International Journal of Computer Vision, 16, 147–182 (1995)

3D Motion Recovery via Affine Epipolar Geometry

LARRY S. SHAPIRO shapiro@robots.ox.ac.uk
ANDREW ZISSERMAN az@robots.ox.ac.uk
MICHAEL BRADY jmb@robots.ox.ac.uk
Robotics Research Group, Department of Engineering Science, Oxford University, Parks Road, Oxford, OX1 3PJ, UK

Abstract. Algorithms to perform point-based motion estimation under orthographic and scaled orthographic projection abound in the literature. A key limitation of many existing algorithms is that they operate on the minimum amount of data required, often requiring the selection of a suitable minimal set from the available data to serve as a "local coordinate frame". Such approaches are extremely sensitive to errors and noise in the minimal set, and forfeit the advantages of using the full data set. Furthermore, attention is seldom paid to the statistical performance of the algorithms.

We present a new framework that allows *all* available features to be used in the motion computations, without the need to select a frame explicitly. This theory is derived in the context of the *affine camera*, which preserves parallelism and generalises the orthographic, scaled orthographic and para-perspective models. We define the affine epipolar geometry for two such cameras, giving the fundamental matrix in this case. The noise resistant computation of the epipolar geometry is discussed, and a statistical noise model constructed so that confidence in the results can be assessed.

The rigid motion parameters are then determined *directly* from the epipolar geometry, using the novel rotation representation of Koenderink and van Doorn (1991). The two-view partial motion solution comprises the scale factor between views, the projection of the 3D axis of rotation and the cyclotorsion angle, while the addition of a third view allows the true 3D rotation axis to be computed (up to a Necker reversal). The computed uncertainties in these parameters permit optimal estimates to be obtained over time by means of a linear Kalman filter. Our theory extends work by Huang and Lee (1989), Harris (1990), and Koenderink and van Doorn (1991), and results are given on both simulated and real data.

1 Introduction

Orthographic and scaled orthographic projection are widely used in computer vision to model the imaging process (Ullman 1979; Bennett et al. 1989; Huang and Lee 1989; Harris 1990; Koenderink and van Doorn 1991; Longuet-Higgins 1991; Ullman and Basri 1991; Tomasi and Kanade 1992). They provide a good approximation to the perspective projection model (the pinhole camera) when, as shown in Fig. 1, the field of view is small and the variation in depth of the scene along the line of sight is small compared to its average distance from the camera (Thompson and Mundy 1987). More importantly, they expose the ambiguities that arise when perspective effects diminish. In such cases, it is not only *advantageous* to use these simplified models but also *advisable* to do so, for by explicitly incorporating these ambiguities into the algorithm, one avoids computing parameters that are inherently ill-conditioned (Harris 1990).

The *affine camera*, introduced by Mundy and Zisserman (1992), generalises the orthographic, scaled orthographic and para-perspective models. It is the natural projection of a 3D affine space to a 2D affine image. For example, parallelism is preserved, so that parallel lines in the scene project to parallel lines in the image. One particularly useful way to envisage the affine camera is as an *uncalibrated* scaled orthographic camera, for although Euclidean measurements (e.g., image angles and distances) are only meaningful with a calibrated camera (i.e., known focal length, aspect ratio and principal point), various affine properties can still be measured without requiring arduous and often ill-conditioned calibration (such as parallelism, ratios of lengths in parallel directions and ratios of areas on parallel planes (Berger 1980)). Such properties are often sufficient for vision tasks. For instance, Beardsley et al. (1994) navigate on the basis of affine structure, Hollinghurst and Cipolla (1993) grasp and manipulate objects using affine stereo, and Reid et al. (1993) define

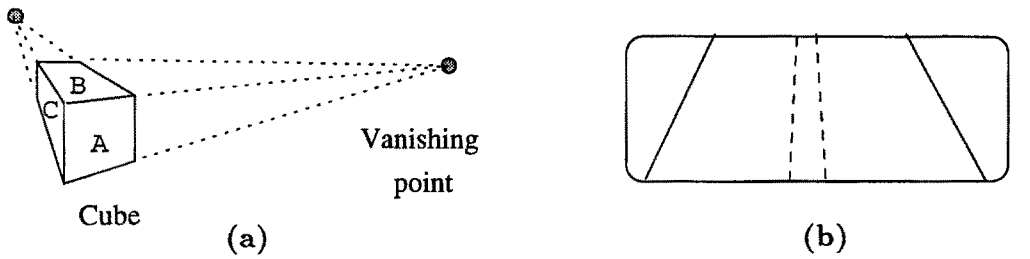

Fig. 1. Perspective projection preserves parallel lines when the object is relatively shallow and the field of view is small: (a) face A is more parallelogram-like than faces B and C, since it has a smaller depth variation; (b) road markings viewed through a wind screen show that parallelism is distorted less for a narrow field of view (dashed lines) than for a wide one (solid lines).

a gaze point for fixating a robot head using only affine constructions.

The use of the affine camera therefore enables the identification of quantities that can be computed under (scaled) parallel projection without calibration, such as affine epipolar geometry (cf. Section 3) and new views of an object (Koenderink and van Doorn 1991; Demey et al. 1992; Shapiro 1993). When calibration *is* needed, the precise stage at which it must be introduced into the computation can be determined, e.g., aspect ratio is required to compute rigid motion parameters (Section 4). This approach echoes the "stratification" philosophy of Koenderink and van Doorn (1991).

This paper investigates the *motion estimation* problem in the context of the affine camera. Given m distinct views of n points located on a rigid object viewed under parallel projection, the task is to compute 3D motion without any prior 3D knowledge. There are several reasons why many existing point-based motion algorithms are of limited practical use. First, the inevitable presence of 'noise' and matching errors is often ignored (Longuet-Higgins 1981; Koenderink and van Doorn 1991). Second, unreasonable demands are often made on prior processing, e.g., a "suitable" perceptual frame must first be selected (Koenderink and van Doorn 1991; Weinshall and Tomasi 1993), or the features must be tracked through every frame (Weinshall and Tomasi 1993). Third, algorithms often only work in special cases, e.g., no rotation about a fixed axis (Huang and Lee 1989; Hu and Ahuja 1991; Longuet-Higgins 1991). Finally, some algorithms require batch processing (Tomasi and Kanade 1992), rather than more natural sequential processing.

The tool we use to tackle this problem is affine epipolar geometry. Although the epipolar constraint has been widely used in perspective and projective motion applications (e.g., to aid correspondence (Charnley et al. 1988; Beardsley et al. 1993), compute rigid motion parameters (Longuet-Higgins 1981), segment moving objects (Torr et al. 1993), and determine calibration parameters (Faugeras et al. 1992; Hartley 1992)), it has seldom been used under *affine* viewing conditions (though see Huang and Lee 1989; Xu et al. 1993). Least squares formulations are employed throughout to utilise *all* available points, not just a minimum set. This improves the accuracy of the motion solution (by providing immunity to noise and enabling detection of outliers) and also obviates the need to *select* a local coordinate frame (LCF). The resulting n-point frame work subsumes the results for minimum configurations.

This paper makes the following contributions:

- The affine camera is related to other familiar models and shown to subsume the orthographic, scaled orthographic and para-perspective camera models (Section 2).
- The epipolar geometry of the *affine camera* is defined and its special fundamental matrix derived (Section 3). No camera calibration is needed at this juncture.
- A robust solution is provided for the epipolar geometry parameters, which underpin subsequent rigid motion computations. Three least squares algorithms based on image distances are evaluated, and a 4D linear method is shown to be both optimal and equivalent to the cost function minimised by Tomasi and Kanade (1992). A noise model is presented along with criteria for existence of a solution, and experiments on real images are described.
- Affine epipolar geometry is related to the rigid motion parameters (Section 4), and Koenderink and van Doorn's novel motion representation (Koenderink and van Doorn 1991) is formalised. The scale, cyclotorsion angle and projected axis of rotation are

then computed *directly* from the epipolar geometry (i.e., using two views). The only camera calibration parameter required here is aspect ratio. A suitable error model is also derived.

- A linear Kalman filter is defined for the multiple-view case to determine optimal estimates over long sequences. Images are processed in successive pairs of frames, facilitating extension to the m-view case in a sequential (rather than batch) processing mode. Unlike some previous point-based structure and motion schemes (e.g., Charnley et al. 1988), we do not assign an individual Kalman filter to each 3D feature. This liberates us from having to track 3D points through the sequence.
- A new method of computing the remaining turn angle about the 3D axis (using three or more distinct views) is presented (Section 5), and its performance contrasted with existing methods.

This research was inspired by the work of Koenderink and van Doorn (1991), who recovered the 3D motion of an object simply by observing the local coordinate frame (LCF) they attached to it. Their entire computation relied on this LCF and was thus extremely sensitive to errors in the points defining the frame. Furthermore, their motion algorithm used only four points and was not extendable when additional points were available. This paper builds on their work in an attempt to redress these shortcomings.

2 Camera Models

A camera projects a 3D world point $\mathbf{X} = (X, Y, Z)^\top$ onto a 2D image point $\mathbf{x} = (x, y)^\top$. This section defines the affine camera and explains its relation to the projective, perspective, para-perspective, weak perspective (scaled orthographic) and orthographic cameras.

The general mapping from $\mathcal{P}^3$ to $\mathcal{P}^2$ can be written in terms of a projection matrix $\mathbf{P} = [P_{ij}]$,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}, \quad (1)$$

where (x_1, x_2, x_3) and (X_1, X_2, X_3, X_4) are homogeneous coordinates related to $\mathbf{x}$ and $\mathbf{X}$ by $\mathbf{x} = (x_1/x_3, x_2/x_3)$ and $\mathbf{X} = (X_1/X_4, X_2/X_4, X_3/X_4)$. Mundy and Zisserman (1992) termed this a *projective camera*. Equation (1) places no restriction on the coordinate systems in which $\mathbf{x}$ and $\mathbf{X}$ are measured: neither frame has to be orthogonal, and the two frames need not be aligned.

2.1 The Affine Camera

An *affine camera* is a special case of the projective camera in Eq. (1), with the constraints $P_{31} = P_{32} = P_{33} = 0$:

$$\mathbf{P}_{\text{aff}} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & P_{34} \end{bmatrix}. \quad (2)$$

It corresponds to a projective camera with its optical center on the plane at infinity; consequently, all projection rays are parallel. Since scale is arbitrary for homogeneous coordinates, only the *ratios* of the elements P_{ij} are important, so $\mathbf{P}_{\text{aff}}$ has eight degrees of freedom.

It is convenient to decompose $\mathbf{P}_{\text{aff}}$ in the same manner that Faugeras et al. (1992) and Luong et al. (1993) decomposed $\mathbf{P}$, namely:

$$\mathbf{P}_{\text{aff}} = \mathbf{C}\mathbf{P}_{\|}\mathbf{G} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ 0 & 0 & C_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ 0 & 0 & 0 & G_{44} \end{bmatrix}. \quad (3)$$

The 3×3 matrix $\mathbf{C}$ accounts for intrinsic camera parameters and represents a 2D affine transformation (hence $C_{31} = C_{32} = 0$). It encodes camera calibration and has a variable number of unknowns (usually up to five), depending on the sophistication of the camera model. We assume there is no shear in the camera axes and use four parameters,

$$\mathbf{C} = \begin{bmatrix} f\xi & 0 & o_x \\ 0 & f & o_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

where ξ is the camera aspect ratio, f the focal length and (o_x, o_y) the principal point (where the optic axis intersects the image plane). The camera is said to be "calibrated" when $\mathbf{C}$ is known. The 3×4 matrix $\mathbf{P}_{\|}$ performs the parallel projection operation, and the 4×4 matrix $\mathbf{G}$ accounts for extrinsic camera parameters, encoding the relative position and orientation between the world and camera coordinate systems. The affine camera therefore covers the composed effects of: (i) a 3D

affine transformation between world and camera coordinate systems; (ii) parallel projection onto the image plane; and (iii) a 2D affine transformation of the image.

In terms of inhomogeneous image and world coordinates, the affine camera is written

$$\mathbf{x} = \mathbf{M}\mathbf{X} + \mathbf{t}, \tag{5}$$

where $\mathbf{M} = [M_{ij}]$ is a 2×3 matrix (with elements $M_{ij} = P_{ij}/P_{34}$) and $\mathbf{t} = (P_{14}/P_{34}, P_{24}/P_{34})^\top$ is a 2-vector giving the projection of the origin of the world coordinate frame ($\mathbf{X} = \mathbf{0}$). A key property of the affine camera is that it *preserves parallelism*: lines that are parallel in the world remain parallel in the image. The proof is simple: two parallel world lines $\mathbf{X}_1(\lambda) = \mathbf{X}_a + \lambda\mathbf{U}$ and $\mathbf{X}_2(\mu) = \mathbf{X}_b + \mu\mathbf{U}$ project to the image lines $\mathbf{x}_1(\lambda) = (\mathbf{M}\mathbf{X}_a + \mathbf{t}) + \lambda\mathbf{M}\mathbf{U}$ and $\mathbf{x}_2(\mu) = (\mathbf{M}\mathbf{X}_b + \mathbf{t}) + \mu\mathbf{M}\mathbf{U}$, which are clearly parallel.

The following two subsections describe two special cases of the affine camera, viz. the weak perspective and para-perspective cameras.

2.2 *The Weak Perspective Camera*

Consider the familiar camera-centered perspective equations, where each point is scaled by its individual depth Z_i^c and all projection rays converge to the optical centre:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{Z_i^c} \begin{bmatrix} X_i^c \\ Y_i^c \end{bmatrix}.$$

Here, $\mathbf{X}^c = (X^c, Y^c, Z^c)^\top$ denotes coordinates in the camera frame. Fig. 2 illustrates the 1D case, where the "image" is a line rather than a plane, and Fig. 3(a) shows the 2D case. When the camera field of view is small and the depth variation of the object $\Delta Z_i^c = Z_i^c - Z_{\text{ave}}^c$ is small compared to the average distance of the object from the camera Z_{ave}^c (a ratio of 1:10 is typically used (Thompson and Mundy 1987)), the individual depths Z_i^c may be approximated by Z_{ave}^c, giving a *weak perspective* or *scaled orthographic* camera[1]:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \frac{f}{Z_{\text{ave}}} \begin{bmatrix} X_i^c \\ Y_i^c \end{bmatrix}.$$

Figs. 2 and 3(c) illustrate this model: points are first projected orthographically onto the average depth plane $Z^c = Z_{\text{ave}}^c$ and then projected perspectively from this fronto-parallel plane onto the image. The weak perspective camera therefore combines orthographic and perspective projection. The latter operation simply introduces a scale factor, accounting for changes in image size when an object looms towards, or recedes from, the camera.

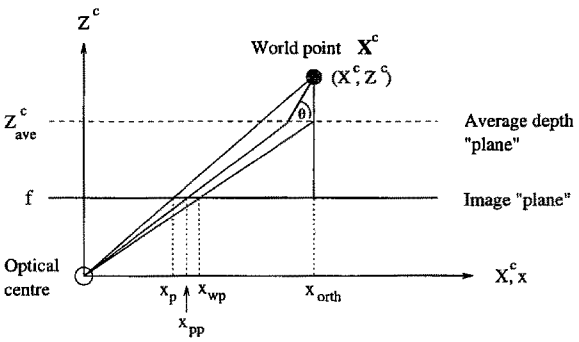

Fig. 2. One-dimensional image formation. The image is the line $Z^c = f$. For x_p (perspective), projection is along the ray connecting the world point $\mathbf{X}^c$ to the optical centre. For x_{orth} (orthographic), projection is perpendicular to the image. For x_{pp} (para-perspective), $\mathbf{X}^c$ is first projected onto the average depth plane at angle θ, and then projected perspectively onto the image plane; x_{wp} (weak perspective) is a special case of x_{pp} with $\theta = 90°$ (i.e., orthographic projection onto the average depth plane).

