Virtual-parallax based structure recovery : a framework for viewpoint selection

D. Sinclair, Roger Mohr

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Virtual-Parallax Based Structure Recovery: a Framework for Viewpoint Selection

David SINCLAIR
Roger MOHR

N° 2203
Février 1994

PROGRAMME 4

Robotique, image
et
vision

Virtual-parallax based structure recovery:- a framework for viewpoint selection

Estimation de structure par une parallaxe virtuelle: un cadre pour le choix d'un point de vue

David Sinclair, Roger Mohr
LIFIA–INRIA,
46 Av. Félix Viallet,
F- 38031 Grenoble Cedex, France

Abstract :

Traditionally solutions to the structure from motion problem with in computer vision concentrate on recovering distances from the camera or stereo pair to isolated points in the world. These depth estimates are are derived from point matches between two or more images. The structural information is typically returned in a camera centered coordinate frame and the data are often noisy. Recently however work has been done on relative positioning which does provide more robust results.

This report details an integrated system designed specifically for structure recovery and new view point selection. The method automaticaly selects a Euclidean basis for structure recovery centered on the object being studied. The basis is naturally aligned with the ground plane on which the object rests. Structure recovery is performed in an object centered way using '*virtual parallax*' between tracked corners on the object and the *virtual* motion of the coincident point on the ground plane. Tracking of the ground plane between viewpoints provides an ideal mechanism for combining structural information derived from multiple viewpoints.

This framework also makes it possible to determine whether the structural information the machine has estimated about the object is complete and if not permits a new viewing direction to be selected.
Keywords: structure from motion, view point selection, active vision, parallax.

Résumé :

L'approche traditionnelle de la perception du volume à partir du mouvement détermine les distances entre les points observés et la caméra. Cette estimation se fait à partir des correspondances dans une ou plusieurs images. L'information est donc centrée dans le repère du capteur et ces données sont largement bruitées. Il a

été montré plus récement que le positionnement relatif est une alternative robuste à cette approche.

Ce rapport fournit les détails d'un système conçu pour la détermination de la repésentation tridimensionnelle de la scène perçue et la détermination d'un nouveau point d'observation. L'approche fournit naturellement une base euclidienne centrée sur l'objet observé. Elle s'appuie sur le plan de base sur lequel repose cet objet. La structure tridimensionnelle est déterminée en utilisant la *parallaxe virtuelle* entre les points suivis sur l'objet et le mouvement virtuel des correspondants sur le plan de base. La poursuite de points de ce plan fournit un mécanisme élégant permettant de combiner l'information issue de plusieurs vues.

Le cadre développé ici permet aussi de déterminer si la structure tridimensionnelle estimée est complète, et sinon de déterminer un nouveau point d'observation.

Mots clés : s'election de point de vue, perception tridimensionnelle à partir du mouvement, vision active, parallaxe.

1 Introduction

The aim of this report is to provide the basis for a working structure from motion system that is able to look into a work-cell, segment any object in the cell from the base of the cell, recover the structure of the object, decide if viewing the object from another position will provide additional information, select a new viewing position and then integrate the information gained from the new position with the existing structural information. This report brings together several ideas from the structure from motion literature and combines them with some new ideas to produce a robust and flexible system for combining structure from motion type information from a series of viewpoints.

Existing structure from motion systems like 'Droid' [4, 5] effectively only provide a sparse camera centered depth map. The depth map dœs not use any surface models to improve its robustness and once a point has disappeared any depth information derived from it is discarded. The depth map exists instantaneously for a series of time steps and dœs not build up a model of the environment. It is desirable therefore to perform object reconstruction in an object centered rather than camera centered frame.

A more appropriate method might be based on the current work on projective reconstruction [3, 1, 7]. Five points within the scene are selected as a projective basis for the scene and reconstruction performed up to a colineation of P(3). The points must be chosen so that no four are coplanar. This method naturally incorporates information from multiple views provided the basis vectors remain visible and requires no camera calibration. This type of method dœs not provide a ready framework for object segmentation and is easily perturbed by false corner motion estimates. It is also unclear how new viewpoint selection would be determined when a scene is only reconstructed up to a colineation of P(3).

What is required is a structure recovery system that is able to combine structural information garnered from various viewpoints, which naturally models the environment onto which the robot looks (a work-cell with a planar base) and performs reconstruction in an object centered basis. The system must be able to cope with the output of the current generation of corner detectors/trackers and should be flexible enough to use the output generated by either a single camera or a stereo pair.

This report is laid out as follows. **Section 2** gives the notation used. **Section 3** describes the method used to solve the correspondence problem; corners are matched between frames using a *hysteresis* corner matcher developed in [9]. **Section 4** describes the recursive planer region detection method used to find the base-plane of the work cell. **Section 5** gives the method used for estimating the translation direction. The method uses a regression fit and hence also rejects any bad flow vectors that the tracker may provide. **Section 6** shows how the plane normal may be estimated unambiguously once the translation direction has been found. **Section 7** derives a novel and powerful result that relates *virtual parallax* to point height above the base-plane. Heights are recovered as a fraction of the camera height from the base-plane. In this method structure can be estimated without ever estimating camera rotation or the magnitude of translation. **Section 8** deals with the selection of a completely Euclidean basis for reconstruction. This basis is aligned with the base-plane of the robot work cell. Results for the proposed

method are laid out in **section 9** and a discussion given in **section 10.**

2 Notation

In this section all the notation used in this document is set out. The camera model is a **pin hole** and is assumed to be calibrated. The image plane is assumed to be in front of the center of projection. The additional assumption that the object to be reconstructed lies on the base-plane of a robot work cell is also made.

- $\mathbf{x}_i$ the position of the th corner in homogeneous coordinates (by default in image 1)
- $\mathbf{x}'_i$ the matched position of the th corner in homogeneous coordinates (by default in image 2)
- $\mathbf{X}$ the position of a point in the scene with respect to a camera centered coordinate frame
- $\mathbf{e}_z$ the optical axis of the camera
- $\mathcal{P}$ the projectivity describing the transformation of two images of the base- plane
- $\mathbf{T}$ the translation between camera center positions
- $\mathbf{t}$ the translation direction given in a specified camera centered frame
- $\mathbf{N}$ the normal to the base-plane of the robot work-cell
- $\mathbf{n}$ the direction of $\mathbf{N}$
- $\mathbf{R}$ the rotation matrix describing the 3D rotation between consecutive camera frames
- $\mathbf{I}$ the identity matrix
- $\mathbf{l}_i = \frac{\mathcal{P}\mathbf{x}_i \times \mathbf{x}'_i}{|\mathcal{P}\mathbf{x}_i||\mathbf{x}'_i|}$ the parallax vector associated with the motion of a tracked point off the base-plane
- h the height of a point above the ground plane

3 Solving the correspondence problem

The first stage in any structure from motion system is to estimate visual motion. Currently the most reliable means of doing this is to match corners between frames and feed the results to a tracker. In the system documented here corners are matched between frames using the hysteresis corner matcher detailed in [9]. The image based outlier rejection scheme given in [9] is also used.

