (see also [8]). The estimated $\mathtt{F}$ is used to generate epipolar lines, and the search for a correspondence for an unmatched corner can be restricted to a band about its epipolar line.

Use of projected structure: corners in I_n (image n) which have associated 3D coordinates and which are matched to corners in I_{n+1} provide a correspondence between 3D points and the image corners in I_{n+1}. Hence, it is possible to compute the perspective projection matrix for I_{n+1}, and we employ an iterative process similar to the one used for $\mathtt{F}$. Once $\mathtt{P}$ has been found, it is used to project unmatched 3D points onto I_{n+1}, and the search area for a corner's match is confined to a circle around the projected point.

The final set of matches is used to recompute $\mathtt{F}$ and $\mathtt{P}$, firstly using the linear methods above, and then employing a non-linear method which minimises an error based on image plane distance, rather than algebraic distance as in the linear case. In addition, $\mathbf{rank}(\mathtt{F}) = 2$ is enforced in the non-linear case, which cannot be done in the linear computation [2].

System parameters such as search radii and match strength thresholds which are used in the above processing are supplied as *a priori* values at the start of the

sequence and are then updated at the end of the frame according to the current matching statistics. 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 200-250.

3.4 Results

The experimental setup is a camera mounted on a robot arm, moving in a horizontal plane and rotating around a vertical axis. Figure 1 shows a reference object made of two perpendicular Tsai calibration grids. Recovered structure is assessed by *(1)* measuring projective invariants, and *(2)* transforming to a Euclidean coordinate frame to measure Euclidean invariants.

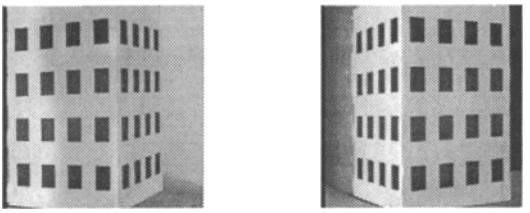

Fig. 1. *First and last images from a sequence of the reference object.*

Four equally spaced collinear points have a cross-ratio of 4/3. Thirty-two such cross-ratios are computed for the reference object at each frame, and the results plotted in Figure 2. The recovered structure shows monotonic improvement.

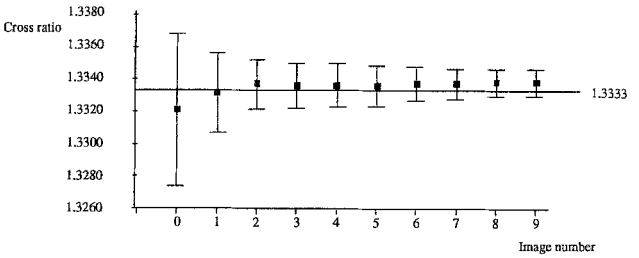

Fig. 2. *Mean value and one sigma standard deviation for cross ratios (actual value 4/3) computed from the recovered projective structure, against frame number.*

Transformation of the structure into a Euclidean frame requires the coordinates of five or more points in the projective and Euclidean frames [12], where direct physical measurement on the reference object is carried out to obtain the Euclidean positions. See Figures 3 and 4.

Measurements on the structure are given in Table 1. The collinearity measure $L = \sqrt{\sigma_y^2 + \sigma_z^2}/\sigma_x$ and the coplanarity measure $P = \sigma_z/\sqrt{\sigma_x^2 + \sigma_y^2}$ for a selected

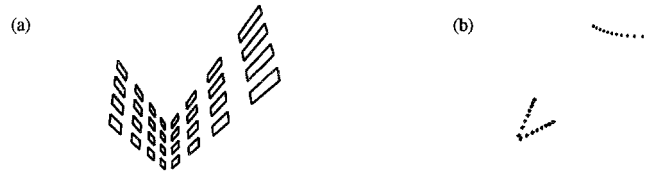

Fig. 3. (a) *Structure of the reference object in an arbitrary projective frame - coplanarity and collinearity are preserved as expected, but the structure is projectively skewed along one plane, and the angle between the two planes is less than* 90° *(connectivity has been added to the point structure for illustration).* (b) *Plan view of the computed structure viewed edge-on along the planes of the reference object (lower left) and showing the computed camera positions (upper right) in the arbitrary projective frame. Compare with the plan view* **after** *transformation to the Euclidean frame in Figure 4.*