Coordinates measured in a world coordinate system ($\mathbf{X}$) are related to $\mathbf{X}^c$ by a (6 degree-of-freedom) *rigid* transformation

$$\mathbf{X}^c = \mathbf{R}\mathbf{X} + \mathbf{T}, \tag{6}$$

where $\mathbf{R}$ is a 3×3 rotation matrix with rows $\{\mathbf{R}_1^\top, \mathbf{R}_2^\top, \mathbf{R}_3^\top\}$, and $\mathbf{T} = (T_x, T_y, T_z)^\top$ is a translation vector representing the origin of the world frame. The depth of a point $\mathbf{X}_i$ measured along the line of sight in the camera frame is then $Z_i^c = \mathbf{R}_3^\top \mathbf{X}_i + T_z$. The centroid of the point set is denoted $\mathbf{X}_{\text{ave}}$ and the depth variation of the object is given by $\Delta Z_i^c = \mathbf{R}_3^\top(\mathbf{X}_i - \mathbf{X}_{\text{ave}})$. The weak perspective projection equations are then

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{R}_1^\top \mathbf{X} + T_x \\ \mathbf{R}_2^\top \mathbf{X} + T_y \\ Z_{\text{ave}}^c \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{R}_1^\top & T_x \\ \mathbf{R}_2^\top & T_y \\ \mathbf{0}^\top & Z_{\text{ave}}^c \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

$$\text{i.e., } \mathbf{P}_{\text{wp}} = \mathbf{C}\mathbf{P}_{\parallel} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^\top & Z_{\text{ave}}^c \end{bmatrix}. \tag{7}$$

In inhomogeneous image and world coordinates,

$$\mathbf{x} = \frac{f}{Z_{\text{ave}}^c} \begin{bmatrix} \xi\mathbf{R}_1^\top \\ \mathbf{R}_2^\top \end{bmatrix} \mathbf{X} + \frac{f}{Z_{\text{ave}}^c} \begin{bmatrix} \xi T_x \\ T_y \end{bmatrix} + \begin{bmatrix} o_x \\ o_y \end{bmatrix}$$

$$= \mathbf{M}_{\text{wp}}\mathbf{X} + \mathbf{t}_{\text{wp}}, \tag{8}$$

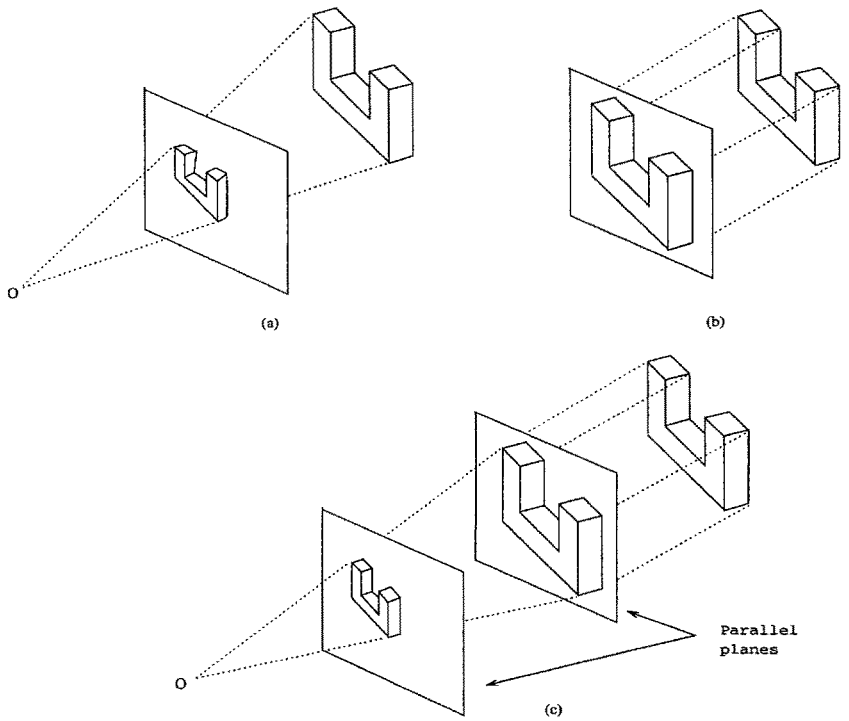

Fig. 3. Camera models: (a) perspective (all rays pass through a single projection point O, and the intersection of the ray star with the image plane generates the image); (b) orthographic (all rays are parallel, with the optical centre O at infinity); (c) weak perspective (combined orthographic and perspective projection). For (b) and (c), parallel lines in the scene remain parallel in the image; this isn't true for (a).

where $\mathbf{M}_{\text{wp}}$ is a 2×3 matrix whose rows are the uniformly scaled rows of a rotation matrix, and $\mathbf{t}_{\text{wp}}$ is a 2-vector.

2.3 *The Para-Perspective Camera*

In the weak perspective case, projection of the scene point onto the average depth plane occurs parallel to the optic axis. The *para-perspective* camera (Aloimonos 1986; Aloimonos 1992) generalises this by projecting parallel to an *arbitrary* (but fixed and known) projection direction. Since the average depth plane remains parallel to the image plane, the perspective projection stage simply introduces a scale factor (as in the weak perspective model). The 1D case takes the form

$$x_{\text{pp}} = \frac{f}{Z^c_{\text{ave}}}(X^c - \Delta Z^c \cot\theta),$$

where θ denotes the angle between the projection direction and the positive X-axis (see Fig. 2). In the 2D case, the projection direction is described by two angles (θ_x, θ_y), where θ_x lies in the X-Z plane (θ in Fig. 2) and θ_y is the equivalent angle in the Y-Z plane. Factoring in camera calibration parameters and the rigid transformation between the camera and world coordinate frames gives

$$\mathbf{P}_{\text{pp}} = \mathbf{C}\begin{bmatrix} 1 & 0 & -\cot\theta_x & \cot\theta_x \\ 0 & 1 & -\cot\theta_y & \cot\theta_y \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^\top & Z^c_{\text{ave}} \end{bmatrix}, \qquad (9)$$

a more general form of the expressions in Aloimonos (1992).

2.4 *Discussion*

The affine camera in Eq. (2) clearly generalises the scaled orthographic and para-perspective models of Eqs. (7) and (9). This generalisation takes two forms. First, *non-rigid* deformation of the object is permitted by the 3D affine transformation. Second, calibration is unnecessary, for the relations amongst the matrix elements that arise in Eqs. (7) and (9) from $\mathbf{C}$ are not applied in the affine camera. The hierarchy of models is summarised in Fig. 4.

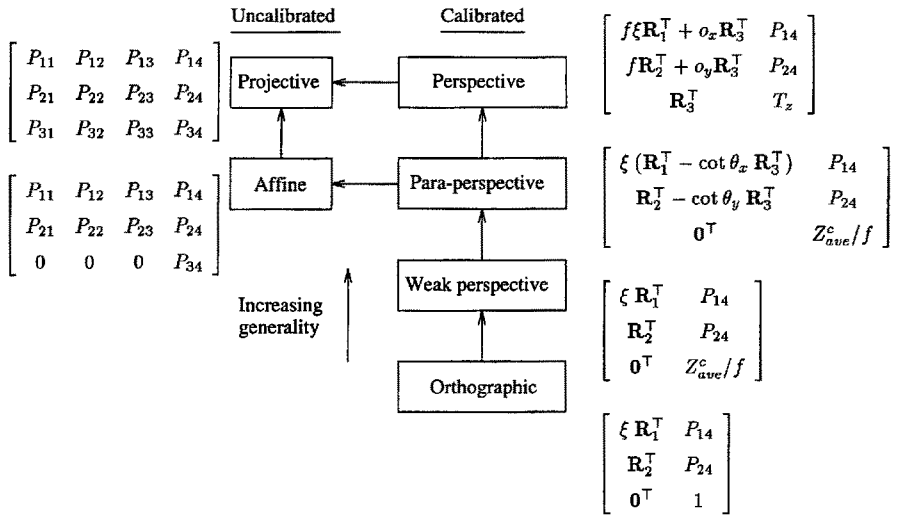

Fig. 4. Hierarchy of camera models and their **P** matrices. The projective camera is the most general and the orthographic camera the least general.

3 Affine Epipolar Geometry

The concept of an epipolar line is familiar in the stereo and motion literature (Arnold and Binford 1980; Barnard and Thompson 1980; Longuet-Higgins 1981; Faugeras et al. 1992; Hartley 1992; Pollard 1985). Figure 5 shows the classical construction for a perspective stereo-pair. The epipolar constraint relates a point in one image to a line in the other image depending on the intrinsic and extrinsic camera parameters. Thus, given $\mathbf{x}_i$ in I, the corresponding point $\mathbf{x}'_i$ in I' is constrained to lie on an *epipolar line* in I', namely the intersection between the epipolar plane and I' (the image in I' of the ray through $\mathbf{x}_i$).

We first consider the geometry of two affine cameras, a situation arising due to stereo viewing or relative motion between camera and scene. We then derive the affine epipolar equations and describe a method to solve them.

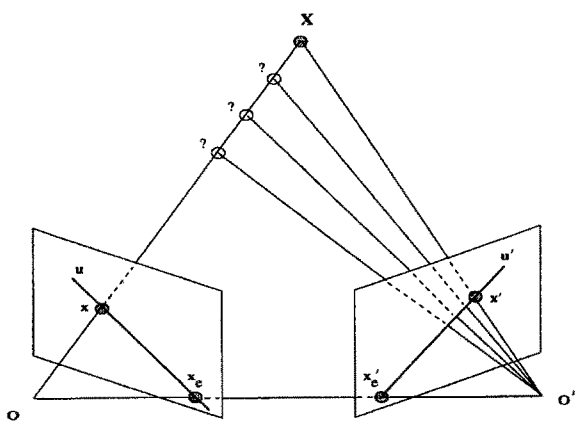

Fig. 5. Perspective epipolar geometry. The baseline connecting optical centres **O** and **O'** intersects the image planes at the epipoles $\mathbf{x}_e$ and $\mathbf{x}'_e$. A point **X**, together with the baseline, defines an epipolar plane, which cuts the image planes in the epipolar lines **u** and **u'**. The point **x'** corresponding to **x** must lie on the epipolar line **u'**. All epipolar lines pass through the epipole.

3.1 *Scene Transformations and Relative Coordinates*

An important property of the affine camera is that it retains its form when the scene undergoes a 3D affine transformation. Consider a world point $\mathbf{X}_i$ projected by an affine camera $\{\mathbf{M}, \mathbf{t}\}$ to the image point $\mathbf{x}_i$, i.e.

$$\mathbf{x}_i = \mathbf{M}\mathbf{X}_i + \mathbf{t}. \tag{10}$$

Let the scene (or camera) then move according to

$$\mathbf{X}'_i = \mathbf{A}\mathbf{X}_i + \mathbf{T}, \tag{11}$$

where $\mathbf{X}'_i$ is the new world position, **A** is a 3×3 matrix and **T** is a 3-vector. This *scene transformation* encodes relative motion between the camera and the world as a 3D affine transformation (12 degrees of freedom, not necessarily a rigid motion). The new world point $\mathbf{X}'_i$ then projects to $\mathbf{x}'_i = (x'_i, y'_i)^\top$, where

$$\begin{aligned}\mathbf{x}'_i &= \mathbf{M}\mathbf{X}'_i + \mathbf{t} = \mathbf{M}(\mathbf{A}\mathbf{X}_i + \mathbf{T}) + \mathbf{t} \\ &= \mathbf{M}\mathbf{A}\mathbf{X}_i + (\mathbf{M}\mathbf{D} + \mathbf{t}) = \mathbf{M}'\mathbf{X}_i + \mathbf{t}'.\end{aligned} \tag{12}$$

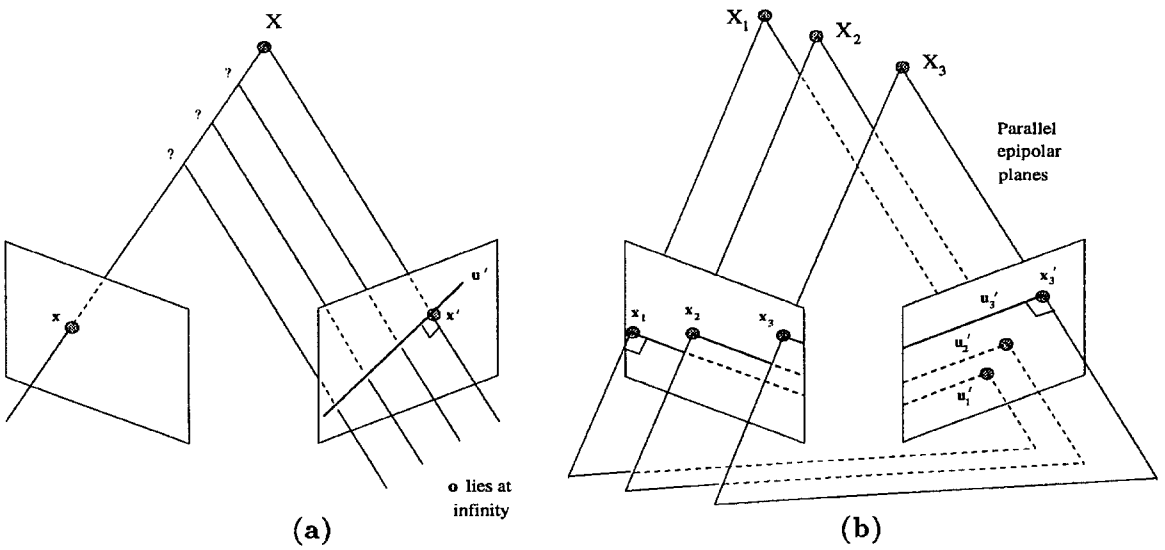

Fig. 6. Affine epipolar geometry: (a) all projection rays are parallel to one another (the optical centre lies at infinity) and perpendicular to the image plane. $\mathbf{X}$ lies along the ray from the optical centre through $\mathbf{x}$, and the projection of this ray into I' gives the epipolar line $\mathbf{u}'$, along which $\mathbf{x}'$ must lie; (b) all epipolar planes (and hence lines) are parallel, and their normal is parallel to the line of intersection of the two image planes.

This can be interpreted as a second affine camera $\{\mathbf{M}', \mathbf{t}'\}$ observing the original scene, with $\{\mathbf{M}', \mathbf{t}'\}$ accounting for changes in both the extrinsic *and* intrinsic parameters (i.e., pose *and* calibration). The form of the affine camera is thus preserved.

A second important property of the affine camera model is that *relative coordinates* cancel out translation effects, and this will be used frequently in subsequent computations. If $\mathbf{X}_0$ is designated a reference point (or *origin*), then vector differencing in the scene gives

$$\Delta\mathbf{X} = \mathbf{X} - \mathbf{X}_0 \quad \text{and} \quad \Delta\mathbf{X}' = \mathbf{X}' - \mathbf{X}'_0 = \mathbf{A}\Delta\mathbf{X},$$

which are clearly independent of $\mathbf{T}$. More importantly, in the *image*, registering the points gives

$$\begin{aligned} \Delta\mathbf{x} &= \mathbf{x} - \mathbf{x}_0 = \mathbf{M}\Delta\mathbf{X} \quad \text{and} \\ \Delta\mathbf{x}' &= \mathbf{x}' - \mathbf{x}'_0 = \mathbf{M}'\Delta\mathbf{X} = \mathbf{M}\mathbf{A}\Delta\mathbf{X}, \end{aligned} \tag{13}$$

which are again independent of $\mathbf{T}$, $\mathbf{t}$ and $\mathbf{t}'$. This cancellation relies crucially on linearity and is not possible in general under perspective projection.

3.2 *The Affine Epipolar Line*

Fig. 6 shows the epipolar geometry for two *affine* cameras. All projection rays are parallel (the optical centre lies at infinity), and because the affine camera preserves parallelism, the epipolar lines are also parallel, i.e., the epipoles are situated at infinity in the image planes.

The equation of an affine epipolar line is obtained by partitioning $\mathbf{M}$ as $(\mathbf{B} \mid \mathbf{b})$, where $\mathbf{B}$ is a (non-singular) 2×2 matrix and $\mathbf{b}$ a 2×1 vector. From Eq. (5),

$$\mathbf{x}_i = \mathbf{B}\begin{bmatrix} X_i \\ Y_i \end{bmatrix} + Z_i\mathbf{b} + \mathbf{t}. \tag{14}$$

Similarly, partitioning $\mathbf{M}'$ into $(\mathbf{B}' \mid \mathbf{b}')$, Eq. (12) gives

$$\mathbf{x}'_i = \mathbf{B}'\begin{bmatrix} X_i \\ Y_i \end{bmatrix} + Z_i\mathbf{b}' + \mathbf{t}'. \tag{15}$$

Eliminating the world coordinates $(X_i, Y_i)^\top$ between these two equations yields the desired relation,

$$\boxed{\mathbf{x}'_i = \Gamma\mathbf{x}_i + Z_i\mathbf{d} + \varepsilon} \tag{16}$$

with $\Gamma = \mathbf{B}'\mathbf{B}^{-1}$, $\mathbf{d} = \mathbf{b}' - \Gamma\mathbf{b}$ and $\varepsilon = \mathbf{t}' - \Gamma t$. Quantities Γ, $\mathbf{d}$ and ε depend only on the cameras $\{\mathbf{M}, \mathbf{t}, \mathbf{M}', \mathbf{t}'\}$—not on the scene structure. Eq. (16) shows that the point $\mathbf{x}'_i$ associated with $\mathbf{x}_i$ lies on a line in the second image with offset $\Gamma\mathbf{x}_i + \varepsilon$ and direction $\hat{\mathbf{d}} = \mathbf{d}/|\mathbf{d}|$ (Fig. 7). The unknown depth Z_i determines how far along this line $\mathbf{x}'_i$ lies. The epipolar lines are clearly parallel ($\hat{\mathbf{d}}$ is constant) and have different offsets (depending on $\mathbf{x}_i$).