4 Base-plane detection

At the heart of this structure from motion system is a robust plane detection algorithm [8]. Rather than use the seed point region growing method suggested in [8], the assumption that at least part of the lower half of the field of view of the initial camera position is occupied by the ground plane is made. A projectivity is then fitted to the point motions in the lower half of the field of view. This projectivity is then used as the starting point for a regression fit for the projectivity describing the motion of the base of the robot work-cell.

If the projectivity between the two frames is $\mathcal{P}$ then,

$$\hat{\mathbf{x}} = \lambda \mathcal{P} \mathbf{x}, \tag{1}$$

with λ constrained so that $\hat{\mathbf{x}} \cdot \mathbf{e}_z = f$ where f is the focal length in pixels and $\mathbf{e}_z$ the optical axis of the camera. This makes the following analysis convenient because corner position variances may then be given in pixels. The linearised variance of the x component of the predicted position $\hat{\mathbf{x}}$ is

$$\mathrm{Var}(\hat{x}) = \begin{pmatrix} \frac{\partial \hat{x}}{\partial x} & \frac{\partial \hat{x}}{\partial y} \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} \frac{\partial \hat{x}}{\partial x} \\ \\ \frac{\partial \hat{x}}{\partial y} \end{pmatrix}, \tag{2}$$

where

$$\frac{\partial \hat{x}}{\partial x} = \frac{f\mathcal{P}_{11}}{\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f} - (\mathcal{P}_{11}x + \mathcal{P}_{12}y + \mathcal{P}_{13}f)\frac{f\mathcal{P}_{31}}{(\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f)^2} \tag{3}$$

$$\frac{\partial \hat{x}}{\partial y} = \frac{f\mathcal{P}_{12}}{\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f} - (\mathcal{P}_{11}x + \mathcal{P}_{12}y + \mathcal{P}_{13}f)\frac{f\mathcal{P}_{32}}{(\mathcal{P}_{31}x + \mathcal{P}_{32}y + \mathcal{P}_{33}f)^2}. \tag{4}$$

The corresponding variance in $\hat{y}$ is derived analogously. Points are accepted as being on the plane associated with the projectivity if the difference between actual and predicted position in the second frame is less than two standard deviations in the x or y directions. This assumes that the error in a corner's position is isotropic and Gaussian distributed. That is,

$$|\hat{x} - x'| \leq 2\sqrt{\mathrm{Var}(\hat{x})} \tag{5}$$

and

$$|\hat{y} - y'| \leq 2\sqrt{\mathrm{Var}(\hat{y})}. \tag{6}$$

The parameter σ represents the acceptable error in a points position and this parameter is reduced during the regression fit. For the example in the results section of this paper initially all the points in the lower half of the field of view were assumed to lie on the base-plane and there motions used to compute an initial estimate of $\mathcal{P}$. With this estimate of $\mathcal{P}$ the regression fitting process converged after 2 iterations.

The list of matched points has now been divided into points that are part of the base-plane and points off the base-plane. For the purposes of this report points off the base-plane are assumed to be part of the object of interest.

5 Solution for translation direction

A weighted least squares regression fit is used to solve for the motion direction. The method used follows [2] in using virtual parallax between points off the base-plane and a points virtual position on the base-plane under camera translation.

Again $\mathbf{x}_i$ is the position of corner i in image 1, $\mathbf{x}'_i$ corresponding position of corner i in image 2, $\mathcal{P}\mathbf{x}_i$ virtual position of corner i assuming it is on the base-plane in image 2, if $\mathbf{t}$ is the translation direction in the coordinate frame of camera 2 then all the vectors $\mathcal{P}\mathbf{x}_i - \mathbf{x}'_i$ lie in epipolar planes. It follows that

$$(\mathcal{P}\mathbf{x}_i - \mathbf{x}'_i) \times \mathbf{x}'_i \cdot \mathbf{t} = 0 \tag{7}$$

and simplifying

$$(\mathcal{P}\mathbf{x}_i \times \mathbf{x}'_i) \cdot \mathbf{t} = 0. \tag{8}$$

The parallax vector $\mathbf{l}_i$ is then defined to be

$$\mathbf{l}_i = \frac{\mathcal{P}\mathbf{x}_i \times \mathbf{x}'_i}{|\mathcal{P}\mathbf{x}_i|\,|\mathbf{x}'_i|}. \tag{9}$$

The magnitude of the parallax vectors $\mathbf{l}_i$ forms a natural weighting for the following least squares solution for the translation direction $\mathbf{t}$,

$$\sum_i \mathbf{t}^T(\mathbf{l}_i\mathbf{l}_i^T)\mathbf{t} = 0, \tag{10}$$

or $\mathbf{t}$ is the eigenvector of the matrix $\sum_i \mathbf{l}_i\mathbf{l}_i^T$ with the smallest eigenvalue.

In practice a regression fit is made for $\mathbf{t}$ using the all corner motions that are not on the base-plane. If the angle between $\mathbf{t}$ and $\mathbf{l}_i$ differs from π by more than a threshold then $\mathbf{l}_i$ is not used in the estimation of $\mathbf{t}$.

$$\frac{\mathbf{t} \cdot \mathbf{l}_i}{|\mathbf{t}|\,|\mathbf{l}_i|} < \text{thresh} \tag{11}$$

This does not necessarily mean that a point is an outlier as the direction of small parallax vectors may will be vulnerable to noise in corner location. This does not mean that the height information encoded in the parallax vector is unreliable. Outlier rejection must take into account the magnitude of $\mathbf{l}_i$. Outlier rejection using this method takes the form

$$\frac{\mathbf{t} \cdot \mathbf{l}_i}{|\mathbf{t}|} < \text{thresh}_2, \tag{12}$$

where thresh_2 embodies an estimate of the uncertainty in a corners position

6 Solution for base-plane normal

Two pieces of information have now been estimated, the projectivity $\mathcal{P}$ describing the motion of the base-plane and the translation direction $\mathbf{t}$. The expression for the projectivity

$\mathcal{P}$ that maps one image of the plane with normal $\mathbf{N}$ into its image taken after a translation $\mathbf{T}$ and a rotation R is derived as follows ([6]. Let $\mathbf{X}$ be the world coordinates of a point and $\mathbf{x}$ be the homogeneous image coordinates. If the coordinate frame is translated by $\mathbf{T}$ and rotated by R then $\mathbf{X}$', the new coordinates are as follows