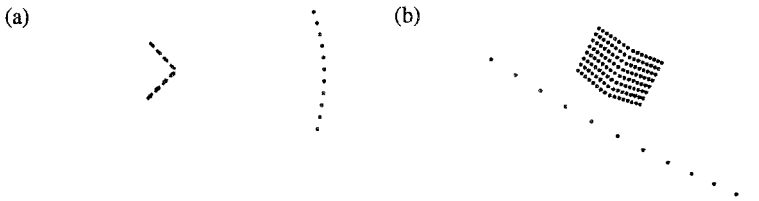

Fig. 4. (a) *Plan view of the reference object structure and camera motion transformed to a Euclidean frame - at left is the reference object with its planes edge-on, and at right the arc of dots shows successive camera positions in a circle.* (b) *View from behind the arc of camera positions.*

set of 3D points are obtained by using Singular Value Decomposition to obtain the principal axes of the set's spatial distribution together with the variance σ_{xyz} of point positions along each axis. A straight line has $L = 0$ and a perfect plane has $P = 0$. The table also includes a comparison with a local implementation of the DROID system [3, 4] which computes Euclidean structure directly. The results indicate that there is no significant difference between the quality of the projective and Euclidean systems, even though no camera calibration is employed in the projective case.

Section 3.1 introduced the Quasi-Euclidean frame which is obtained using approximate knowledge of the camera intrinsic parameters and ego-motion. Experiments in [1] investigate the effect of varying the estimates of the intrinsic parameters, and show that numerical instabilities causing degradation in the computed structure only arise with *extreme* projective frames which are far from Euclidean. Figure 5 shows results in a Quasi-Euclidean frame for an indoor scene.

Measure	Expected value	Projective	Affine	DROID
	After 2 frames			
Point error (mm)	0.0	0.5	0.7	0.7
Collinearity	0.0	0.003	0.005	0.006
Coplanarity	0.0	0.004	0.006	0.007
Cross-ratio	4/3	1.332 ± 0.006	1.333 ± 0.003	1.332 ± 0.005
Distance ratio	1.0	0.999 ± 0.012	1.002 ± 0.009	1.000 ± 0.013
	After 20 frames			
Point error (mm)	0.0	0.4	0.4	0.5
Collinearity	0.0	0.002	0.002	0.004
Coplanarity	0.0	0.002	0.003	0.004
Cross-ratio	4/3	1.333 ± 0.002	1.333 ± 0.001	1.333 ± 0.002
Distance ratio	1.0	1.000 ± 0.004	1.000 ± 0.006	0.999 ± 0.007

Table 1. *For the projective structure, the cross-ratio measurement was made before transformation to the Euclidean frame, and the remaining measures after. For the affine structure, the cross-ratio and ratio measurements were made before transformation to the Euclidean frame, and the remaining measures after. 128 points were used to compute the transformation to the Euclidean frame. The point error is the average distance between a measured point and the veridical Euclidean point, in the Euclidean frame. Coplanarity is a mean value for the two faces of the reference object.*

4 Affine Structure From Motion

The objective of the affine SFM processing is to recover the 3D position of scene points and the optical centre up to an *affine* transformation of three-space, i.e. if $\mathbf{X}_E = (X, Y, Z, 1)^\top$ is a homogeneous vector representing the Euclidean position (X, Y, Z) of a point, then the recovered position is $\mathbf{X} = \mathtt{H}_A \mathbf{X}_E$ where $\mathtt{H}_A$ is an affine transformation which is unknown but the same for all points,

$$\mathtt{H}_A = \begin{bmatrix} \mathtt{A} & \mathbf{T} \\ 0^\top & 1 \end{bmatrix} \tag{8}$$

with $\mathtt{A}$ a non-singular 3×3 matrix and $\mathbf{T}$ a general 3-vector.

Moons *et al* [10] have shown that affine structure can be obtained from a perspective camera undergoing pure translational motion with fixed internal parameters. We capitalise on this observation, using a single pure translation to determine the plane at infinity π_∞ in the current projective coordinate frame. Subsequently, we make occasional pure translational motions to update the plane at infinity in case of drift over time. Measurement of π_∞ follows from the lemma [1]:

> Given two camera matrices $\mathtt{P}_1 = (\mathtt{I}|0)$ and $\mathtt{P}_2 = (\mathtt{M}|\mathbf{t}^*)$ for identical cameras related by a pure translation, then $\mathtt{M} = k\mathtt{I} + \mathbf{t}^*\mathbf{v}^\top$ where $\pi_\infty = (\mathbf{v}^\top, 1)$ is the equation of the plane at infinity.