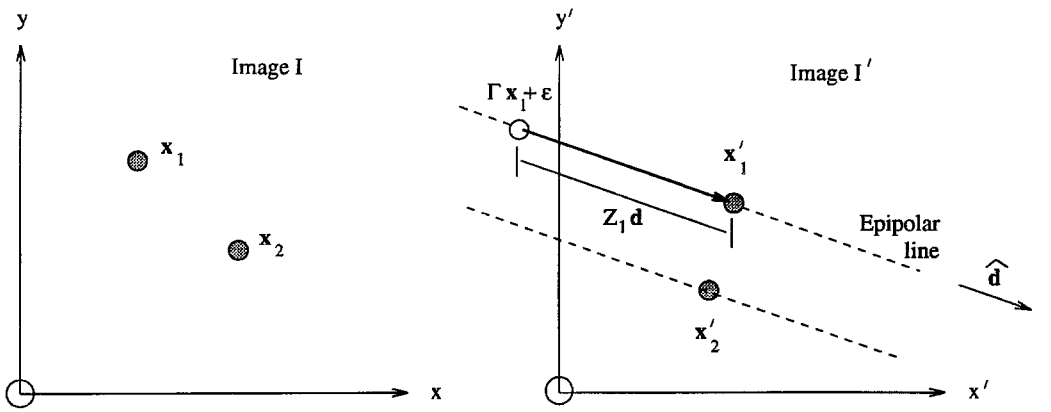

Fig. 7. The affine epipolar line: $\mathbf{x}'_i = \mathbf{\Gamma}\mathbf{x}_i + \varepsilon + Z_i\mathbf{d}$.

Inverting Eq. (16) yields the epipolar line in the first image for $\mathbf{x}'_i$, with offset $\mathbf{\Gamma}^{-1}(\mathbf{x}'_i - \varepsilon)$ and direction parallel to $\mathbf{\Gamma}^{-1}\mathbf{d}$:

$$\boxed{\mathbf{x}_i = \mathbf{\Gamma}^{-1}(\mathbf{x}'_i - \varepsilon) - Z_i\mathbf{\Gamma}^{-1}\mathbf{d}} \tag{17}$$

Finally, taking difference vectors $\mathbf{\Delta x}_i = \mathbf{x}_i - \mathbf{x}_0$ and $\mathbf{\Delta x}'_i = \mathbf{x}'_i - \mathbf{x}'_0$ gives translation-invariant versions of these formulae[2], eliminating ε (and hence $\mathbf{T}$, $\mathbf{t}$ and $\mathbf{t}'$):

$$\mathbf{\Delta x}'_i = \mathbf{B}'\begin{bmatrix}\Delta X_i \\ \Delta Y_i\end{bmatrix} + \Delta Z_i\mathbf{b}' = \mathbf{\Gamma}\mathbf{\Delta x}_i + \Delta Z_i\mathbf{d} \tag{18}$$

$$\mathbf{\Delta x}_i = \mathbf{B}\begin{bmatrix}\Delta X_i \\ \Delta Y_i\end{bmatrix} + \Delta Z_i\mathbf{b} = \mathbf{\Gamma}^{-1}\mathbf{\Delta x}'_i - \Delta Z_i\mathbf{\Gamma}^{-1}\mathbf{d}. \tag{19}$$

3.3 *The Affine Fundamental Matrix*

Eq. (16) defined the epipolar line in parametric form; an implicit form is obtained by eliminating the depths Z_i, giving a single equation in the *image measurables*,

$$(\mathbf{x}'_i - \mathbf{\Gamma}\mathbf{x}_i - \varepsilon)^\top\mathbf{d}^\perp = 0, \tag{20}$$

where $\mathbf{d}^\perp$ is the perpendicular[3] to $\mathbf{d}$. Equation (20) can also be written

$$\boxed{ax'_i + by'_i + cx_i + dy_i + e = 0} \tag{21}$$

with $(a, b)^\top = \mathbf{d}^\perp$, $(c, d)^\top = -\mathbf{\Gamma}^\top\mathbf{d}^\perp$ and $e = -\varepsilon^\top\mathbf{d}^\perp$. This *affine epipolar constraint equation* (Zisserman 1992) is a linear equation in the unknown constants $a \cdots e$, which depend only on the camera parameters and the relative motion. Evidently, only the *ratios* of the five parameters $a \cdots e$ can be computed, so Eq. (21) has *four* independent degrees of freedom. Solving this equation does not require a calibrated camera, since an affine model has been used throughout. The difference vector form is

$$a\Delta x'_i + b\Delta y'_i + c\Delta x_i + d\Delta y_i = 0. \tag{22}$$

Eq. (21) may also be expressed in the form of a *fundamental matrix* $\mathbf{F}_A$,

$$\mathbf{p}'^\top\mathbf{F}_A\mathbf{p} = [x'_i \quad y'_i \quad 1]\begin{bmatrix}0 & 0 & a \\ 0 & 0 & b \\ c & d & e\end{bmatrix}\begin{bmatrix}x_i \\ y_i \\ 1\end{bmatrix} = 0, \tag{23}$$

where $\mathbf{p}' = (x', y', 1)^\top$ and $\mathbf{p} = (x, y, 1)^\top$ are homogeneous 3-vectors representing points in the image plane. The matrix $\mathbf{F}_A$ has maximum rank two. The epipolar lines corresponding to $\mathbf{p}$ and $\mathbf{p}'$ are $\mathbf{u}' = \mathbf{F}_A\mathbf{p}$ and $\mathbf{u} = \mathbf{F}_A^\top\mathbf{p}'$ respectively, where $\mathbf{u} = (u_1, u_2, u_3)^\top$ represents the line $u_1x + u_2y + u_3 = 0$ (and similarly for $\mathbf{u}'$). The *normal* to $\mathbf{u}$ thus lies in the direction $(u_1, u_2) = (c, d)$, while for $\mathbf{u}'$, the normal is $(u'_1, u'_2) = (a, b)$. An example with synthetic data is shown in Fig. 8.

The form of $\mathbf{F}_A$ in Eq. (23) is a special case of the general 3×3 fundamental matrix $\mathbf{F}$ used in stereo and motion algorithms. Faugeras et al. (1992) and Luong et al. (1993) showed that the particular parameterisation of $\mathbf{F}$ was of some importance. This is not the case for the affine fundamental matrix $\mathbf{F}_A$, since unlike $\mathbf{F}$, the number of independent degrees of freedom in $\mathbf{F}_A$ are encoded exactly by the number of non-zero matrix elements: there are five parameters a, b, c, d and e, and four independent degrees of freedom plus an arbitrary scale factor (since it is a homogeneous matrix). The

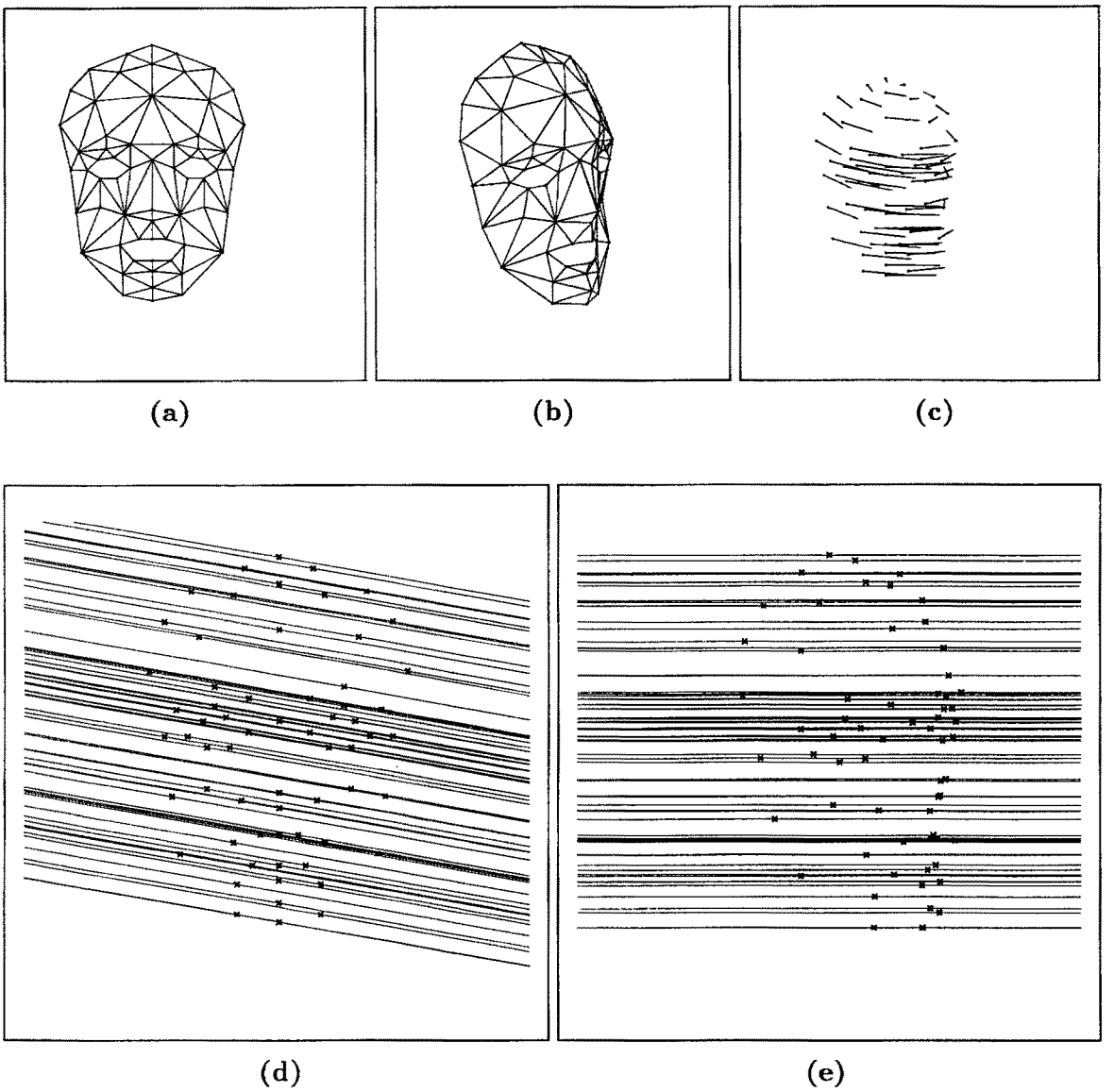

Fig. 8. Epipolar lines for noise-free synthetic data with $\mathbf{n} = (0.00, 0.70, -0.13, -0.71)$: (a), (b) images I and I' of a wire-frame face; (c) motion vectors; (d), (e) epipolar lines for (a) and (b). Note that the epipolar lines are parallel whereas the motion vectors are not.

scale factor is deliberately not removed, say, by fixing one of the parameters (e.g., $e = 1$) since any parameter could validly be zero.

Note that the general fundamental matrix $\mathbf{F}$ arising from two *projective* cameras with *parallel epipolar lines* will not necessarily be of the affine form (see Appendix A); special case projective camera geometries have parallel epipolar lines without satisfying the requirements for validity of the affine camera. For instance, a perspective camera translating parallel to the image plane will have parallel epipolar lines, regardless of the depth of field and field of view of the scene being observed (Torr 1993).

3.4 *Solving the Affine Epipolar Equation*

Eq. (21) is defined up to a scale factor, so only four point correspondences are needed to solve for the four independent degrees of freedom (conditions for *existence* of a solution are discussed later). When n correspondences are available ($n > 4$), it is advantageous to use

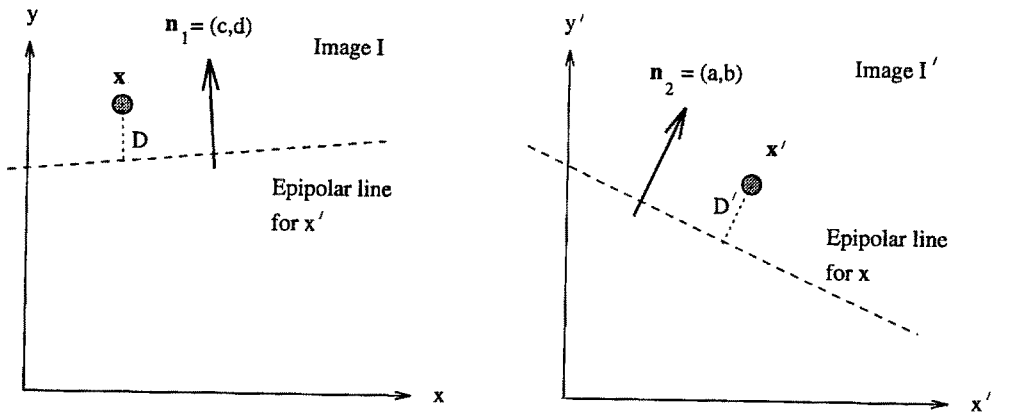

Fig. 9. The normals to the epipolar lines are $\mathbf{n}_1$ and $\mathbf{n}_2$. Noise displaces a point $\mathbf{x}'$ in I' from the epipolar line associated with its counterpart $\mathbf{x}$ by perpendicular distance D'. A similar displacement by D occurs in I.

all the points, since this improves the accuracy of the solution, allows detection of (and hence provides immunity to) outliers, and obviates the need to select a minimal point set.

However, the presence of "noise" in the over determined system means that points won't lie exactly on their epipolar lines (Fig. 9), and an appropriate minimisation is required. Optimal estimation requires knowledge of the image noise distribution, which depends on the specific camera, its lens and the image processing performed. It may also depend on the camera aspect ratio, though with suitable scaling, on no other calibration parameter. When necessary, we assume image noise to be Gaussian and isotropic[4] and return to the validity of this assumption later.

The following sections discuss three minimum variance cost functions involving the epipolar parameters, and differing in the image distances minimised. These functions are assessed in terms of accuracy and complexity.

3.4.1 Notation. The two image coordinates of a corresponding point are combined into a vector $\mathbf{r}_i = (x_i', y_i', x_i, y_i)^\top$. The centroid of these 4-vectors is $\bar{\mathbf{r}}$ and the centred points are denoted $\mathbf{v}_i = \mathbf{r}_i - \bar{\mathbf{r}}$. The 4D normal vector is defined as $\mathbf{n} = (a, b, c, d)^\top$, with $\mathbf{n}_1 = (c, d)^\top$ and $\mathbf{n}_2 = (a, b)^\top$ being the 2D normal vectors to the epipolars in I and I' respectively. The perpendicular distance D_i' between $\mathbf{x}_i'$ and its associated epipolar line in I' is

$$D_i' = \frac{ax_i' + by_i' + cx_i + dy_i + e}{\sqrt{a^2 + b^2}}$$

and the counterpart distance D_i in I is

$$D_i = \frac{ax_i' + by_i' + cx_i + dy_i + e}{\sqrt{c^2 + d^2}},$$

as in Fig. 9. The real, symmetric scatter matrix $\mathbf{W}$ is partitioned into four 2×2 matrices as follows:

$$\mathbf{W} = \sum_{i=0}^{n-1} \mathbf{v}_i \mathbf{v}_i^\top = \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \\ \mathbf{K}_2^\top & \mathbf{K}_3 \end{bmatrix}$$
$$= \begin{bmatrix} \sum_i (\mathbf{x}_i' - \bar{\mathbf{x}}')(\mathbf{x}_i' - \bar{\mathbf{x}}')^\top & \sum_i (\mathbf{x}_i' - \bar{\mathbf{x}}')(\mathbf{x}_i - \bar{\mathbf{x}})^\top \\ \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i' - \bar{\mathbf{x}}')^\top & \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \end{bmatrix}.$$

3.4.2 Cost Functions. We consider the following three cost functions:

$$E_1(\mathbf{n}, e) = \left(\frac{1}{a^2 + b^2} + \frac{1}{c^2 + d^2}\right) \times \sum_{i=0}^{n-1} (ax_i' + by_i' + cx_i + dy_i + e)^2 \quad (24)$$

$$E_2(\mathbf{n}, e) = \frac{1}{a^2 + b^2} \times \sum_{i=0}^{n-1} (ax_i' + by_i' + cx_i + dy_i + e)^2 \quad (25)$$

$$E_3(\mathbf{n}, e) = \frac{1}{a^2 + b^2 + c^2 + d^2} \times \sum_{i=0}^{n-1} (ax_i' + by_i' + cx_i + dy_i + e)^2 \quad (26)$$

All three functions minimise the sum of squares of a perpendicular distance measure, all are *scale-invariant*

(i.e., if $\{\mathbf{n}, e\}$ is a solution, then so is $\{k\mathbf{n}, ke\}$ where k is a non-zero scalar), and all can be minimised over e directly, giving $e = -\mathbf{n}^\top \bar{\mathbf{r}}$ (see Appendix B).

1. *Cost function* E_1. E_1 sums the squared perpendicular image distances over I and I': $E_1 = \sum_{i=0}^{n-1} D_i^2 + (D_i')^2$. The solution satisfies a system of non-linear simultaneous equations (see Appendix in Shapiro (1993)):

$$\left(1 + \frac{1}{s^2}\right)\begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \\ s^4\mathbf{K}_2^\top & s^4\mathbf{K}_3 \end{bmatrix}\begin{bmatrix} \mathbf{n}_2 \\ \mathbf{n}_1 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{n}_2 \\ \mathbf{n}_1 \end{bmatrix}, \quad |\mathbf{n}_2|^2 = 1,$$

where $s = |\mathbf{n}_1|/|\mathbf{n}_2|$ is a scale factor[5] and λ a Lagrange multiplier. The solution requires non-linear minimisation by numerical methods, the minimum cost being $E_{1,\min} = \lambda(1 + 1/s^2)$.