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{T}) \tag{13}$$

substituting $\mathbf{X} \cdot \mathbf{N} = 1$, (the equation of the base-plane in world coordinates), gives

$$\mathbf{X}' = R(\mathbf{X} - \mathbf{T}\mathbf{X} \cdot \mathbf{N}) \tag{14}$$

and taking homogeneous coordinates

$$\mathbf{x}' = R(I - \mathbf{T}\mathbf{N}^T)\mathbf{x}, \tag{15}$$

hence

$$\mathcal{P} = R(I - \mathbf{T}\mathbf{N}^T). \tag{16}$$

This expression for the projectivity $\mathcal{P}$ makes the speed scale ambiguity explicit. So long as the magnitudes of $\mathbf{T}$ and $\mathbf{N}$ are scaled in inverse proportion $\mathcal{P}$ is unchanged. The substitution

$$\mathbf{t}\mathbf{n}^T = \mathbf{T}\mathbf{N}^T \tag{17}$$

is made, where $\mathbf{t}$ and $\mathbf{n}$ are the scaled translation direction and plane normal giving

$$\mathcal{P} = R(I - \mathbf{t}\mathbf{n}^T). \tag{18}$$

The plane normal is estimated as follows; if $\mathbf{u}$ is a unit vector and

$$\mathbf{u} \perp \mathbf{n}, \tag{19}$$

then

$$\mathbf{u}^T\mathcal{P}^T\mathcal{P}\mathbf{u} = \mathrm{c}, \tag{20}$$

where c is a constant as $\mathbf{u} \cdot \mathbf{n} = 0$. There are three cases to consider based on the eigenvalues of $\mathcal{P}^T\mathcal{P}$. Let λ_1, λ_2, λ_2 and $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ respectively be the eigenvalues and eigenvectors of $\mathcal{P}^T\mathcal{P}$. Assume additionally that $\mathcal{P}^T\mathcal{P}$ has been rescaled so that $\lambda_2 = 1$ and that the eigenvalues are arranged in ascending order. Let α be a vector

$$\alpha = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \alpha_3\mathbf{e}_3, \tag{21}$$

then

$$\alpha^T\mathcal{P}^T\mathcal{P}\alpha = \alpha_1^2\lambda_1 + \alpha_2^2\lambda_2 + \alpha_3^2\lambda_3 = 1 \tag{22}$$

is an ellipsoid centered on the origin. The major axes of the ellipsoid are given by the eigenvectors of $\mathcal{P}^T\mathcal{P}$. The plane of directions of $\mathbf{u}$ will be the plane through the origin that intersects the ellipsoid in a circle. Depending on the degeneracy of the eigenvalues there are three possible cases.

The first case is that all the eigenvalues are equal and no solution exists (this case corresponds to zero translation). The second case is when two of the eigenvalues are

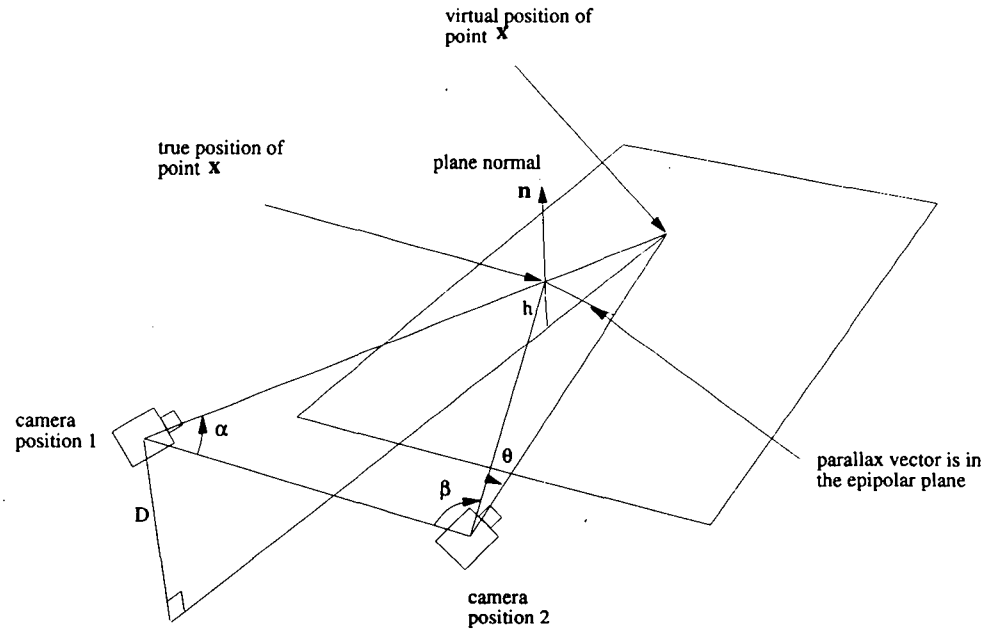

Figure 1: *Schematic representation of the parallax of a point relative to a plane. The angles α, β and θ are respectively the angel between a point* $\mathbf{x}$ *and the motion direction in frame 1, the angle between the motion direction and the true position of point* $\mathbf{x}$ *in frame 2 and the angle or parallax between the virtual position and real position of* $\mathbf{x}$ *in frame 2. D is the distance of the camera from the base-plane and h the height of the point* $\mathbf{x}$ *above the plane.*

equal, in which case there is one solution (there is only one plane through the origin that intersects the figure in a circle). The normal to this plane, and hence $\mathbf{n}$, is the non degenerate eigenvector. The third and final case is when all three eigenvalues are distinct. There are two possible planes which cut the ellipsoid in circles. Each of the two possible planes passes through $\mathbf{e}_2$ the eigenvector of the middle eigenvalue. The angle between the plane and $\mathbf{e}_3$ in the plane of $\mathbf{e}_1$ and $\mathbf{e}_3$ is θ where,

$$\cos(\theta) = \pm\sqrt{\frac{1-\lambda_1}{\lambda_3-\lambda_1}}. \tag{23}$$

The directions of the two possible plane normals $\mathbf{n}$ are given by,

$$\mathbf{n} = \mathbf{e}_2 \times (\cos(\theta)\mathbf{e}_3 \pm \sin(\theta)\mathbf{e}_1) \tag{24}$$

Assuming that the third case is true then two possible plane normal directions have been found. It remains to choose the correct one and fix its scale with respect to $\mathbf{t}$ and the sign of its direction.

The sign of the direction of $\mathbf{n}$ is easily fixed by employing the visibility constraint: $\mathbf{n} \cdot \mathbf{e}_z < 0$. For each of the two candidate directions for $\mathbf{n}$ the magnitude of $\mathbf{n}$ relative to $\mathbf{t}$ may be determined by noting that

$$\text{Trace}(\mathcal{P}^T\mathcal{P}) = (3 - 2\mathbf{t}\cdot\mathbf{n} - \mathbf{n}\cdot\mathbf{n}), \tag{25}$$

or alternatively,

$$\mathbf{t}^T(\mathcal{P}^T\mathcal{P})\mathbf{t} = |\mathbf{t}|^2(1 - \mathbf{n}\cdot\mathbf{t})^2 \tag{26}$$

and substituting for $\mathbf{t}$ and the direction of $\mathbf{n}$ and then solving for the magnitude of $\mathbf{n}$. With the two pairs of estimated values of $\mathbf{t}$ and $\mathbf{n}$ two projectivities $\hat{\mathcal{P}}$ can be constructed and the dot product between the following two matrices calculated ,

$$j(\mathbf{n}) = \frac{(\mathcal{P}^T\mathcal{P})}{|(\mathcal{P}^T\mathcal{P})|} \cdot \frac{(\hat{\mathcal{P}}^T\hat{\mathcal{P}})}{\left|(\hat{\mathcal{P}}^T\hat{\mathcal{P}})\right|}. \tag{27}$$

The correct $\mathbf{n}$ will be the one that produces the larger dot product j.