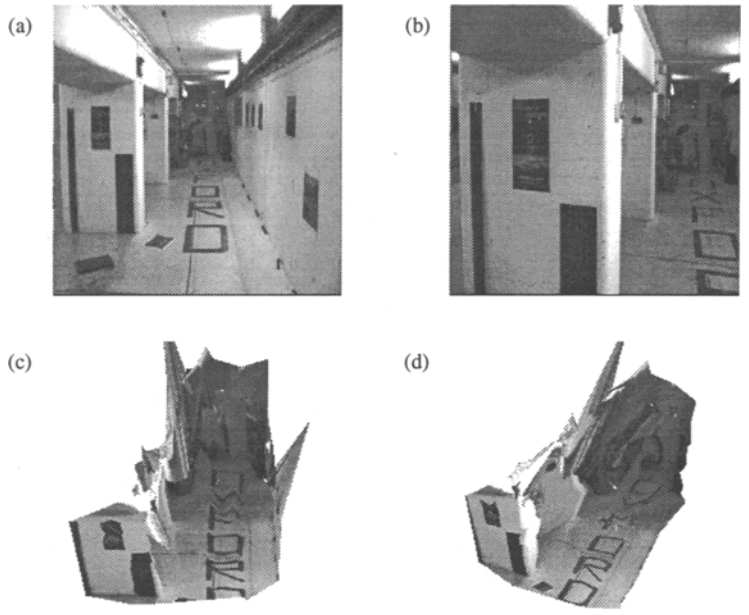

Fig. 5. **(a),(b)** *Example images taken during translation and rotation of an AGV (Autonomous Guided Vehicle) in a laboratory.* **(c),(d)** *Images constructed by mapping image intensity onto the computed 3D structure, and viewing from different positions. The mapping process is based on a Delaunay triangulation of the 2D image corners. The structure is in a quasi-Euclidean frame.*

We determine when $\mathtt{D} = \lambda \mathtt{I} - \mathtt{M}$ drops rank to 1 as a function of λ; then $\mathtt{D} = \mathbf{t}^* \mathbf{v}^\top$, and $\mathbf{t}^*$ is known, so it is possible to solve for $\mathbf{v}$.

As well as updating the plane at infinity, the case of pure translation is exploited in determining the fundamental matrix. If the camera is undergoing pure translation and the intrinsic parameters are fixed, then equation (6) has the special form $\mathtt{F} = \mathtt{M}^{-\top}[\mathbf{t}_1 - \mathbf{t}_2]_\times \mathtt{M}^{-1}$. Thus $\mathtt{F}$ is skew and has only two DOF. This reduces processing in the computation of $\mathtt{F}$ and leads to a more accurate result than the general case.

4.1 Results

Table 1 includes quantitative results for computed affine structure. Figure 6 shows results for an indoor sequence.

5 Navigation in Affine Space

The affine SFM scheme is utilised for navigation by incrementally computing free space regions, and planning paths through these regions in order to reach a specified target. For both the robot arm used in our experiments and an AGV, it is sufficient to project recovered affine structure onto the ground plane, and to compute free space and plan motions in 2D on this plane.

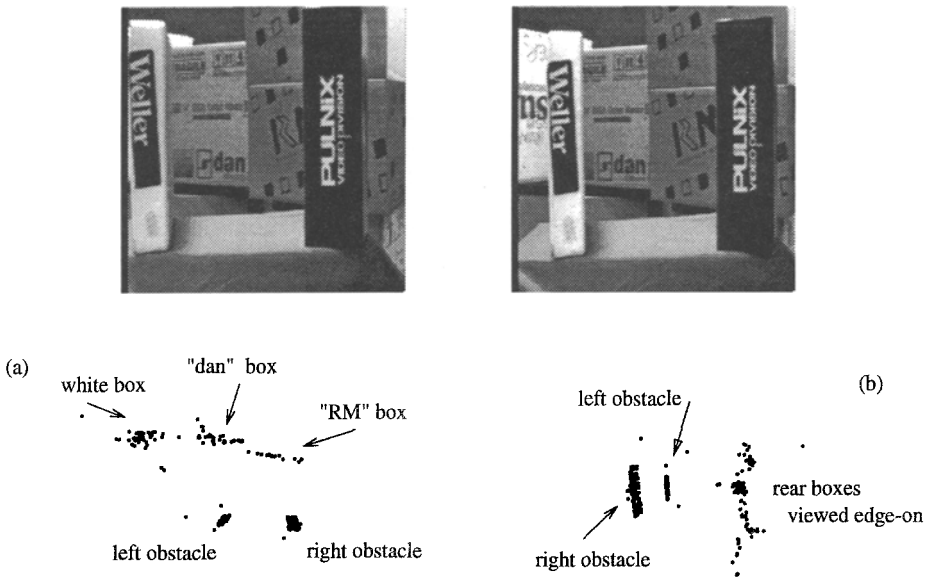

Fig. 6. *Two images from a sequence with structure recovered in a Quasi-Euclidean affine frame. (a) Plan view of recovered structure. (b) View from the right and to the rear of the obstacles. See the schematic plan view in Figure 8.*