2. *Cost function* E_2. E_2 sums the squared perpendicular distances in a single image (I'): $E_2 = \sum_i (D_i')^2$. The normal $\mathbf{n}_2$ satisfies a 2D eigenvector equation (see Appendix in Shapiro (1993)):

$$\left(\mathbf{K}_1 - \mathbf{K}_2\mathbf{K}_3^{-1}\mathbf{K}_2^\top\right)\mathbf{n}_2 = \lambda_1\mathbf{n}_2, \quad |\mathbf{n}_2|^2 = 1,$$

and is the eigenvector associated with the minimum eigenvalue λ_1. Then $\mathbf{n}_1 = -\mathbf{K}_3^{-1}\mathbf{K}_2^\top\mathbf{n}_2$, giving minimum cost $E_{2,\min} = \lambda_1$. A similar algorithm can be obtained by minimising the distances in I, namely $\sum_i D_i^2$.

3. *Cost function* E_3. Equation (21) can be written as $\mathbf{r}_i^\top\mathbf{n} + e = 0$, where $\mathbf{n}$ is the normal to a hyperplane in 4D and $(\mathbf{r}_i^\top\mathbf{n} + e)/|\mathbf{n}|$ is the 4D perpendicular distance from $\mathbf{r}_i$ to this hyperplane. E_3 sums these squared 4D perpendicular distances,

$$E_3 = \sum_{i=0}^{n-1} \frac{\left(\mathbf{r}_i^\top\mathbf{n} + e\right)^2}{|\mathbf{n}|^2} = \sum_{i=0}^{n-1} \frac{\left(\mathbf{v}_i^\top\mathbf{n}\right)^2}{|\mathbf{n}|^2} = \frac{\mathbf{n}^\top\mathbf{W}\mathbf{n}}{|\mathbf{n}|^2}. \tag{27}$$

This is the classic linear least squares problem. The solution employs *orthogonal regression* and satisfies the eigenvector equation (see Appendix B)

$$\mathbf{W}\mathbf{n} = \lambda_1\mathbf{n}, \quad |\mathbf{n}|^2 = 1$$

where $\mathbf{n}$ is the unit eigenvector corresponding to the minimum eigenvalue λ_1. The minimum cost is $E_{3,\min} = \lambda_1$.We show below that minimising E_3 is equivalent to minimising point-to-point image distances.

3.4.3 Discussion and Previous Work. The three above-mentioned functions all involve image distances; this is important since the observations are made in the image and the system noise originates there (Harris 1990). Various arguments can then be advanced in favour of the different cost functions; we show that E_3 is superior to E_1 and E_2 in several respects.

Faugeras et al. (1992) and Luong et al. (1993) evaluated candidate cost functions for computing the fundamental matrix $\mathbf{F}$ of a *projective* camera; E_1 is the affine analogue of their favoured non-linear criterion (using distances to epipolar lines[6]) and E_3 is the analogue of their linear criterion (using the eigenvector method). They criticised the latter approach for failing to impose the rank constraint[7] on $\mathbf{F}$ and for introducing bias into the computation by shifting the epipole towards the image centre. In the *affine* case, however, $\mathbf{F}_A$ is *guaranteed* to have a maximum rank of two (cf. Eq. (23)) and the epipole lies at infinity, removing these two objections against the linear method. Moreover, E_3 is the affine analogue of the non-linear normalised gradient criterion introduced in Luong et al. (1993), with the gradient here simply being $\mathbf{n}$ (which, unlike the projective case, is constant over all the points).

Furthermore, although E_3 may be interpreted as a 4D algebraic distance measure, it is actually equivalent to an *image* distance measure based on point-to-point (rather than point-to-line) distances. It measures the distance between the observed image location and the location predicted by projecting the computed affine structure $\mathbf{X}_i$ onto an image using the computed affine cameras $\{\mathbf{M}, \mathbf{t}, \mathbf{M}', \mathbf{t}'\}$ (see Fig. 10), i.e.

$$\begin{aligned} E_{\mathrm{TK}} &= \min_{\{\mathbf{M},\mathbf{M}',\mathbf{X},\mathbf{t},\mathbf{t}'\}} \sum_{i=0}^{n-1} |\mathbf{x}_i - \mathbf{M}\mathbf{X}_i - \mathbf{t}|^2 \\ &\quad + \sum_{i=0}^{n-1} |\mathbf{x}_i' - \mathbf{M}'\mathbf{X}_i - \mathbf{t}'|^2 \\ &= \min_{\{\mathbf{n}\}} \sum_{i=0}^{n-1} \frac{\left(\mathbf{v}_i^\top\mathbf{n}\right)^2}{|\mathbf{n}|^2} = E_3 \end{aligned} \tag{28}$$

Reid and Murray (1994) showed Eq. (28) to be the cost function minimised by Tomasi and Kanade (1992), hence the TK subscript. This cost function is a sensible one to minimise because: (i) it involves the exact number of degrees of freedom of the system ($\mathbf{t}$, $\mathbf{t}'$, $\mathbf{M}$, $\mathbf{M}'$ and $\mathbf{X}_i$); and (ii) it minimises the noise residual that originates in the image plane (where the observations arise). A similar approach was adopted by Weng et al. (1993). It is shown in Appendix C that after

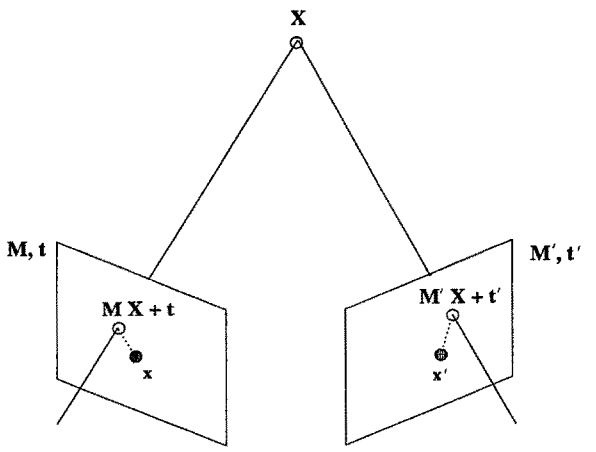

Fig. 10. The minimisation expression E_{TK}, summing the image distance between the computed point (open circle) and the observed point (solid circle).

differentiating E_{TK} with respect to $\mathbf{t}$, $\mathbf{t}'$ and $\mathbf{X}_i$ and resubstituting, E_3 obtains. That is, E_3 is simply E_{TK} with the structure-dependent terms factored out optimally.

Finally, when $s = 1$ (so $|\mathbf{n}_1| = |\mathbf{n}_2|$), the solutions returned by *all three methods* coincide[8]. Thus, when the scale change is small, the stationary point $\mathbf{n} = (a, b, c, d)^\top$ of E_3 is an excellent approximation to that of E_1. The unity scale condition occurs frequently with an affine camera, which itself is only valid over small scale changes (otherwise perspective effects become significant). When the scale change is large, E_3 provides a good initial estimate for E_1. In short, for an affine camera, the linear method (E_3) not only minimises a meaningful quantity, but also generally performs as well as its non-linear counterpart (E_1).

Like E_3, E_2 also arises from a point-based minimisation (see Shapiro 1993), namely

$$E_H = \sum_{i=0}^{n-1} |\mathbf{x}_i' - \mathbf{\Gamma}\mathbf{x}_i - Z_i\mathbf{d} - \varepsilon|^2, \qquad (29)$$

which minimises the distance in I' between the observed location and the location predicted by the computed motion parameters $\{\mathbf{\Gamma}, \mathbf{d}\}$ and affine coordinates Z_i (cf. Eq. (16)). This is the affine version of the expression minimised by Harris (1990), and has several drawbacks. First, minimising the noise in only *one* image leaves the errors unevenly distributed between I and I': a set of epipolars which fits one image well may not do likewise in the other image, causing discrepancies in the epipolar geometry (Faugeras et al. 1992; Luong et al. 1993). Second, Shapiro (1993) showed that an obscure constraint is imposed implicitly on I, namely $\sum_i D_i \Delta x_i = \mathbf{0}$ (the distribution of points about the centroid, weighted by the distances from their epipolar lines, must sum to zero). Finally, unlike E_{TK}, there is no straightforward extension of E_H to more than two images. The E_2 method is therefore unattractive.

3.4.4 Cost Function Results. The cost functions are compared using simulated data. A set of 63 scene points (wire frame triangle vertices) are projected into two images of size 256 × 256 pixels, and the images are then perturbed by Gaussian noise with variance σ^2. Two forms of E_2 are examined (E_{2a} uses I distances and E_{2b} uses I' distances), and a Newton-Raphson method is employed to minimise E_1, starting from the E_3 solution.

First consider the example in Fig. 8, which has near unity scale change between views ($s = 1.036$). The epipolar geometry parameters are computed using each of the four cost functions, and Table 1 shows the values of the cost functions for these four solutions[9] (with $\sigma = 1$ and $\sigma = 2$ pixels). Evidently, the costs are very similar with the four different solutions (within 1% of one another). The distances in I' (E_{2b}) are slightly larger than in I (E_{2a}) since $s > 1$ (the theoretical ratio of the distances is s^2).

A more useful way to compare the results is to examine the actual solutions (Table 2), i.e., the epipolar parameters (a, b, c, d). To aid understanding, we interpret these parameters in terms of rigid motion parameters, namely s (the scale factor), ϕ (the orientation of projection of the axis of rotation) and θ (the cyclotorsion angle between views). These parameters and their derivation are explained in Section 4.3. The solutions are clearly very similar, as expected from the similarity of the cost function values above. However, the difficulty with E_{2a} and E_{2b} becomes apparent, since by minimising the noise in only a single image, they describe the extreme values of the parameter range. E_1 and E_3 give more balanced solutions, and are in fact very similar over a wide range of noise values ($\sigma = 1$ and $\sigma = 2$ are shown). Notice that the variation in s and θ is smaller than that in ϕ, indicating that the ϕ parameter is more sensitive to noise (a fact borne out by the noise models of Section 4.5).

A second example, with a larger scale change ($s = 3.015$), is given in Fig. 11 and Table 3. The minimum values of the cost functions are shown to illustrate the range of values (e.g., there are much larger errors in I' than in I due to the large scale). Importantly, the E_3

Table 1. A comparison of the four cost functions for the example of Fig. 8 with Gaussian noise: (a) $\sigma = 1$ pixel; (b) $\sigma = 2$ pixels. The four columns show the cost functions evaluated with the solution obtained by minimising the cost function listed for that row. For instance, E_1 was minimised for the first row, and the obtained (a, b, c, d, e) values were then substituted into E_1, E_{2a}, E_{2b} and E_3 to give the costs shown in the appropriate columns.

Minimisation criterion	Evaluated cost				Evaluated cost			
	E_1	E_{2a}	E_{2b}	E_3	E_1	E_{2a}	E_{2b}	E_3
	(a)				(b)			
E_1	429.963	206.958	223.005	107.341	1719.423	826.195	893.228	429.202
E_{2a}	430.145	206.870	223.275	107.380	1722.391	824.770	897.622	429.827
E_{2b}	430.121	207.197	222.923	107.387	1722.022	830.128	891.894	429.952
E_3	429.964	206.946	223.018	107.341	1719.440	825.987	893.452	429.198

Table 2. A comparison of the cost functions for Fig. 8. The columns give the computed solution $\mathbf{n}$ for each cost function and their rigid motion interpretations in terms of s, θ and ϕ: (a) $\sigma = 1$ pixel; (b) $\sigma = 2$ pixels.

Function	Computed $\mathbf{n} = (a, b, c, d)$	ϕ	θ	s
	(a)			
Fiducial	(0.000, 0.695, −0.125, −0.709)	90.00	10.00	1.036
E_1	(0.004, 0.694, −0.126, −0.709)	89.68	9.75	1.038
E_{2a}	(0.003, 0.693, −0.125, −0.709)	89.61	9.72	1.037
E_{2b}	(0.005, 0.694, −0.126, −0.709)	89.76	9.77	1.039
E_3	(0.004, 0.694, −0.126, −0.709)	89.69	9.75	1.038
	(b)			
E_1	(0.009, 0.693, −0.127, −0.709)	89.27	9.46	1.040
E_{2a}	(0.005, 0.692, −0.125, −0.711)	89.61	9.57	1.043
E_{2b}	(0.013, 0.694, −0.130, −0.708)	88.94	9.35	1.037
E_3	(0.008, 0.693, −0.127, −0.710)	89.30	9.47	1.040

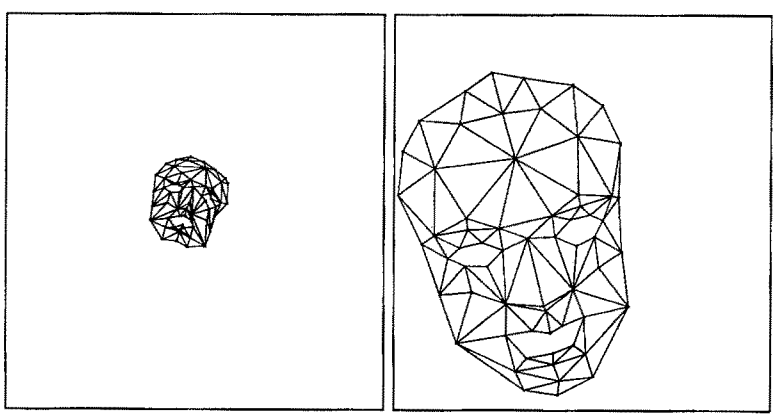

Fig. 11. Images I and I' for a large scale change ($s = 3.015$).

solution is still similar to that obtained with E_1, despite the large value of s.

In summary, the four cost functions give similar solutions for the affine epipolar geometry. In-depth analysis reveals that E_1 and E_3 are most alike, and that E_2 is unsuitable. We adopt E_3 since it is both simpler to compute than E_1 (having a closed-form solution) and has a sounder theoretical justification.

Table 3. The example from Fig. 11: (a) $\sigma = 1$ pixel; (b) $\sigma = 2$ pixels.

Function	Value	Computed $\mathbf{n} = (a, b, c, d)$	ϕ	θ	s
		(a)			
Fiducial	—	(0.310, 0.055, −0.892, 0.325)	10.00	30.00	3.015
E_1	794.342	(0.311, 0.044, −0.882, 0.352)	8.04	29.80	3.023
E_{2a}	77.890	(0.309, 0.044, −0.883, 0.350)	8.12	29.77	3.042
E_{2b}	715.963	(0.311, 0.044, −0.882, 0.352)	8.04	29.81	3.021
E_3	70.291	(0.309, 0.044, −0.883, 0.351)	8.11	29.77	3.040
		(b)			
E_1	3169.137	(0.315, 0.033, −0.868, 0.382)	5.97	29.71	2.997
E_{2a}	310.419	(0.308, 0.033, −0.873, 0.378)	6.17	29.56	3.073
E_{2b}	2850.863	(0.316, 0.033, −0.868, 0.382)	5.95	29.73	2.989
E_3	280.622	(0.308, 0.033, −0.872, 0.378)	6.15	29.58	3.065

3.4.5 Existence of Solutions and Noise Models. This section investigates the conditions for existence of a solution $\{\mathbf{n}, e\}$ and provides a noise model for the solution obtained by minimising E_3. Let $\mathbf{V}$ be the $4 \times n$ matrix $[\mathbf{v}_0 \mid \mathbf{v}_1 \mid \ldots \mid \mathbf{v}_{n-1}]$. The equations $\mathbf{v}_i^\top \mathbf{n} = 0$ (from Eq. (22)) then give $\mathbf{n}^\top \mathbf{V} = \mathbf{0}^\top$:

$$\mathbf{n}^\top \left[\begin{array}{c|c|c|c} \mathbf{x}'_0 - \bar{\mathbf{x}}' & \mathbf{x}'_1 - \bar{\mathbf{x}}' & \ldots & \mathbf{x}'_{n-1} - \bar{\mathbf{x}}' \\ \mathbf{x}_0 - \bar{\mathbf{x}} & \mathbf{x}_1 - \bar{\mathbf{x}} & \ldots & \mathbf{x}_{n-1} - \bar{\mathbf{x}} \end{array} \right] = \mathbf{0}^\top.$$