7 Virtual-parallax based structure recovery

The information now available is the projectivity $\mathcal{P}$ describing the motion of the base-plane the relatively scaled values of camera translation $\mathbf{t}$ and plane normal $\mathbf{n}$ and the parallax of points on the object. It remains to determine the height h of a given point on the object above the plane and the coordinates of the point wrt the base-plane. The relation between the height of an object from the base plane and the parallax of a point is derived in this section. Figure 1 shows a representation of the parallax of a point at height h above the base-plane.

Using the sine rule and similar triangle the ratio between height h and height D is, (appendix A),

$$\frac{h}{D} = \frac{\sin(\theta)\sin(\alpha)}{\sin(\alpha+\beta)\sin(\beta+\theta)} \tag{28}$$

where,

$$\theta = \sin^{-1}(\left|\mathbf{l}_i \times \left(\frac{\mathbf{t}}{|\mathbf{t}|}\right)\right|) \tag{29}$$

the parallax (projection of the angle between real and virtual image position of point $\mathbf{x}$ in image 2 onto the epipolar plane through $\mathbf{x}$'), α is the angle between point $\mathbf{x}$ and the motion direction in frame 1 and β is the angle between the motion direction and the true position of point $\mathbf{x}$ in frame 2.

If $\mathbf{t}$ is the motion direction relative to frame 2 then

$$\mathbf{t}' = \mathcal{P}^{-1}\mathbf{t} \tag{30}$$

will be the motion direction relative to frame 1.

This method of structure estimation has the advantage that the rotation of the camera is never explicitly computed.

8 Euclidean basis selection for reconstruction

The 'natural' basis to chose for object reconstruction is an orthogonal set of axes with two of the axes parallel to the base-plane of the work-cell and the third representing

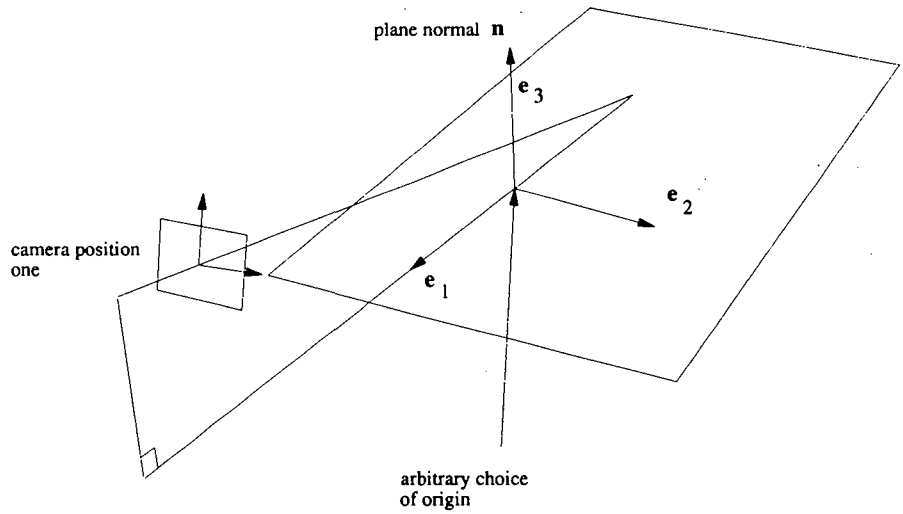

Figure 2: *The choice of origin for reconstruction is arbitrary. Orthogonal axes are chosen. Axis* $\mathbf{e}_1$ *is in the base plane and its projection is parallel to the plane normal* $\mathbf{n}$. *Axis* $\mathbf{e}_2$ *is in the base plane and its projection is perpendicular to the projection of* $\mathbf{e}_1$. *Axis* $\mathbf{e}_3$ *is parallel to the plane normal* $\mathbf{n}$.

height, figure 2. The coordinate system for reconstruction must be chosen in the first frame. Orthogonality of chosen axes can be guaranteed by aligning the first axis ($\mathbf{e}_1$) with the projection of the plane normal in frame 1 but in the base-plane. The second axis ($\mathbf{e}_2$) is chosen to be perpendicular to the projection of the first but also in the base-plane. The third axis ($\mathbf{e}_3$) is then the normal to the base-plane. The relative scale of the axes remains to be fixed.

8.1 Recovering a point's position relative to the base-plane

An expression for the scaled height of a point above the base-plane was derived in section 7. In this section the point of intersection between the vector through a point $\mathbf{x}$ parallel to the base-plane normal $\mathbf{n}$ and the base-plane is determined. This is then the coordinates of a point with respect to the base-plane, figure 3. It is sufficient to give these coordinates in terms of image coordinates as all images of the base plane are related by projectivities. The projectivity that transforms a particular image of the base plane to the chosen reconstruction frame will also transform the plane coordinates of a point into this frame. The angle η (equivalent to the base-plane coordinate of $\mathbf{x}$) in figure 2 is related to h, D and ϕ by, (appendix B),

$$\eta = \tan^{-1}\left(\tan(\phi)\frac{1}{(1-\frac{h}{D})}\right) - \phi \tag{31}$$

where the ratio $\frac{h}{D}$ is defined in (28).

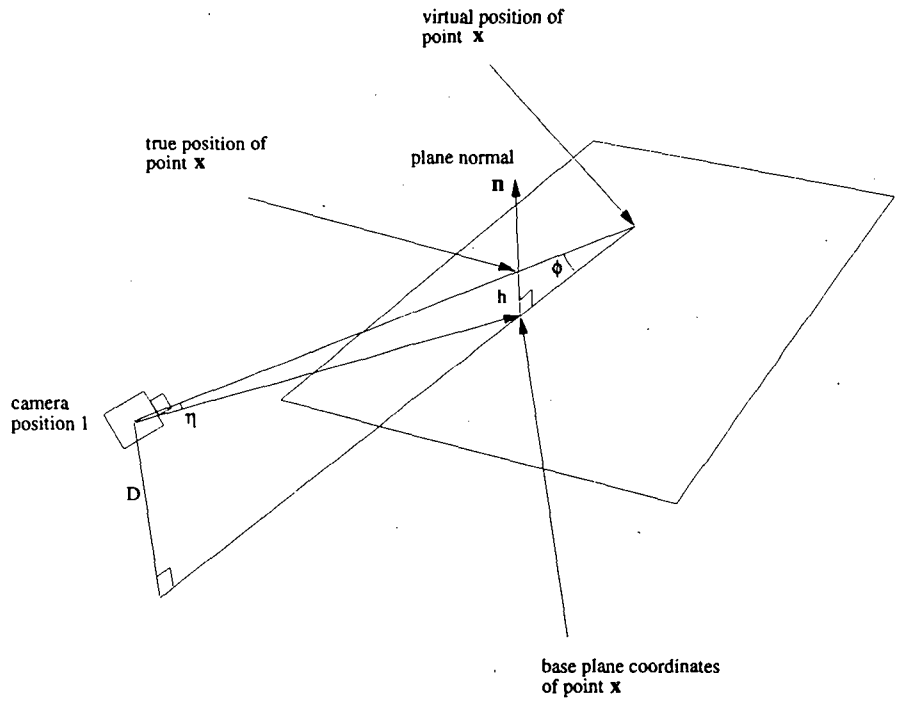

Figure 3: *The angle η specifies the plane coordinates of a point* $\mathbf{x}$ *at height h and ϕ is the inclination of the base-plane.*