At the most basic level the robot could be controlled by *servoing* alone i.e. with no calibration the robot could be driven to rotate until a distinguished affine feature such as a centroid is at the middle of the image. Instead we transform between the affine coordinate system and the robot Euclidean frame. Determining the transformation requires the coordinates of the camera centre in the affine frame and the robot Euclidean frame at a minimum of four non-coplanar positions. Once in the robot frame, projection to the ground plane is straightforward.

A limitation of the 3D structure consisting solely of points is that there is no representation of continuous surfaces and thus no notion of the free space between objects. We use a simple occlusion test to detect free space as illustrated in Figure 7. If a 3D point is visible continuously over a number of frames, then there is no occluding surface in the region defined by the 3D point and the moving optical centre. This free space region is projected onto the ground plane defining a free space triangle. If the triangle contains any other points from the projected structure, it is rejected. The free space map is the union of all accepted triangles.

Processing begins with a check on the free space map to determine whether there is an unobstructed route to the target. If an obstruction is present, a choice of alternative direction is made by selecting the free space "lobe" of largest area, and moving along its mid-line. As long as no obstruction is present, the camera

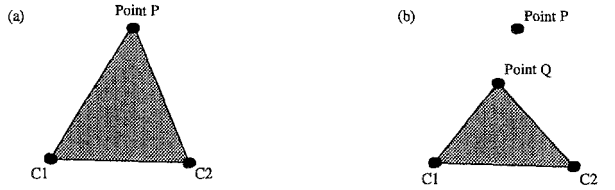

Fig. 7. *3D points projected onto the ground plane. The camera moves from C_1 to C_2. (a) Point P is visible continuously and C_1PC_2 does not contain any other projected points, so it is marked as free space. (b) Triangle C_1PC_2 is unacceptable as free space because of the presence of Q; however C_1QC_2 is accepted.*

moves in a straight line and two checks are carried out - firstly on time-to-contact measurements to detect potential collision in the forward direction; secondly on the line of sight to the target to detect whether the obstacle has been passed and an adjustment in the direction to the target can take place - the robot then proceeds to the target.

5.1 Results

Figure 8 is a schematic plan view of the experimental setup. Figure 9 shows plan views of the computed free space, and the projection of the structure onto the ground plane. The robot is unable to proceed directly to the target position because it would strike the left obstacle, so it moves first to the gap and then alters trajectory toward the target. Typical images from the sequence are shown in Figure 6.

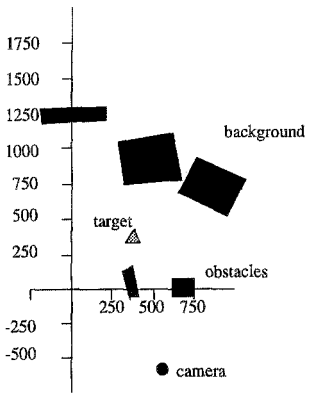

Fig. 8. *Plan view of the experimental layout (c.f. Figures 6 and 9). Axes in mm.*

6 Discussion of Implementation

Wide-angle lens: use of a wide angle lens leads to better camera localisation because rays to scene points have good divergence; it also makes it easier to fix