To obtain a unique, non-trivial solution for $\mathbf{n}$, $\mathbf{V}$ must have rank three (as noted in an equivalent formulation by Tomasi and Kanade (1992), who solved for rigid structure and motion under orthographic projection, although they didn't investigate rank deficiency). No unique solution can be found when the rank of $\mathbf{V}$ is *less than* 3. The rank 2 case typically arises when there is a 2D affine transformation, $\mathbf{x}'_i = \Upsilon \mathbf{x}_i + \mathbf{f}$ (where Υ and $\mathbf{f}$ are fixed), so $\mathbf{V}$ only has two independent rows. This occurs, for instance, when there is pure translation parallel to the image plane; when there is only rotation about the optic axis; or when the object is planar. The rank 1 case arises in the unlikely event that the points are additionally collinear. If $\mathbf{V}$ has *full* rank (rank 4), and this cannot be attributed to noise, then either the camera model is invalid (e.g., there are perspective effects) or there are independent motions in the scene. The noise characteristics of linear least squares solutions were analysed in (Weng et al. 1989; Kanatani 1993; Shapiro and Brady 1993). The noise present in $\mathbf{r}_i$ propagates through to the solution vector $\mathbf{n}$, whose covariance matrix $\Lambda_\mathbf{n}$ provides a confidence measure in the parameters of the epipolar fit. Suppose each data point $\mathbf{r}_i$ is perturbed by independent, isotropic, additive, Gaussian noise $\delta\mathbf{r}_i$. The noise has zero mean ($E\{\delta\mathbf{r}_i\} = \mathbf{0}$) with variance σ^2, so $E\{\delta\mathbf{r}_i \delta\mathbf{r}_j^\top\} = \delta_{ij}\sigma^2\mathbf{I}_4$, where δ_{ij} is the Kronecker delta function and $\mathbf{I}_4$ the 4×4 identity matrix. The eigenvalues $\{\lambda_1, \ldots, \lambda_4\}$ of $\mathbf{W}$ are arranged in increasing order with corresponding eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. The eigenvector corresponding to the minimum eigenvalue, $\mathbf{u}_1$, gives the solution vector $\mathbf{n}$. The covariance matrix for $\mathbf{n}$ is (Shapiro and Brady 1993):

$$\Lambda_\mathbf{n} = E\{\delta\mathbf{n}\delta\mathbf{n}^\top\} = \sigma^2 \sum_{k=2}^{4} \frac{\mathbf{u}_k \mathbf{u}_k^\top}{\lambda_k}. \qquad (30)$$

This covariance matrix is important for it provides a confidence measure in the parameters of the epipolar fit, which could in turn be used to compute uncertainty regions around the epipolar lines. Furthermore, it facilitates the rejection of *outliers*, "rogue observations" which plague data analysis techniques such as linear least squares regression. The outlier problem arises when a given data set actually comprises two subsets: a large, dominant body of valid data and a relatively small set of contaminants. Removing the outliers is crucial since an analysis based on the contaminated data set distorts the underlying parameters. The task is particularly complicated when (as is normally the case) the data are also noisy, and few existing structure from motion algorithms have outlier immunity (though see Torr et al. 1993). We employ the eigenvalue-based regression diagnostic of Shapiro and Brady (1993), which pinpoints potentially unsuitable points by means of an *influence function*. It computes the extent to which a particular point influences the fit by determining the change in solution when the data point is omitted, and uses a χ^2 test to ascertain when

all outliers have been removed (to a specified confidence level). The final algorithm is summarised below.

Task

Given two sets of n image points, $\mathbf{x}_i$ and $\mathbf{x}'_i$ ($i = 0 \cdots n-1$), compute the parameters of the affine epipolar geometry $\{\mathbf{n}, e\}$, if possible.

Algorithm

1. Compute the data centroid $\bar{\mathbf{r}}$ and centre the points, giving $\mathbf{v}_i = \mathbf{r}_i - \bar{\mathbf{r}}$.
2. Construct the real, symmetric scatter matrix $\mathbf{W} = \sum_{i=0}^{n-1} \mathbf{v}_i \mathbf{v}_i^\top$. Compute its rank, and stop unless rank $(\mathbf{W}) = 3$ (within the bounds of noise).
3. Solve for $\mathbf{n}$ by minimising E_3 (linear eigenvector computation).
4. Calculate $e = -\mathbf{n}^\top \bar{\mathbf{r}}$.
5. Calculate the covariance matrix $\Lambda_{\mathbf{n}}$ for $\mathbf{n}$ (Eq. (30)).

Affine epipolar geometry algorithm.

3.5 *Experimental Results*

Fig. 12 shows a sequence taken at frame rate (25 Hz) from a robot head rotating in a static environment. Corner features were extracted (Fig. 12(c)) and tracked, and motion vectors inconsistent with the main statistical distribution (i.e. outliers) were removed using the scheme in (Shapiro and Brady 1993). Fig. 12(d) shows the motion vectors that remain; in general, these need not be parallel (cf. Fig. 8(c)). Finally, Figs. 12(e) and (f) show the epipolar lines computed using the algorithm in Section 3.4.5. The sum of squared perpendicular distances in I and I' is 141.55, giving a mean perpendicular distance of 0.55 pixels between each corner and its epipolar line. This error value depends on the 'noise' in the system, which is a function of the particular camera, the feature detection algorithm (which is this case localises corners to the nearest pixel), and the presence of outliers.

Fig. 13 shows two other sequences, with a camera moving in a static world and with an object moving relative to a stationary camera. The mean perpendicular distances are 0.76 and 0.49 pixels respectively. The epipolar lines are thus typically within a pixel accuracy (on 256×256 images), despite the lack of subpixel accuracy in the corner detection stage. These epipolar lines therefore provide effective constraints for correspondence.

Finally, we illustrate the advantage of using *all* available points when computing the epipolar geometry. A synthetic scene with 63 points had its images corrupted by independent, isotropic, Gaussian noise ($\sigma = 0.6$ pixels). Subsets of the data comprising p points (where p varied from 4 to 63) were randomly selected and a fit $\{\mathbf{n}, e\}$ computed using this subset. The E_1 distance was then computed for the whole point set, summing the squared perpendicular image distances from each point to its computed epipolar line. For each value of p, 500 experiments were run. Fig. 14 shows the median distance and the standard deviation of the distances for each value of p. Both decrease as p increases, showing that the use of more points leads not only to better fits but also to more consistent ones. The facility to perform outlier rejection is another good justification for using all the available points, since outliers cannot be identified using minimal point sets; redundancy in the data set is required.

4 Rigid Motion: Two Views

It is well-known that two distinct views of four non-coplanar, *rigid* points generate a one-parameter family of structure and motion solutions under parallel projection (Bennett et al. 1989; Huang and Lee 1989; Koenderink and van Doorn 1991). Koenderink and van Doorn (1991) showed further that from two views, the change in scale, the cyclorotation angle and the projections of the axes of rotation can be recovered. Their algorithm used the minimum set of points, required a perceptual frame and involved a succession of stages. These shortcomings are overcome here by deriving this partial motion solution *directly from the affine epipolar geometry.* The resulting algorithm uses the full set of points and requires no perceptual frame.

4.1 *Motion Ambiguities*

Ambiguous interpretations of object motion (and structure) arise when two or more unknown parameters are confounded and there are insufficient data to identify their individual contributions to the imaged effect. For instance, the "depth-scale" ambiguity means that while the algorithm can recover scaled depth $\Delta Z_i^c / Z_{\text{ave}}^c$, the true depth ΔZ_i^c cannot be determined unless Z_{ave}^c is known. The true depth of at least one point on the object is therefore needed to resolve this ambiguity.

Two well-known ambiguities specific to parallel projection are *Necker reversal* and the *bas-relief ambiguity.* Necker reversal occurs because an object rotating by ρ and its mirror object (reflected about the image plane)

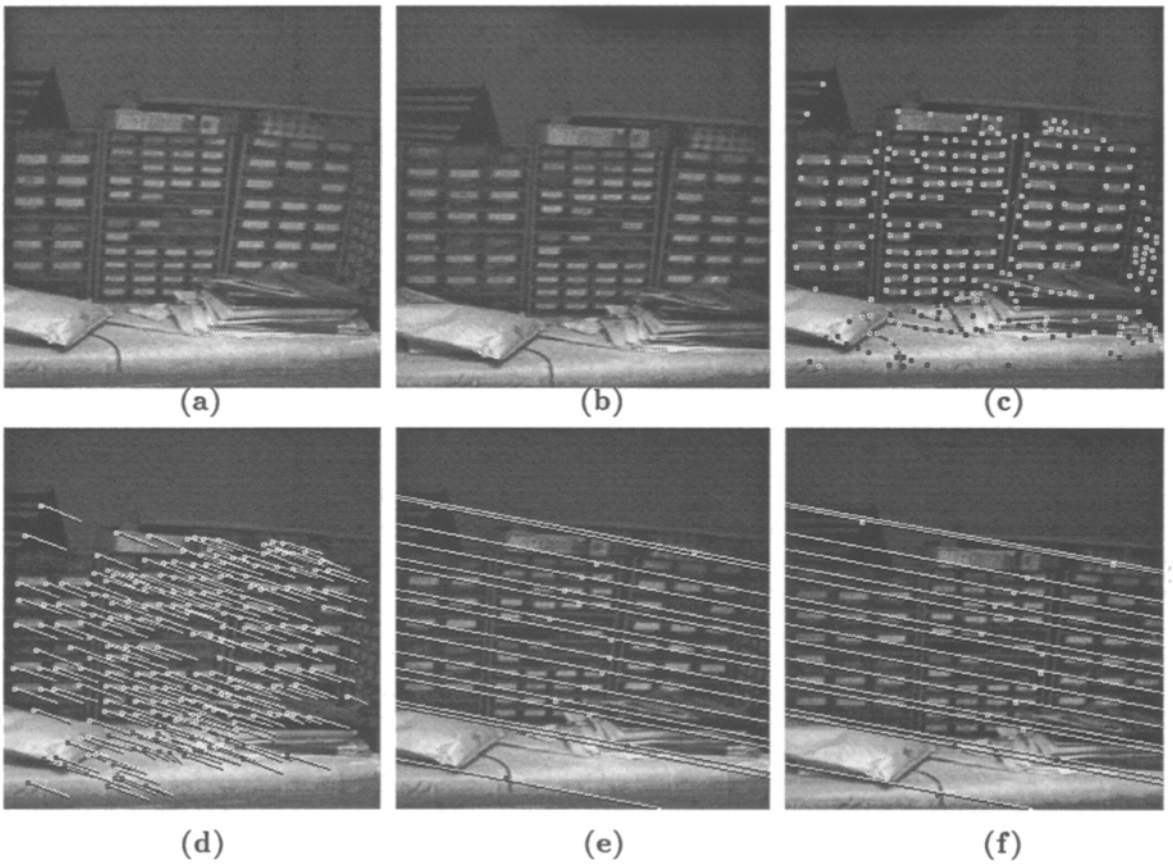

Fig. 12. Computing epipolar geometry for a real sequence: (a), (b) frames 0 and 5 (I and I'); (c) corner features; (d) 231 motion vectors remain after outlier rejection; (e), (f) epipolar lines for I and I' (every 15$^{\text{th}}$ line shown).

rotating by $(-\rho)$ generate the same image under parallel projection. Thus, structure can only be recovered up to a reflection about the frontal plane. The bas-relief ambiguity is illustrated in Figs. 15(a) and (b), where a rigid rod of length ℓ starts in position OA and rotates about O in the X-Z plane through angle ρ. Clearly, $X'_A - X_A = \ell \sin \rho$, and if $X'_A - X_A$ is observed without any prior knowledge of the rod or its motion, then it is impossible to deduce both ℓ and ρ, because an infinite family of (ℓ, ρ) solutions could generate the same image.

This bas-relief (or "depth-turn") ambiguity is so-named because a shallow object experiencing a large turn (i.e., small ℓ and big ρ) generates the same image as a deep object experiencing a small turn (i.e., big ℓ and small ρ). Extra points cannot resolve this ambiguity since each new point adds one new piece of information (X) and one new unknown (its depth). Fixing any *one* depth (or the angle ρ) determines the structure and the motion uniquely, i.e., there is a *one-parameter* family of solutions (Bennett et al. 1989; Huang and Lee 1989; Koenderink and van Doorn 1991). This ambiguity can only be resolved from two views when perspective effects are significant; otherwise, three views are needed.

4.2 *Weak Perspective Epipolar Geometry*

Rigidity is imposed on the world motion parameters $\{\mathbf{A}_m, \mathbf{D}_m\}$ by requiring $\mathbf{A}_m$ to be a rotation matrix[10], denoted $\mathbf{A}_m = \mathbf{R}_m$. Equation (16) then becomes

$$\mathbf{X}' = \mathbf{R}_m \mathbf{X} + \mathbf{T}_m. \tag{31}$$

Rigidity reduces the degrees of freedom in the motion parameters from 12 to 6, and the use of difference vectors eliminates $\mathbf{T}_m$,

$$\Delta \mathbf{X}' = \mathbf{X}' - \mathbf{X}'_0 = \mathbf{R}_m \Delta \mathbf{X},$$

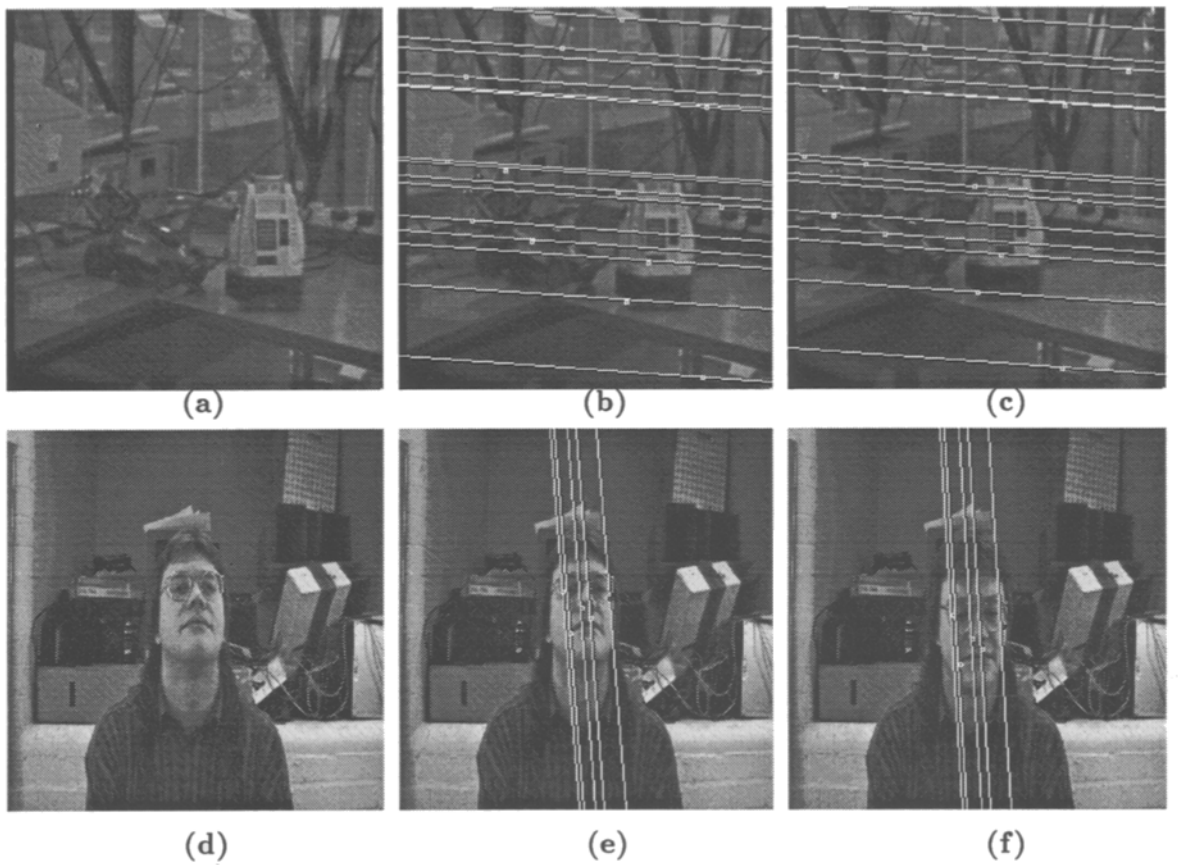

Fig. 13. Affine epipolar lines: (a)–(c) the camera moves (every10$^{\text{th}}$ line shown); (d)–(f) the object moves (every 2$^{\text{nd}}$ line shown).

leaving the three rotational degrees of freedom. There are various ways to parameterise these rotation angles (see Appendix D), the most popular being Euler angles and the angle-axis form. Solving for $\mathbf{R}$ requires the measurement of image angles, which are not affine invariants; it is therefore necessary to use *weak perspective* cameras here, namely $\mathbf{M}$ and $\mathbf{M}'$ (see Eq. (8)):

$$\mathbf{M} = \frac{f}{Z^c_{\text{ave}}}\begin{bmatrix} \xi \mathbf{R}^\top_{p1} \\ \mathbf{R}^\top_{p2} \end{bmatrix} \quad \text{and} \quad \mathbf{M}' = \frac{f'}{Z^{c'}_{\text{ave}}}\begin{bmatrix} \xi' \mathbf{R}'^\top_{p1} \\ \mathbf{R}'^\top_{p2} \end{bmatrix}.$$