The method so far recovers, the projectivity $\mathcal{P}$ describing the transformation of the ground plane, the translation direction $\mathbf{t}$, the plane normal $\mathbf{n}$ (scaled wrt $\mathbf{t}$) and the heights h of points $\mathbf{x}$ above the base plane (scaled with $\mathbf{n}$) for a series of point matches between a pair of images. If more than one pair of images are available then the method provides a very natural framework for integrating depth estimates from multiple view points.

9 Results

Figure 4 shows the first image of a motion sequence. The scene consists of a rectangular box standing on top of a textured plane. Corners were detected and tracked through the image sequence using the hysteresis corner tracker mentioned in section 3. Figure 5 shows the second image of the sequence but with detected corners superimposed on it. The corners were detected using the Wang-Brady corner detector [10]. The resulting flow vectors are shown in figure 6. The only camera calibration used was the manufacturers estimate of focal length and aspect ratio. The optical axis was assumed to pass through the center of the field of view.

Figure 7 shows three projections of the reconstructed object based on the matched corner motions in figure 6. If fewer outliers are required then the thresholds described in section 5 may be tightened up.

Structural information derived from a second view point is shown in figure 8. The

Figure 4: *First image of a motion sequence. The scene consists of a rectangular box standing on top of a textured plane.*

base-plane of the work cell was tracked between the first and second vantage points. The structural information is then transformed into the reference frame of image 1 using the method given in section ??. The results of the integration can be seen in figure 9. The registration between the structural derived from the two distinct vantage points is good.

10 Discussion

The results section demonstrates the method for structure recovery and the integration of structural information gathered from two distinct view points. The structure recovery is sufficiently accurate for the reader to recognise the object on the plane as being box-like.

The method has several desirable properties. The first is that camera rotation is never explicitly calculated. The magnitude of the translation direction is not required when estimating point heights for a stereo pair taken from a particular vantage point. The basis for reconstruction is naturally aligned with the base-plane of the work cell and can be chosen to be not only orthogonal but genuinely Euclidean. The tracked corners used for reconstruction need only be seen in two consecutive frames, points disappearing or reappearing through occlusion pose no problem. If points have been tracked through multiple framed then extremely accurate relative depth estimates may be obtained using the resulting larger effective stereo base line.

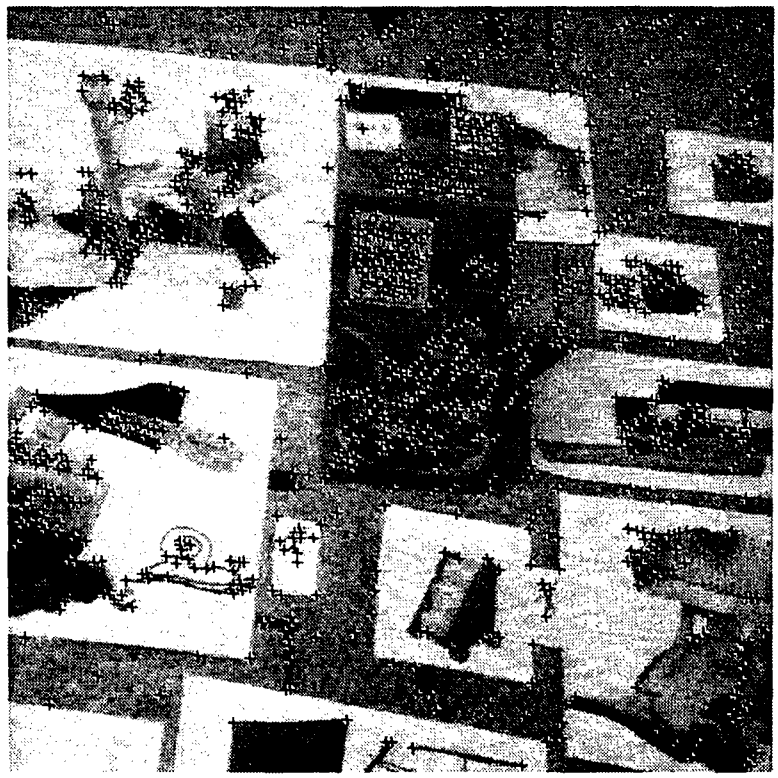

Figure 5: *Second image of the sequence but with detected corners superimposed on it. The corners were detected using the Wang-Brady corner detector.*

The ability to integrate information from multiple view points is a very powerful tool for object reconstruction. No single vantage point will let all of an object be seen. The selection of new vantage points from which to view an object as it is being reconstructed is unfortunately beyond the scope of this report however the structure integration framework proposed here is a very natural starting point.

In this report the frame of the first image is taken as the frame for reconstruction. It would be equally possible to choose the current frame and bring structural information forward from previous frames. This would allow the system to be used by an AGV moving on a flat floor in a cluttered environment for example. This system would permit the AGV to remember obstacles that it was no longer looking at.

The primary disadvantage of the system is that the object to be reconstructed must lie on a plane and there must be trackable features on the plane.

Acknowledgements

We are grateful to the EC for providing the funding for the Human Capital Mobility program. We would like to thank Paul Beardsley and Larry Shapiro for the benefit of their experience and provision of robust C code.