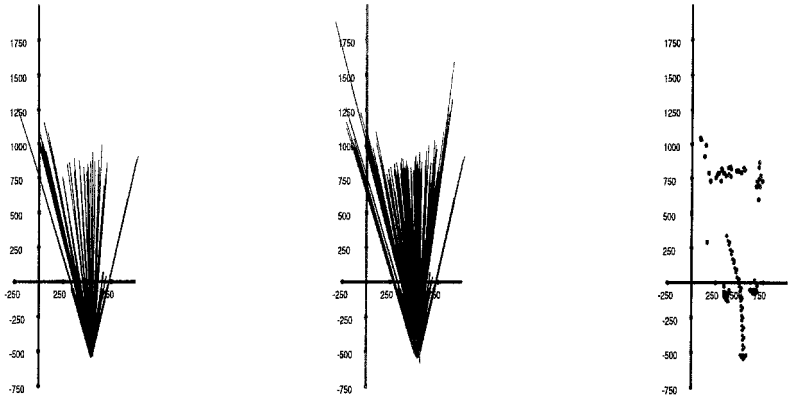

Fig. 9. *Plan view of free space (marked black) maps taken early and late in a sequence, with a consequent filling out of space in the later map. The right-hand figure shows the projection of 3D scene structure and camera motion (vertical line) onto the ground plane (c.f. the schematic plan view in Figure 8). Axes are labelled in mm.*

each new camera position in the ongoing coordinate frame because many points remain in view between images.

Forward motion: simple forward motion produces poor structure because rays from the camera to a scene point change angle slowly (*c.f.* the effect of a lateral motion) resulting in large error in the computed point position. To avoid this situation forward motion paths are "dithered" with lateral movements. Stereo would be of obvious benefit.

Computation of $\mathtt{F}$: accuracy of epipolar computation (average distance of a corner from its epipolar line) is typically 0.4 pixels when using corner detection on indoor scenes, and 0.02 pixels when image points are located by line intersection on the reference object.

Sensitivity to outliers: empirically, the computation of $\mathtt{F}$ is less sensitive to the presence of mismatches than the computation of $\mathtt{P}$, so $\mathtt{F}$ is always computed first and used to eliminate mismatches (Section 3.3).

Homogeneous coordinates: The arbitrary homogeneous component in a homogeneous vector is typically chosen as unity, e.g. an image corner (x, y) is represented as $(x, y, 1)$. Increased stability is achieved if the third component is chosen to be of the same order of magnitude as x and y (as in Section 3.1). The same considerations apply in $\mathcal{P}^3$.

7 Conclusion

The recovery of projective and affine structure is increasingly well-understood, but its use raises interesting problems about what can be achieved when Euclidean measurements are not available. We have demonstrated the recovery of projective and affine structure with an accuracy similar to a system using calibrated cameras, and applied the affine structure in path planning. The use of the translational motion constraint [10] to attain affine structure is part of a

spectrum of possibilities for investigation, ranging from fully calibrated stereo heads through to cameras of unknown intrinsic parameters and motion. We have concentrated on the uncalibrated end of the spectrum but have introduced a mechanism, the Quasi-Euclidean frame, for incorporating poor or partial camera calibration in the structure computation.

This work was supported by SERC Grant No GR/H77668 and Esprit BRA VIVA. Thanks for helpful discussions with Richard Hartley, Jitendra Malik, John Mayhew, Joe Mundy, and to colleagues in the Robotics Research Group, particularly Andrew Blake, Mike Brady, Phil McLauchlan, Ian Reid, Larry Shapiro, and Phil Torr.

References

1. P.A. Beardsley, A.P. Zisserman, and D.W. Murray. Sequential update of projective and affine structure from motion. Technical report OUEL 2012/94, Dept of Eng Science, University of Oxford, 1994.
2. 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.
3. C.G. Harris. Determination of ego-motion from matched points. In *Third Alvey Vision Conference*, pages 189–192, 1987.
4. C.G. Harris and J.M. Pike. 3D positional integration from image sequences. In *Third Alvey Vision Conference*, pages 233–236, 1987.
5. C.G. Harris and M. Stephens. A combined corner and edge detector. In *Fourth Alvey Vision Conference*, pages 147–151, 1988.
6. R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. *Proc. Conference Computer Vision and Pattern Recognition*, 1992.
7. R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conference on Computer Vision*, pages 579–587. Springer-Verlag, 1992.
8. Q.T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix. Technical report 1894, INRIA, Sophia-Antipolis, France, 1993.
9. 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.
10. T. Moons, L. Van Gool, M. Van Diest, and A. Oosterlinck. Affine structure from perspective image pairs under relative translations between object and camera. Technical report KUL/ESAT/M12/9306, Departement Elektrotechniek, Katholieke Universiteit Leuven, 1993.
11. J.L. Mundy and A.P. Zisserman. *Geometric invariance in computer vision.* MIT Press, 1992.
12. J.G. Semple and G.T. Kneebone. *Algebraic projective geometry.* Oxford University Press, 1952.
13. R. Szeliski and S.B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. DEC technical report 93/3, DEC, 1993.
14. Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis.* Springer-Verlag, 1992.