Here $\mathbf{R}_p$ and $\mathbf{R}'_p$ are rotation matrices representing the camera positions relative to the world coordinate frame. The aspect ratios ξ and ξ' must be known in order to compute angles, and the ratio of focal lengths f/f' must be known (or must equal unity if the focal lengths are unknown) in order to determine scale. No other calibration parameters are needed. Without loss of generality, we set $\mathbf{R}_p = \mathbf{I}_3$ (the identity matrix) and $\mathbf{R}'_p = \mathbf{R}$, incorporating the composite camera poses and object motion into $\mathbf{R}$. The scale factor $s = Z^c_{\text{ave}}/Z^{c'}_{\text{ave}}$ is introduced ($s > 1$ for a "looming" object and $s < 1$ for one that is "receding"), and scaled depth is defined as $\Delta z_i = f\Delta Z^c_i / Z^c_{\text{ave}} = f(Z^c_i - Z^c_{\text{ave}})/Z^c_{\text{ave}}$. Eq. (18) then becomes

$$\boxed{\Delta \mathbf{x}' = s\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}\Delta \mathbf{x} + s\Delta z \begin{bmatrix} R_{13} \\ R_{23} \end{bmatrix}} \quad (32)$$

This is the rigid motion form of the epipolar line, with direction $(R_{13}, R_{23})^\top$. Resolving $\Delta \mathbf{x}'$ perpendicular to this direction and using standard cross product equalities for a rotation matrix gives the rigid motion form of the affine epipolar constraint equation (Eq. (22)):

$$\boxed{R_{23}\Delta x' - R_{13}\Delta y' + sR_{32}\Delta x - sR_{31}\Delta y = 0} \quad (33)$$

Equations (32) and (33) generalise the pure orthographic forms ($s = 1$) derived by Huang and Lee

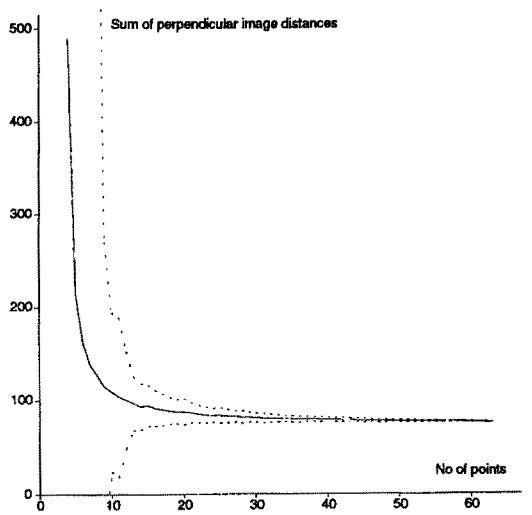

Fig. 14. Improvement in the epipolar geometry as the number of points increases. The solid line shows the median distance (from 500 experiments) and the two dotted lines show the standard deviation ($\pm 1\sigma$ level). See text for details.

(1989) and used in (Chen and Huang 1991; Hu and Ahuja 1991). There are only three independent degrees of freedom in Eq. (33), since only the ratios of the coefficients may be computed; we will show these to be the scale factor s and two rotation angles.

4.3 The KvD Rotation Representation

Koenderink and van Doorn (1991) introduced a novel rotation representation (which we term *KvD* and show in Appendix D to be a variant of Euler angles), and presented a geometric analysis of it. We formalise their representation algebraically to illustrate its advantages. In *KvD*, a rotation matrix $\mathbf{R}$ is decomposed into two parts (Fig. 16),

$$\mathbf{R} = \mathbf{R}_\rho \mathbf{R}_\theta. \tag{34}$$

First, there is a rotation $\mathbf{R}_\theta$ in the image plane through angle θ (i.e., about the line of sight). This is followed by a rotation $\mathbf{R}_\rho$ through an angle ρ about a unit axis $\mathbf{\Phi}$ lying in a plane parallel to the image plane, with $\mathbf{\Phi}$ angled at ϕ to the positive X axis, i.e., a pure rotation *out of* the image plane. We define $\mathbf{\Phi}$ and its perpendicular $\mathbf{\Phi}^\perp$ as $\mathbf{\Phi}^\perp$ as

$$\mathbf{\Phi} = \begin{bmatrix} \cos\phi \\ \sin\phi \end{bmatrix} \quad \text{and} \quad \mathbf{\Phi}^\perp = \begin{bmatrix} \sin\phi \\ -\cos\phi \end{bmatrix}.$$

The *KvD* representation has two main advantages. First, rotation about the optic axis provides no new information about structure, so it makes sense to first remove this "useless" component. Second, *KvD* isolates the bas-relief ambiguity in away that the more popular angle-axis form doesn't—an advantage of Euler forms in general (Harris 1990). Thus, two views determine

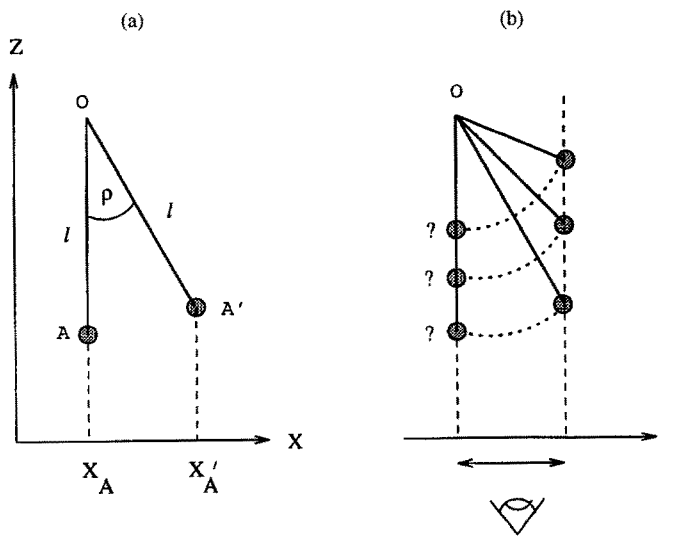

Fig. 15. The bas-relief motion ambiguity under parallel projection: (a) a rigid rod OA rotates through ρ in the X-Z plane; (b) from observing $X'_A - X_A = \ell \sin\rho$, it is impossible to deduce both ℓ and ρ.

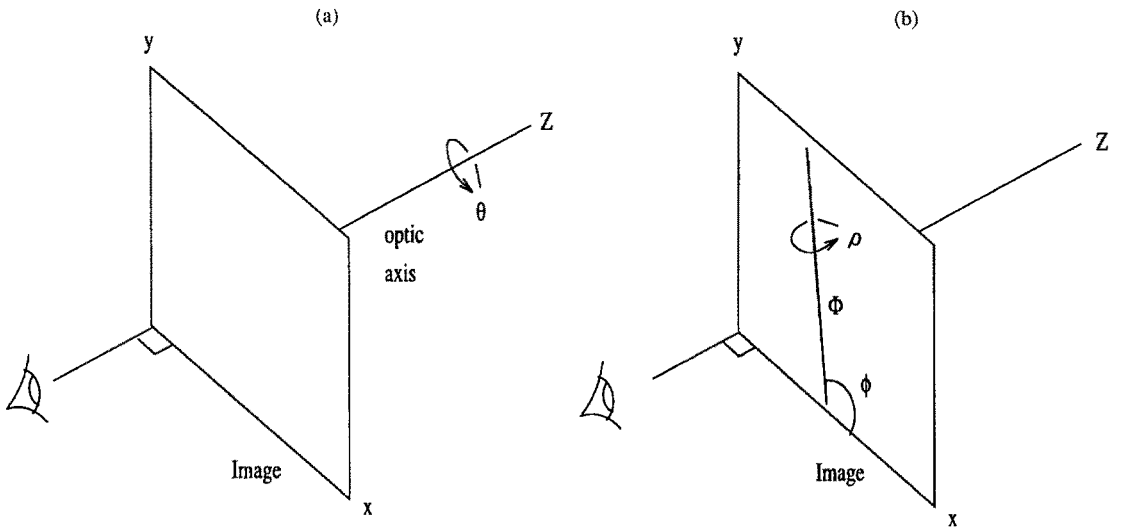

Fig. 16. The *KvD* rotation representation: (a) rotation by θ about the Z axis; (b) subsequent rotation by ρ about a fronto-parallel axis Φ angled at ϕ to the X axis.

the two rotation angles (ϕ and θ), with the third angle (ρ) parameterising the remaining family of solutions. This contrasts with, say, the angle-axis form (Appendix D.1), for which only one angle can be solved from two views, the two remaining angles satisfying a non-linear constraint equation (Bennett et al., 1989). The disadvantage of *KvD* is that the physical interpretation of rotation occurring about a single 3D axis is lost.

4.4 Solving for s, ϕ and θ

It is now shown how to solve for the scale factor (s), the projection of the axis of rotation (ϕ) and the cyclotorsion angle (θ) directly from the affine epipolar geometry. The formal derivations are preceded by a geometric explanation of how the epipolar lines relate to the unknown motion parameters.

4.4.1 Geometry. Consider a camera rotating about an axis $\mathbf{\Phi}$ lying parallel to the image plane (Fig. 17(a)). The epipolar plane π is perpendicular to both this axis and the two images, and intersects the images in the epipolar lines $\mathbf{u}$ and $\mathbf{u}'$. This is because a world point A defines the epipolar plane π and projects onto the images along these lines of intersection (cf. Fig. 6). Consequently:

The projection of the axis of rotation $\mathbf{\Phi}$ is perpendicular to the epipolar lines.

This relation still holds if there is additionally a cyclotorsion θ in the image plane (Fig. 17(b)); the axis $\mathbf{\Phi}$ and intersection $\mathbf{u}'$ remain fixed in space, and are simply observed at a new angle in the image, maintaining the orthogonality between the epipolar lines and the projected axis. The orientations of the epipolars in the two images therefore differ by θ. Importantly, changing the magnitude of the turn angle ρ doesn't alter the epipolar geometry in any way (Fig. 18). This angle is therefore indeterminate from two views, a consequence of the bas-relief ambiguity. Fig. 18 also illustrates the effect of scale. Consider a 3D object to be sliced into parallel epipolar planes, with each plane constraining how a particular slice of the object moves. Altering the effective size of the object (e.g., by moving closer to it) simply changes the relative spacing between successive epipolar planes.

In summary, cyclotorsion simply rotates the epipolars, rotation out of the plane causes foreshortening along the epipolar lines (orthogonal to $\mathbf{\Phi}$), and a scale change uniformly alters the epipolar line spacing (Fig. 19).

4.4.2 Theory. Substituting the *KvD* expressions for R_{ij} into Eq. (32) gives the direction of the epipolar line in I',

$$\begin{bmatrix} R_{13} \\ R_{23} \end{bmatrix} = \sin\rho\,\mathbf{\Phi}^{\perp},$$

proving that the epipolar is perpendicular to the axis of rotation $\mathbf{\Phi}$. Similarly, the epipolar constraint (Eq. (33))

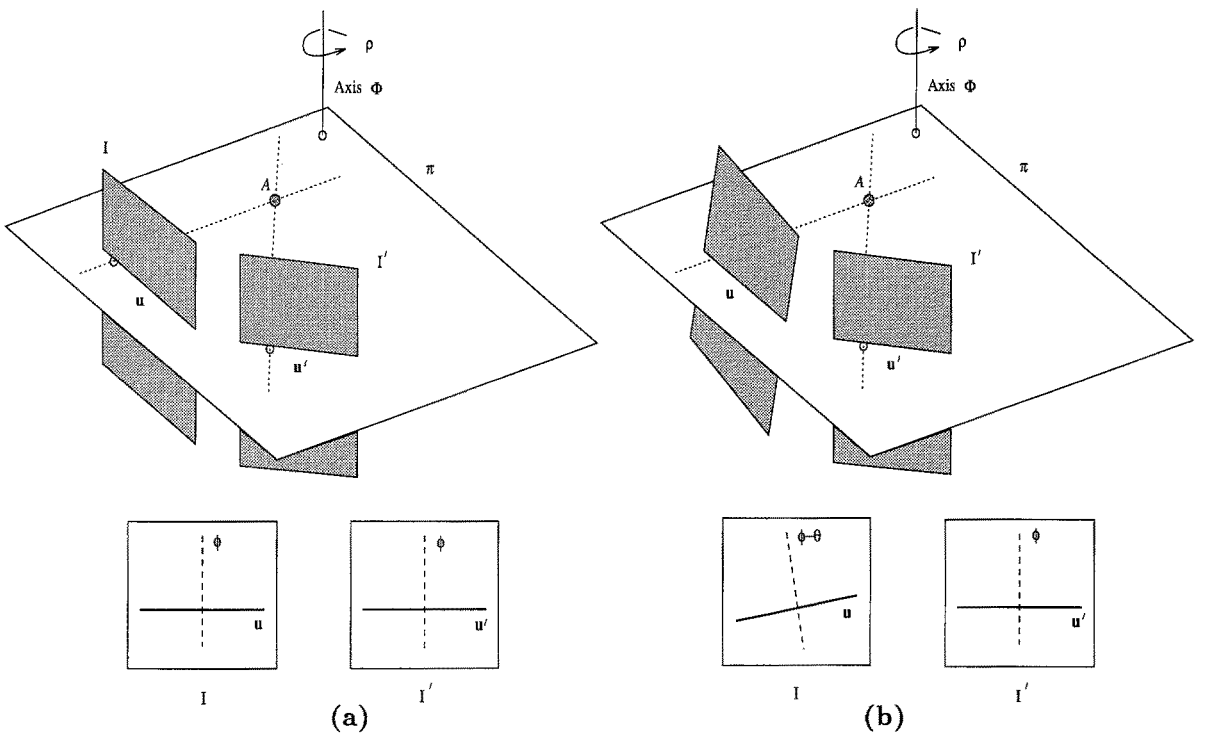

Fig. 17. The camera rotates about the axis Φ which is parallel to the image plane. The intersection of the epipolar plane π with the image planes gives epipolar lines $\mathbf{u}$ and $\mathbf{u}'$, and the projections of Φ in the images are orthogonal to these epipolars: (a) no cyclotorsion occurs ($\theta = 0°$); (b) the camera counter-rotates by θ in I, so the orientations of the epipolars change by θ.

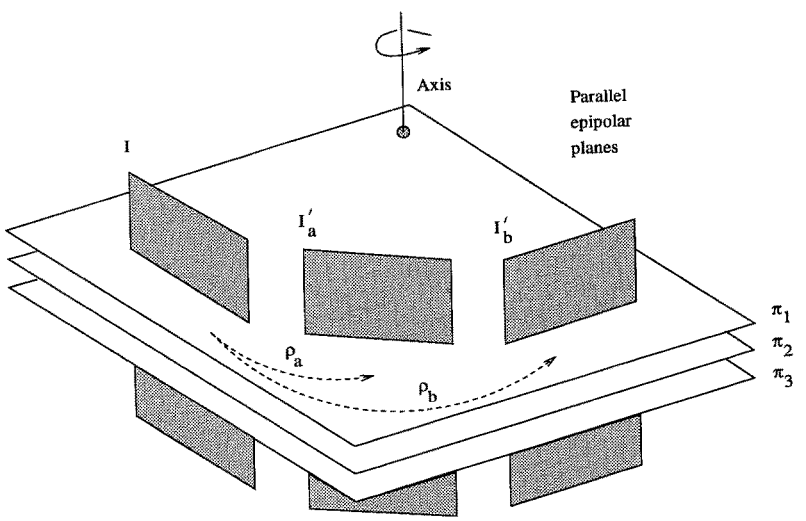

Fig. 18. The scene can be sliced into parallel epipolar planes. The magnitude of ρ has no effect on the epipolar geometry (provided $\rho \neq 0$), so it is indeterminate from two views.

becomes

$$\sin\rho\,[\cos\phi\Delta x_i' + \sin\phi\Delta y_i' - s\cos(\phi-\theta)\Delta x_i - s\sin(\phi-\theta)\Delta y_i] = 0 \tag{35}$$

It is evident from Eq. (35) that s, θ and ϕ can be computed directly from the affine epipolar geometry, because the algorithm in Section 3.4.5 solves

$$a\Delta x_i' + b\Delta y_i' + c\Delta x_i + d\Delta y_i = 0$$

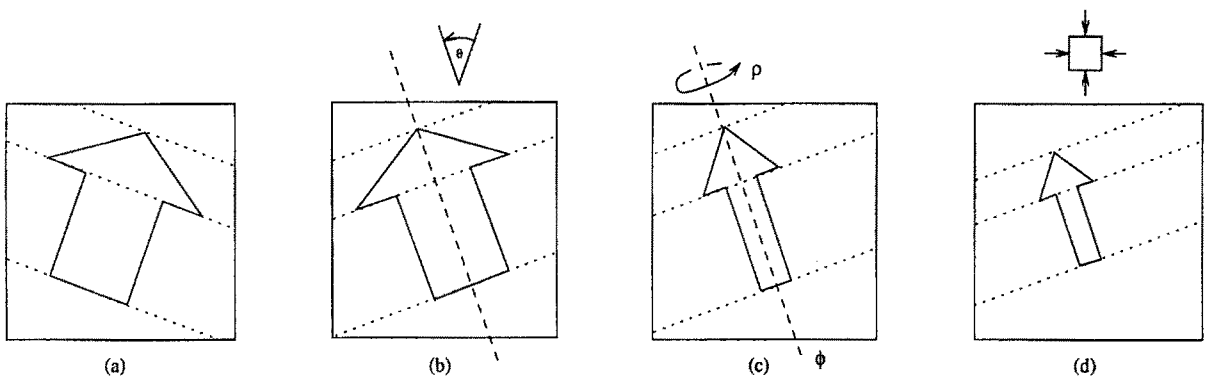

Fig. 19. The effect of scale and *KvD* rotation angles on the epipolar lines for an object moving relative to a stationary camera. This also illustrates the assumed sequence of events accounting for the transition from I to I': (a) I; (b) cyclotorsion (θ); (c) rotation out of the plane (ϕ and ρ); (d) scaling, giving I'.

for the ratios of a, b, c and d. A direct comparison with Eq. (35) yields the central result,

$$\tan\phi = \frac{b}{a}, \quad \tan(\phi - \theta) = \frac{d}{c} \quad \text{and} \quad s^2 = \frac{c^2 + d^2}{a^2 + b^2}, \tag{36}$$

with $s > 0$ (by definition). Note that ϕ is the angle of projection in I' of the axis of rotation out of the plane, while $(\phi - \theta)$ is its angle of projection in I. Equation (35) shows immediately that Eq. (33) has only *two* independent rotation parameters, θ and ϕ, because the angle ρ cancels out (provided it is non-zero). If $\rho = 0$, then there is no rotation out of the image plane and $\mathbf{\Phi}$ is obviously undefined, so this technique cannot be used. Equation (35) is therefore more informative than Eq. (33) since it identifies explicitly what quantities can be computed, and under what circumstances.