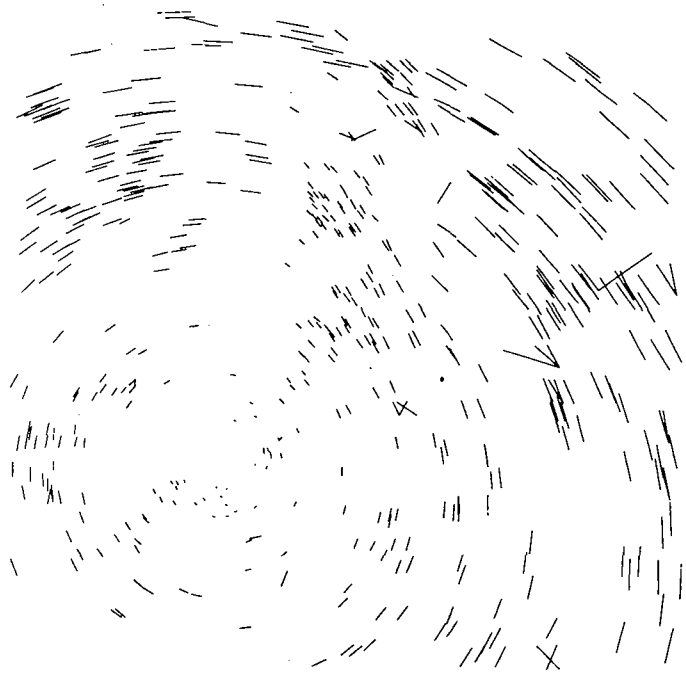

Figure 6: *Corner motions found for tracked corners between the two images in the previous two images.*

References

[1] P.A. Beardsley. *Applications of projective geometry to robot vision.* PhD thesis, Dept. of Engineering Science, University of Oxford, 1992.

[2] P.A. Beardsley, D. Sinclair, and A. Zisserman. Ego-motion from six points. Insight meeting, Catholic University Leuven, 1992.

[3] B. Boufama, R. Mohr, and F. Veillon. Euclidian constraints for uncalibrated reconstruction. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 466–470, 1993.

[4] C.G. Harris. Determination of ego - motion from matched points. In *3rd Alvey Vision Conference*, pages 189–192, 1987.

[5] C.G. Harris and J.M. Pike. 3d positional integration from image sequences. In *3rd Alvey Vision Conference*, pages 233–236, 1987.

[6] H.C. Longuet-Higgins. The visual ambiguity of a moving plane. *Proc.R.Soc.Lond.*, B223:165–175, 1984.

[7] L. Robert and O. Faugéras. Relative 3D positionning and 3D convex hull computation from a weakly calibrat ed stereo pair. In *Proc. 4th Int. Conf. on Computer Vision*, pages 540–544, 1993.

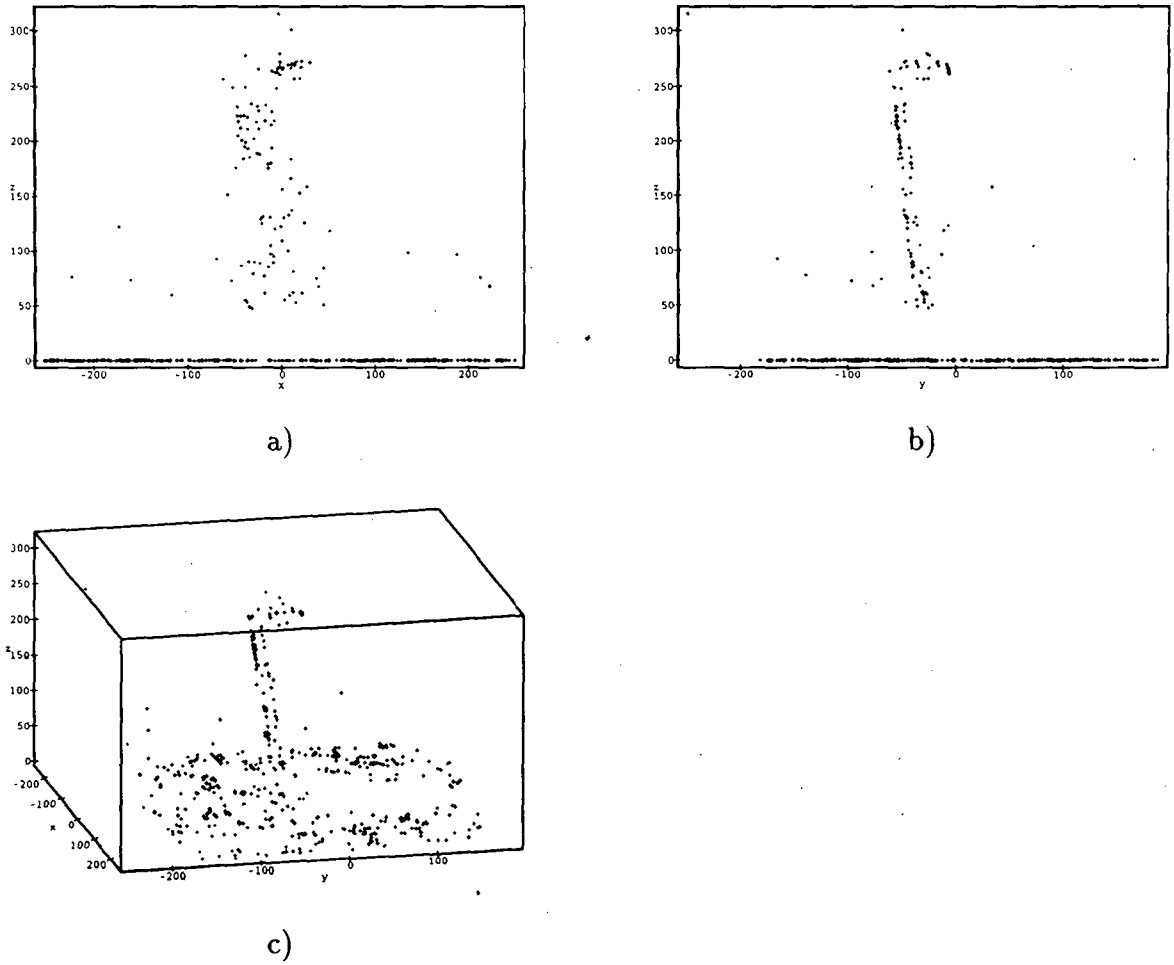

Figure 7: *Three projections of the reconstructed object based on the matched corner motions in figure 6. The structure of the box can clearly be seen as can the gap in the data where the box obscures the base-plane.*

[8] D. Sinclair. *Experiments in Motion and Correspondence.* PhD thesis, Oxford University, 1993.

[9] D. Sinclair, R. Tamstorf, J Crowley, and P Beardsley. Hysteresis corner matching and image based outlier rejection. Technical report, LIFIA - INRIA, 1994.

[10] H. Wang. Corner detection for 3D vision using array processors. In *Proc. BARNAIMAGE 91*, Barcelona, 1991.

A

The relation between a points height above the base plane and the camera height above the plane is derived here. Figure 10 shows the camera base-plane configuration. Using

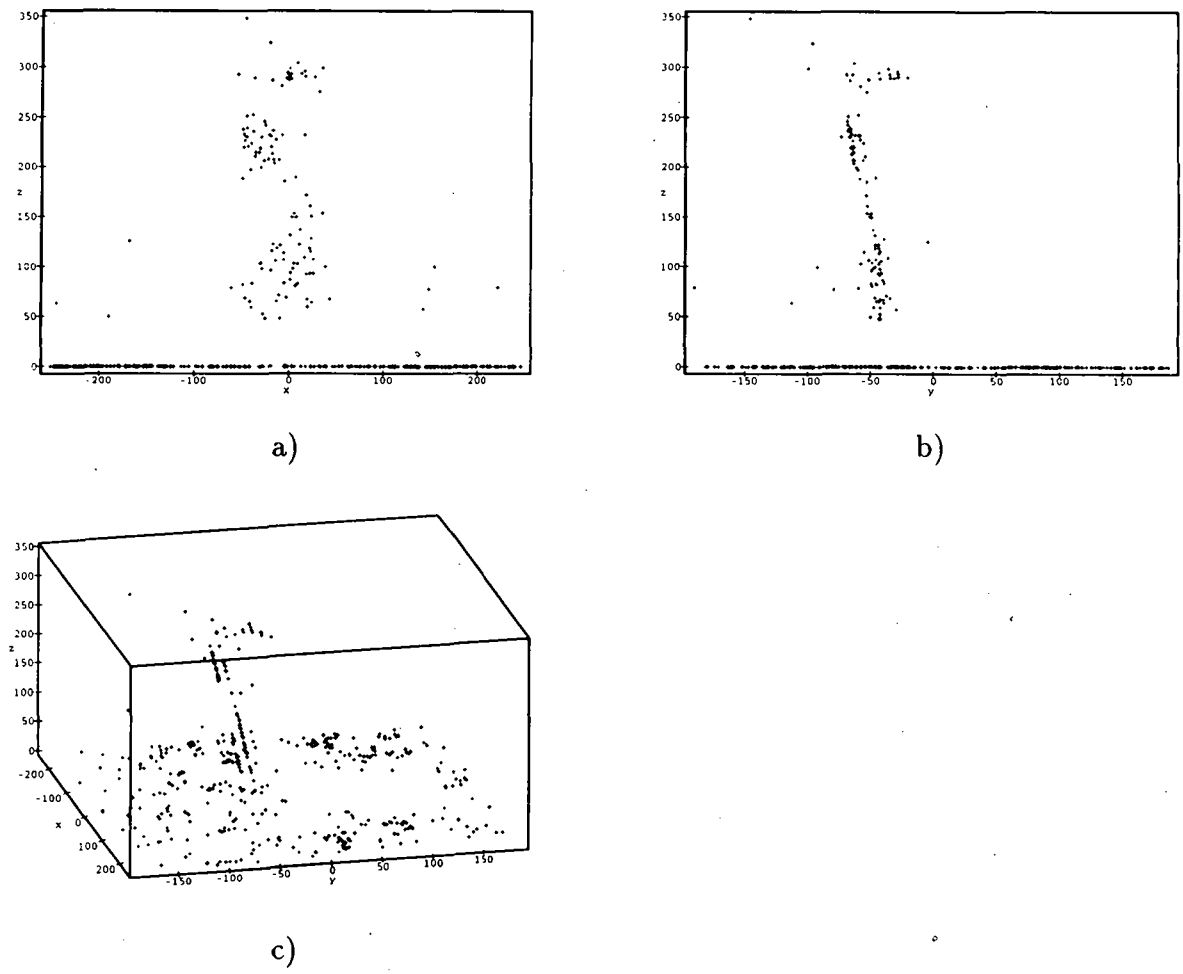

Figure 8: *Three projections of the reconstructed object based on a second set of matched corner motions. Again the reconstruction closely resembles the object. The base-plane of the work cell was tracked between the first and second vantage points.*

the property of similar triangles,

$$\frac{h}{D} = \frac{a}{b} \tag{32}$$

and applying the sine rule to the triangles in the epipolar plane;

$$\frac{a}{c} = \frac{\sin(\theta)}{\sin(\pi - \gamma - \theta)} \tag{33}$$

$$\frac{b}{c} = \frac{\sin(\beta + \theta)}{\sin(\alpha)} \tag{34}$$

where

$$\gamma = \pi - \alpha - \beta - \theta \tag{35}$$

substituting for $\frac{a}{b}$ in (32) and simplifying gives

$$\frac{h}{D} = \frac{\sin(\theta)\sin(\alpha)}{\sin(\alpha + \beta)\sin(\beta + \theta)} \tag{36}$$

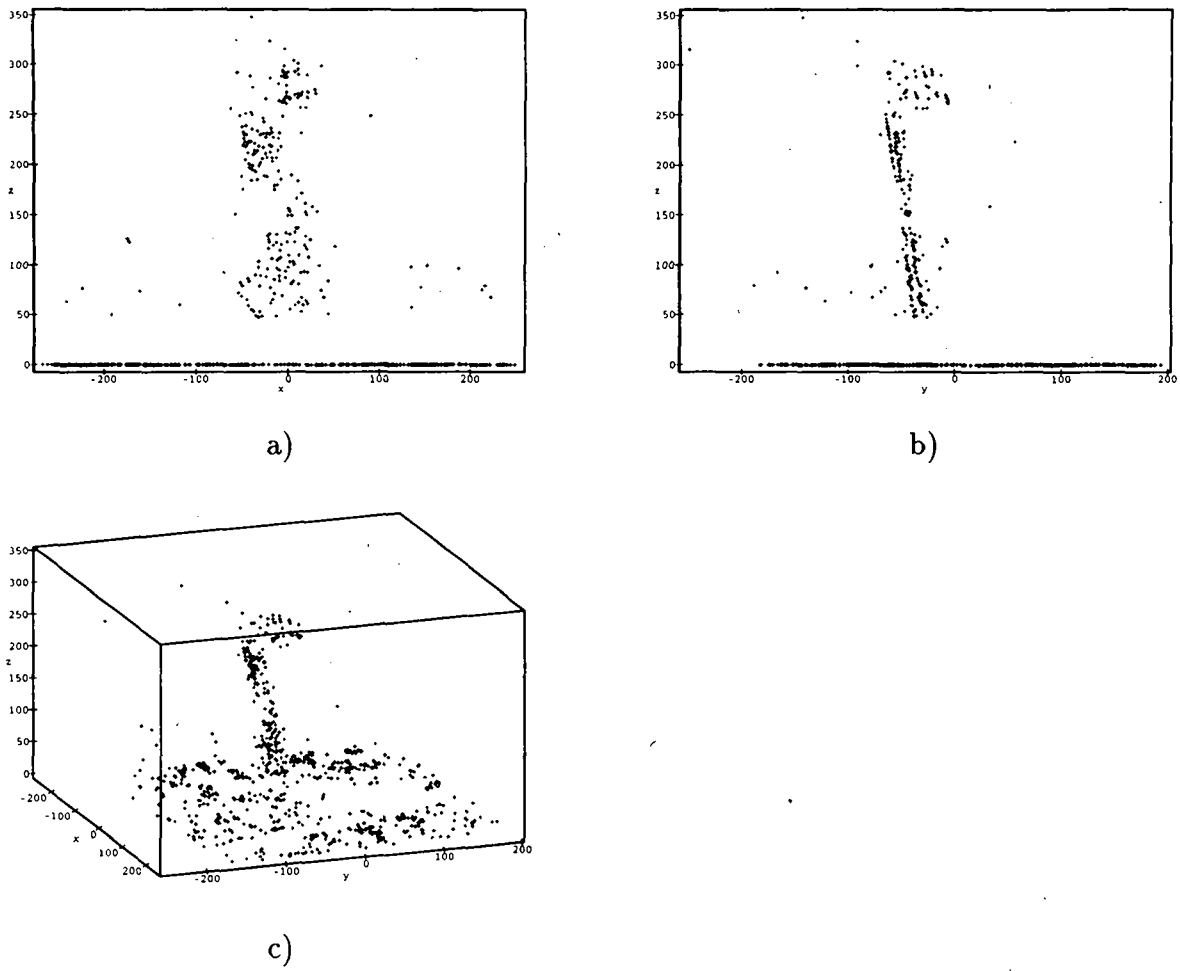

Figure 9: *Results of the integration of the structure derived in figures 7 and 8. The registration between the two estimates of the structure is good.*