The minimisation of Eq. (22) followed by the use of Eq. (36) is equivalent to minimising Eq. (35) over s, ϕ and θ directly, though the former approach is simpler; the functions differ only in a change of coordinate systems.

4.5 *Error Model*

We now compute a noise model for s, ϕ and θ, each of which is a non-linear function of $\mathbf{n} = (a, b, c, d)^\top$, itself a random variable due to image noise. Given the covariance matrix $\mathbf{\Lambda}_\mathbf{n} = [\Lambda_{ij}]$ for $\mathbf{n}$ from Eq. (30), our objective is to compute the means and variances of s, ϕ and θ.

Let the true (i.e., noise-free) value of $\mathbf{n}$ be $\tilde{\mathbf{n}}$, and write $\mathbf{n} = (n_1, n_2, n_3, n_4)^\top$ for convenience. The noise perturbation of $\tilde{\mathbf{n}}$ is $\delta\mathbf{n} = (\delta n_1, \delta n_2, \delta n_3, \delta n_4)^\top$, so that $\mathbf{n} = \tilde{\mathbf{n}} + \delta\mathbf{n}$. The diagonal elements of $\mathbf{\Lambda}_\mathbf{n}$ define the variance of $\delta\mathbf{n}(\Lambda_{ii} = E\{\delta n_i^2\})$ while the off-diagonal elements give the covariances $(\Lambda_{ij} = \Lambda_{ji} = E\{\delta n_i \delta n_j\})$. The Taylor series for a function $q(\tilde{\mathbf{n}})$ expanded about $\mathbf{n}$ is[11]

$$q(\tilde{\mathbf{n}}) = q(\mathbf{n} - \delta\mathbf{n}) = q(\mathbf{n}) - \sum_{i=1}^{4} \frac{\partial q(\mathbf{n})}{\partial \tilde{n}_i} \delta n_i + \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{\partial^2 q(\mathbf{n})}{\partial \tilde{n}_i \partial \tilde{n}_j} \delta n_i \delta n_j - \cdots$$

We ignore terms above second order, assume that $\partial^2 q / \partial n_i \partial n_j = \partial^2 q / \partial n_j \partial n_i$, and note that $E\{\delta\mathbf{n}\} = \mathbf{0}$ and $E\{\tilde{\mathbf{n}}\} = \tilde{\mathbf{n}}$. Then,

$$E\{q(\tilde{\mathbf{n}})\} \approx E\{q(\mathbf{n})\} + \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{\partial^2 q(\mathbf{n})}{\partial \tilde{n}_i \partial \tilde{n}_j} E\{\delta n_i \delta n_j\}$$

$$\text{i.e.} \quad q(\tilde{\mathbf{n}}) \approx E\{q(\mathbf{n})\} + \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{\partial^2 q(\mathbf{n})}{\partial \tilde{n}_i \partial \tilde{n}_j} \Lambda_{ij}.$$

This is a *biased* estimate, since the expected value of $q(\mathbf{n})$ does not equal the true value $q(\tilde{\mathbf{n}})$. The systematic error (or *bias*) is

$$B = \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \frac{\partial^2 q(\mathbf{n})}{\partial \tilde{n}_i \partial \tilde{n}_j} \Lambda_{ij} \tag{37}$$

where $E\{q(\mathbf{n})\} = q(\tilde{\mathbf{n}}) - B$. Since $\mathbf{n}$ and $\mathbf{\Lambda}_\mathbf{n}$ are known, B is fully determined and is added to the

computed value $q(\mathbf{n})$. The variance of q is $E\{[q(\mathbf{n}) - E\{q(\mathbf{n})\}]^2$, i.e.

$$\begin{aligned}\mathrm{var}\{q(\mathbf{n})\} \approx E\Bigg\{&\sum_{i=1}^{4}\frac{\partial q(\mathbf{n})}{\partial \tilde{n}_i}\delta n_i - \frac{1}{2} \\ &\times \sum_{i=1}^{4}\sum_{j=1}^{4}\frac{\partial^2 q(\mathbf{n})}{\partial \tilde{n}_i \partial \tilde{n}_j}\delta n_i \delta n_j + B\Bigg\}^2 \\ =&\sum_{i=1}^{4}\sum_{j=1}^{4}\frac{\partial q(\mathbf{n})}{\partial \tilde{n}_i}\frac{\partial q(\mathbf{n})}{\partial \tilde{n}_j}\Lambda_{ij}. \end{aligned} \tag{38}$$

Similarly, the covariance between two different expressions $q_1(\mathbf{n})$ and $q_2(\mathbf{n})$ is given by $E\{[q_1(\mathbf{n}) - E\{q_1(\mathbf{n})\}][q_2(\mathbf{n}) - E\{q_2(\mathbf{n})\}]\}$, i.e.

$$\mathrm{cov}\{q_1(\mathbf{n}), q_2(\mathbf{n})\} \approx \left(\sum_{i=1}^{4}\sum_{j=1}^{4}\frac{\partial q_1(\mathbf{n})}{\partial \tilde{n}_i}\frac{\partial q_2(\mathbf{n})}{\partial \tilde{n}_j}\Lambda_{ij}\right). \tag{39}$$

The three functions of interest are

$$s(\mathbf{n}) = \frac{\sqrt{n_3^2 + n_4^2}}{\sqrt{n_1^2 + n_2^2}}, \quad \phi(\mathbf{n}) = \frac{n_1}{\sqrt{n_1^2 + n_2^2}}$$

and

$$\theta(\mathbf{n}) = \frac{n_1}{\sqrt{n_1^2 + n_2^2}} - \frac{n_3}{\sqrt{n_3^2 + n_4^2}},$$

and it is straightforward to derive the relevant bias, variance and covariance expressions. These provide confidence regions for the solution, and are used later in the Kalman filter.

4.6 Algorithm

The final algorithm is given below. The *centroid* serves as the reference point rather than some designated feature point (e.g., $\mathbf{x}_0$) since this relates better to the minimisation theory of Section 3.4. The solution for θ in Eq. (36) has a 180° ambiguity since only the *directions* of the projected axes are known, not their positive or negative senses. In other words, it remains to determine "which way" the axis cyclo-rotated between I and I'. Equation (35) is therefore used to check whether $(\phi - \theta)$ is correct.

The algorithm fails in two cases: (i) the object is planar; and (ii) the object is non-planar but doesn't rotate out of the image plane[12]. The rank test of Section 3.4.5 signals these problems because in both cases, a 2D matrix Γ explains the image motion ($\Delta\mathbf{x}' = \Gamma\,\Delta\mathbf{x}$), so the system has rank two (within the bounds of noise).

Task

Given $\{a, b, c, d\}$ (computed from the affine epipolar geometry algorithm using two distinct views of at least four non-coplanar points), determine the scale s, cyclotorsion θ and axis projection ϕ.

Algorithm

1. Set $s = \sqrt{c^2 + d^2}/\sqrt{a^2 + b^2}$.
2. Set $\phi = \arctan(b/a)$ and $\theta = \phi - \arctan(d/c)$.
3. Calculate the sums of square residuals

$$r^- = \sum_{i=0}^{n-1}\left[\Delta\mathbf{x}_i'^{\top}\Phi - s\Delta\mathbf{x}_i^{\top}\mathbf{R}(-\theta)\Phi\right]^2,$$

$$r^+ = \sum_{i=0}^{n-1}\left[\Delta\mathbf{x}_i'^{\top}\Phi + s\Delta\mathbf{x}_i^{\top}\mathbf{R}(-\theta)\Phi\right]^2.$$

If $r^+ < r^-$, set $\theta \leftarrow \theta + 180°$.

Weak perspective epipolar motion algorithm.

Fig. 20(a) graphs successive two-frame estimates of s, θ and ϕ in a synthetic 30-frame sequence with additive Gaussian image noise. The object undergoes two different, constant motions separated by a step change: s is initially unity (no translation in depth) and then increases as the object approaches the camera at constant speed; θ changes from 4° to −2°; and ϕ changes from 82° to 50°. The true parameter values clearly lie within the computed 95% error bounds.

4.7 Kalman Filter

Physical objects have inertia and it is sensible to exploit this temporal continuity to improve the motion estimates. This is achieved by means of a *linear* discrete-time Kalman filter (Bar-Shalom and Fortmann 1988; Durrant-Whyte 1993), a popular framework for weighting observations and predictions. The state-transition equation (*plant* model) describes the evolution of the state $\mathbf{y}$,

$$\mathbf{y}(k+1) = \mathbf{F}\mathbf{y}(k) + \mathbf{g}(k) \tag{40}$$

where $\mathbf{y}(k)$ is the state at time k, $\mathbf{g}(k)$ is the constant additive process noise, $\mathbf{F}$ is the state-transition matrix, and $\mathbf{y}(k+1)$ is the state at the following $k+1$ time step. Observations of the system are made according to the *measurement* model

$$\mathbf{z}(k+1) = \mathbf{H}\mathbf{y}(k+1) + \mathbf{w}(k+1) \tag{41}$$

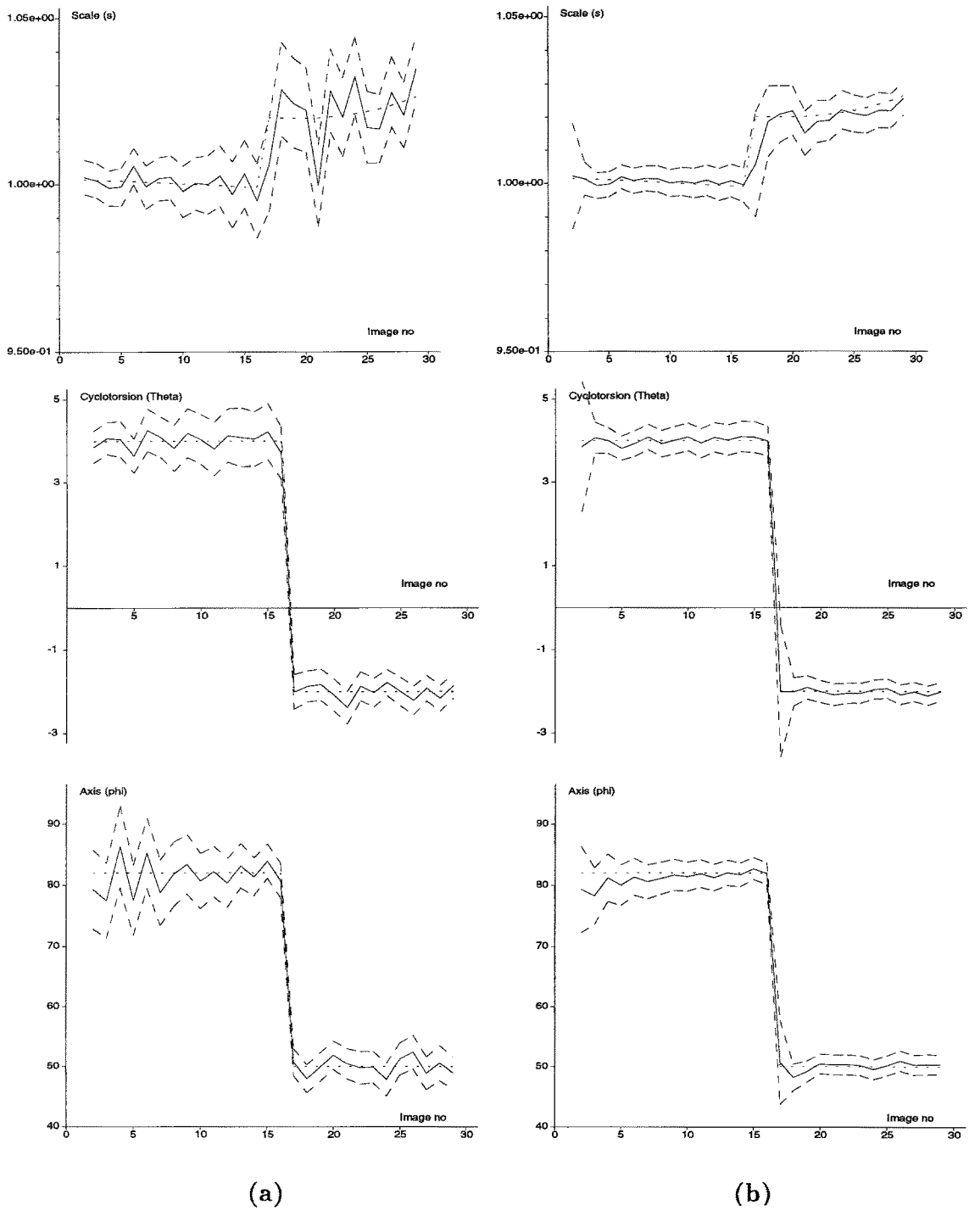

Fig. 20. Estimates of s, θ and ϕ over a 30-frame synthetic sequence with $\sigma = 0.5$ pixels and a discontinuity in constant motion. Solid lines show computed values, dotted lines show true values, and the two dashed lines show the upper and lower 95% confidence intervals: (a) Raw 2-frame estimates with associated error models (input to filter); (b) Kalman filtered estimates with 95% confidence regions (within which the true values lie).

where $\mathbf{z}(k+1)$ is the observation at time $k+1$, $\mathbf{w}(k+1)$ is the additive observation noise, and $\mathbf{H}$ is the observation matrix. The noise vectors $\mathbf{g}(k)$ and $\mathbf{w}(k)$ are assumed to be Gaussian, identically distributed, and temporally uncorrelated with zero mean. Thus, $E\{\mathbf{g}(k)\} = E\{\mathbf{w}(k)\} = \mathbf{0}$, $E\{\mathbf{g}(i)\mathbf{g}(j)^\top\} = \delta_{ij}\mathbf{\Lambda}_\mathbf{g}$ and $E\{\mathbf{w}(i)\mathbf{w}(j)^\top\} = \delta_{ij}\mathbf{\Lambda}_\mathbf{w}(i)$, where $\mathbf{\Lambda}_\mathbf{g}$ and $\mathbf{\Lambda}_\mathbf{w}$ are covariance matrices. The complete predict-observe-update cycle is:

Predict state and variance:

$$\hat{\mathbf{y}}(k+1 \mid k) = \mathbf{F}\hat{\mathbf{y}}(k \mid k) \tag{42}$$

$$\mathbf{P}(k+1 \mid k) = \mathbf{F}\mathbf{P}(k \mid k)\mathbf{F}^\top + \mathbf{\Lambda}_\mathbf{g} \tag{43}$$

Update state and variance:

$$\hat{\mathbf{y}}(k+1 \mid k+1) = \hat{\mathbf{y}}(k+1 \mid k) + \mathbf{W}(k+1) \times [\mathbf{z}(k+1) - \mathbf{H}\hat{\mathbf{y}}(k+1 \mid k)] \tag{44}$$

$$\mathbf{P}(k+1 \mid k+1) = \mathbf{P}(k+1 \mid k) - \mathbf{W}(k+1) \times \mathbf{S}(k+1)\mathbf{W}^\top(k+1) \tag{45}$$

where $\mathbf{W}(k+1) = \mathbf{P}(k+1 \mid k)\mathbf{H}^\top\mathbf{S}^{-1}(k+1)$ and $\mathbf{S}(k+1) = \mathbf{H}\mathbf{P}(k+1 \mid k)\mathbf{H}^\top + \mathbf{\Lambda}_\mathbf{w}(k+1)$.