B

In this section the expression for the base-plane coordinates of a point above the plane given in section 8.1 is derived. Figure 11 shows the camera base-plane configuration and again similar triangles are used. The base-plane coordinates of the point $\mathbf{x}$ are uniquely specified if the angle η and the plane normal $\mathbf{n}$ are known. using similar triangles and the definition of tangent gives,

$$\frac{h}{D} = \frac{a}{b}, \tag{37}$$

$$\tan(\alpha) = \frac{D}{(b-a)}, \tag{38}$$

$$\tan(\phi) = \frac{D}{(b)}, \tag{39}$$

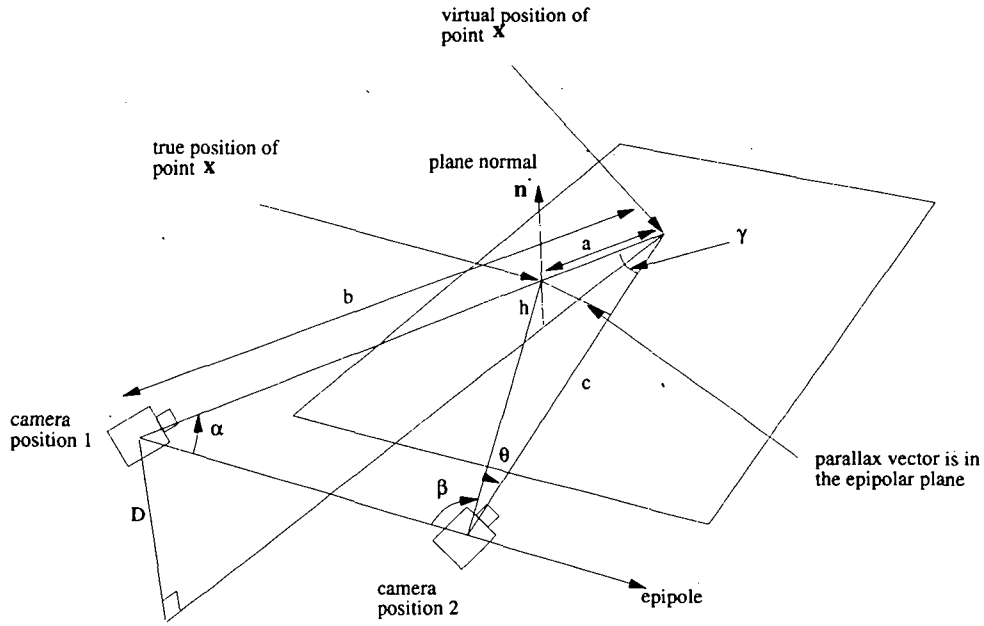

Figure 10: *Schematic representation of the parallax of a point relative to a plane. The angles α, β, γ and θ are as shown. Similar triangles and the sine rule are used to define the ratio $\frac{h}{D}$ in terms of these angles.*

noting that

$$\eta = \alpha - \phi \tag{40}$$

gives

$$\eta = \tan^{-1}\left(\tan(\phi)\frac{1}{(1-\frac{h}{D})}\right) - \phi. \tag{41}$$

The image based coordinates of the base-plane coordinates of the point $\mathbf{x}$ are then found by rotating $\mathbf{x}$ by η in the plane of $\mathbf{x}$ and $\mathbf{n}$. The mapping between image coordinates and any basis chosen for ground plane coordinates will be given by a projectivity. It is this property that enables structural information from differing view points to be effectively combined.

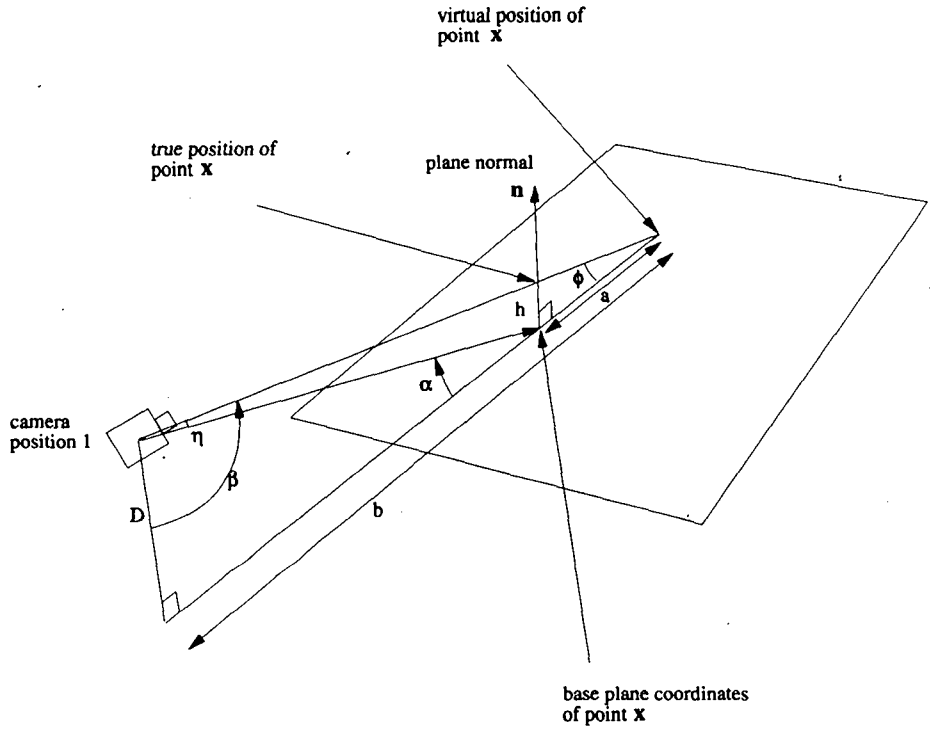

Figure 11: *The angle η specifies the plane coordinates of a point* $\mathbf{x}$ *at height h and ϕ is the inclination of the base-plane. The angles α and β are as shown.*

Unité de Recherche INRIA Rhône-Alpes
46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)

Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rocquencourt Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399