We estimate s, ϕ and θ using a constant position model ($\dot{s} = \dot{\phi} = \dot{\theta} = 0$). The state vector is $\mathbf{y} = (s, \phi, \theta)^\top$ with state transition matrix $\mathbf{F} = \mathbf{I}_3$. We observe[13] s, ϕ and $\phi - \theta$, so $\mathbf{z} = (z_1, z_2, z_3)^\top = (s + B_1, \phi + B_2, \phi - \theta + B_3)^\top$, where B_i are the relevant bias terms from Eq. (37). The observation matrix ($\mathbf{H}$) and its noise covariance matrix ($\mathbf{\Lambda}_\mathbf{w}(k) = E\{\mathbf{w}(k)\mathbf{w}(k)^\top\}$) are then

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{\Lambda}_\mathbf{w} = \begin{bmatrix} \text{var}\{z_1\} & \text{cov}\{z_1, z_2\} & \text{cov}\{z_1, z_3\} \\ \text{cov}\{z_1, z_2\} & \text{var}\{z_2\} & \text{cov}\{z_2, z_3\} \\ \text{cov}\{z_1, z_3\} & \text{cov}\{z_2, z_3\} & \text{var}\{z_3\} \end{bmatrix}.$$

The variance and covariance terms are obtained from Eqs. (38) and (39). Unlike $\mathbf{\Lambda}_\mathbf{g}$, $\mathbf{\Lambda}_\mathbf{w}$ changes over time and must be recomputed at each iteration.

The process noise covariance matrix $\mathbf{\Lambda}_\mathbf{g}$ is calculated by considering potential errors in the model and determining how they propagate through to the state vector. The velocities $(\dot{s}, \dot{\phi}, \dot{\theta})$ will not be exactly zero, so they are modelled as Gaussian, zero-mean, random variables (r_s, r_ϕ, r_θ) with variances (q_s, q_ϕ, q_θ)(i.e., $E\{r_s\} = 0$ and $\text{var}\{r_s\} = q_s$). This noise is assumed uncorrelated over time (e.g., $E\{r_s(i)r_s(j)\} = \delta_{ij}q_s$), giving

$$\mathbf{g} = \Delta t \begin{bmatrix} r_s \\ r_\phi \\ r_\theta \end{bmatrix}, \quad \mathbf{\Lambda}_\mathbf{g} = \Delta t^2 \begin{bmatrix} q_s & 0 & 0 \\ 0 & q_\phi & 0 \\ 0 & 0 & q_\theta \end{bmatrix},$$

with Δt the time period between successive observations.

The filter is initialised by choosing a starting estimate for $\hat{\mathbf{y}}(0|0)$ and its variance $\mathbf{P}(0|0)$. The solution from the first frame-pair $\{I, I'\}$ gives the initial state $\hat{\mathbf{y}}(0 \mid 0)$ and we set $\mathbf{P}(0 \mid 0) = 25\mathbf{\Lambda}_\mathbf{g}$. The effects of these initial estimates diminish with time, though a good initial estimate improves convergence (Durrant-Whyte 1993). Since $\mathbf{\Lambda}_\mathbf{w}$ changes over time, the filter doesn't settle to a steady state. The variances of the velocity perturbation variables (q_s, q_ϕ, q_θ) are determined empirically.

The above Kalman filter is a simple linear filter with no dynamics, directly observing the parameters it estimates. An alternative approach would be to observe the epipolar geometry parameters (a, b, c, d), which provide three non-linear constraints on s, ϕ and θ (after rewriting Eq. (36)). The use of an *extended* Kalman filter to estimate these motion parameters would simplify the noise models, shifting the non-linearity from the observations to the filter itself. Performance comparisons between these two schemes are the subject of current investigation.

4.8 *Results*

Figure 20(b) shows the filtered results of the example from Section 4.6, along with the 95% confidence intervals obtained from the Kalman filter variances. The solution is clearly smoother after filtering and the variances decrease as the filter becomes more confident in its predictions. Note the increase in variances at the motion discontinuity; at that stage, the filter detects that the observation falls outside the validation region and re-initialises itself.

Figure 21 shows a real sequence with a car rotating on a turn-table about a (known) fixed axis. The successive two-frame estimates of the projected axis angle along with the computed errors serve as the filter input, and these are shown alongside the filter output, which gives reliable estimates of the projected axis. Figure 22 shows a second example with a subject shaking his head. Here the true axis is unknown, but it is approximately vertical and the results are qualitatively correct. Finally, results are shown for the images and

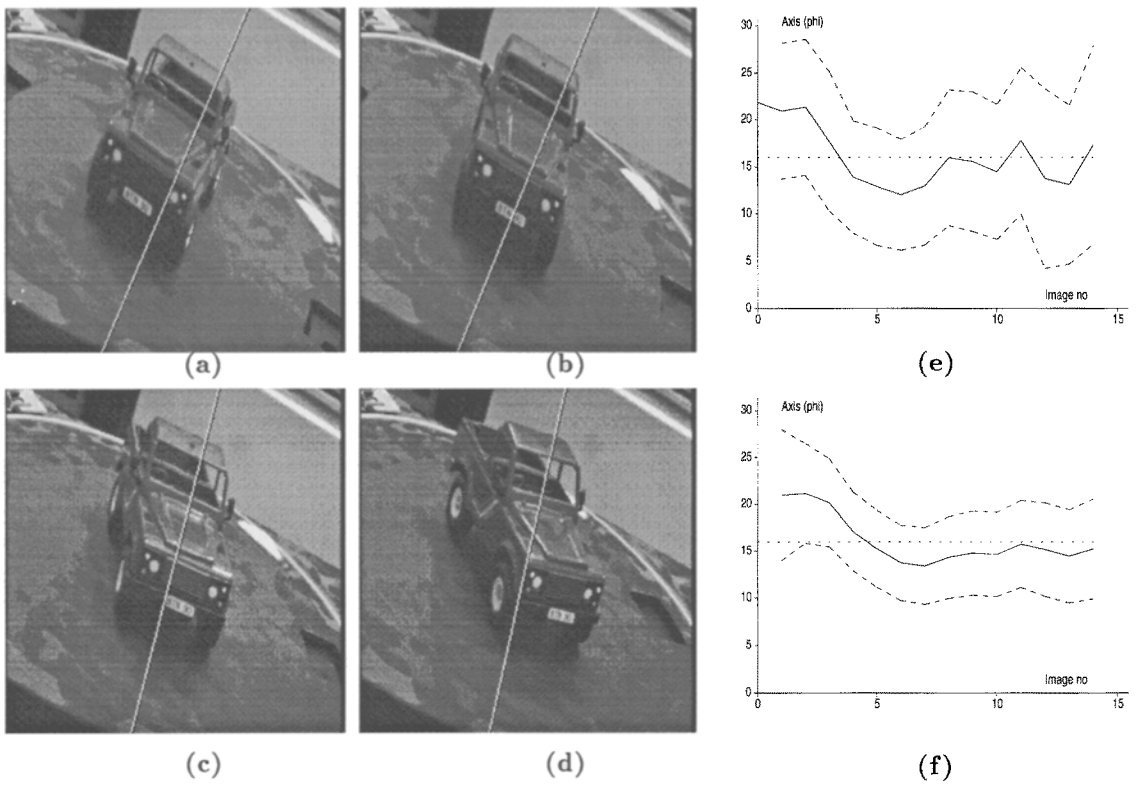

Fig. 21. A buggy rotates on a turn-table angled at 16° off the vertical in the image (true world angle approximately 24° with aspect ratio 0.65): (a)–(d) frames 2, 4, 6, 8 (6° turns between successive frames), with axes drawn through the image centre; (e) unfiltered two-frame estimates of the axis angle. Solid lines show computed values, dotted lines show true values, and dashed lines show 95% confidence intervals; (f) filtered estimates with associated 95% confidence regions.

corner data of Harris (1990) (Figs. 23 and 24). There is no scale change between views, and the fiducial axis is 10°off the vertical. The unfiltered solution (using E_3) is identical to Harris' (using E_2) since the scale here is unity (see Section 3.4). However, here the error model facilitates the filtering operation, whose estimates are clearly superior.

4.9 Previous Work

Harris (1990) used a weak perspective camera and the Euler angle representation to solve for rotation angles over two frames. The weak perspective form of E_H (Eq. (29)), whose shortcomings were outlined in Section 3.4.3, was minimised and shown to be *independent* of the turn angle η (see Appendix D.2), illustrating the bas-relief ambiguity. No confidence estimates in the solution were provided, and only the scale and projected axis were interpreted (not the cyclotorsion angle).

Fig. 22. The subject shakes his head. The true axis is roughly vertical, giving qualitatively correct results.

Huang and Lee (1989) assumed orthographic projection and proposed a linear algorithm to solve the equation $R_{23}\Delta x' - R_{13}\Delta y' + R_{32}\Delta x - R_{31}\Delta y = 0$ (a special case of Eq. (33)). Hu and Ahuja (1991) rightly

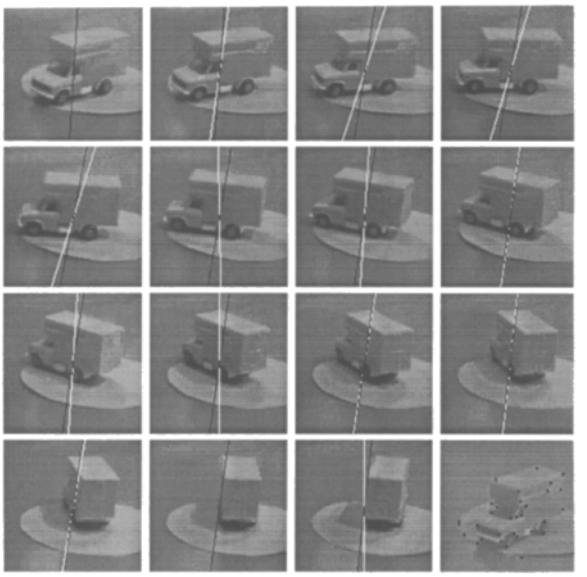

Fig. 23. The Harris truck sequence (Harris 1990) showing the unfiltered and filtered axes (white and black respectively). A typical set of corners is also shown.

criticised their approach, noting that the equation has only *two* independent unknowns, since

$$R_{13}^2 + R_{23}^2 = R_{31}^2 + R_{32}^2 = 1 - R_{33}^2. \qquad (46)$$

Our formulation (see Eq. 33) has *three* independent unknowns (the scale factor s is also accounted for), making a linear solution valid. Chen and Huang (1991) improved on the Huang-Lee solution by incorporating Eq. (46) as a constraint in their minimisation expression; however, this yielded a non-linear equation. None of the above authors noted that the projections of the axis of rotation could be found directly from R_{13}, R_{23}, R_{31} and R_{32}.

Huang and Lee (1989) deduced that two views yield a one-parameter family of motion (and structure) solutions, since R_{13}, R_{23}, R_{31} and R_{32} could only be recovered up to a scale factor. Bennet et al. (1989) proved the existence of this one-parameter family of rigid interpretations using set theory.

In an algorithm involving numerous computational stages, Koenderink and van Doorn (1991) determined the scale factor and the projections of the axes of rotation by means of a LCF (four non-coplanar points). Our algorithm retains their underlying principles, but uses a single set of equations and *all* available points. Lee and Huang (1990) independently devised Koenderink and van Doorn's method. They termed $(R_{13}, R_{23})^\top$ and $(R_{31}, R_{32})^\top$ the "matching directions", denoting them $c\mathbf{l}_1$ and $c\mathbf{l}_2$ respectively (where $\mathbf{l}_1$ and $\mathbf{l}_2$ were unit vectors). It is clear from our notation that $c = \sin\rho$, $\mathbf{l}_1 = (\sin\phi, -\cos\phi)^\top$ and $\mathbf{l}_2 = (\sin(\phi - \theta), -\cos(\phi - \theta))^\top$, corresponding to the epipolar directions.

Aloimonos and Bandyopadhay (1985) claimed that two orthographic projections of four rigidly linked non-coplanar points admitted only *four* interpretations of structure (up to a reflection); this is incorrect, for there are infinitely many solutions. Their error arises from the fact that their three polynomial equations in the relative depths are not independent (as claimed); the Jacobian of their system, used to determine whether there are a finite number of solutions (by the inverse function theorem), actually has zero determinant.

5 Rigid Motion: Three Views

As is well known, three distinct views of four rigidly connected non-coplanar point features are sufficient to extract uniquely the 3D motion and structure of the object up to a Necker reversal (Ullman 1979). That is, the introduction of a distinct third view, I'', removes the bas-relief ambiguity (Huang and Lee 1989; Harris 1990; Koenderink and van Doorn 1991; Ullman and Basri 1991), enabling the computation of the remaining rotation parameter ρ for each pair of views. Two main approaches to this three-view motion computation may be distinguished:

1. Use all three views and solve for all motion parameters simultaneously.
2. Compute partial solutions from the two-view pairs and based on these partial solutions, solve for the remaining parameters (ρ) from three views.

This section reviews the first approach and then presents a new method using the second approach. The resulting algorithm employs the *KvD* representation to solve for ρ directly using n points, and doesn't fail on special case motions (e.g., rotation about a fixed axis).

5.1 *Direct Three-View Approach*

A popular method of computing rotation parameters from three-views is to use the rotation matrix constraint (Huang and Lee 1989; Hu and Ahuja 1991; Longuet-Higgins 1991): the rotation $\mathbf{R}_{13}$ from I to I'' must be a composite of the two rotations from I to I' ($\mathbf{R}_{12}$) and

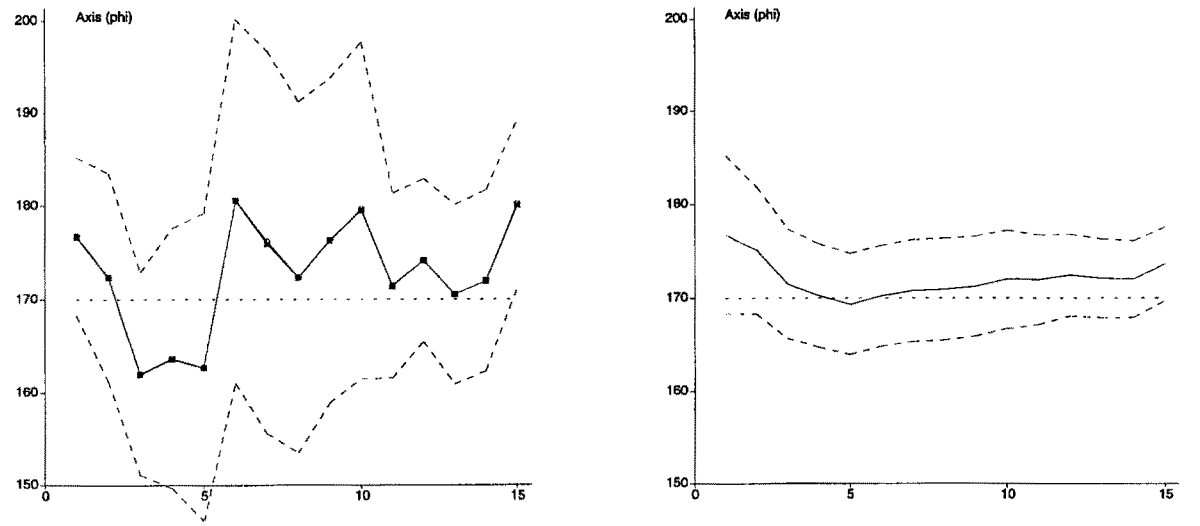

Fig. 24. The Harris truck sequence: solid lines give computed values, dotted lines show "true" values, and dashed lines show 95% confidence intervals: (a) our two-frame solution (E_3, circles) is indistinguishable from that of Harris (E_2, crosses); (b) the improved filtered values (dashed lines show 95% confidence levels).

I' to I'' ($\mathbf{R}_{23}$). That is, $\mathbf{R}_{13} = \mathbf{R}_{23}\mathbf{R}_{12}$, and these 9 equations yield a solution for the angles. The "border elements" (R_{13}, R_{23}, R_{31} and R_{32}), in particular, are frequently used. This approach is flawed in two ways. Firstly, rotation matrices are over-determined, using 9 elements to encode three independent angles, so inconsistencies can arise. Secondly, the constraint fails in quite typical special cases. For instance, Huang and Lee (1989) noted that their algorithm failed when $S_{32}R_{13} - S_{31}R_{23} = 0$ (where $\mathbf{R} = \mathbf{R}_{12} = [R_{ij}]$ and $\mathbf{S} = \mathbf{R}_{23} = [S_{ij}]$), though they didn't interpret what this meant. The *KvD* form of this condition is

$$\sin\rho_{12}\sin\rho_{23}\sin(\phi_{12} - (\phi_{23} - \theta_{23})) = 0,$$

where the subscripts indicate the appropriate view-pair. Thus, apart from the understandable difficulty when either ρ_{12} or ρ_{23} is zero, their algorithm also fails when $\phi_{12} = \phi_{23} - \theta_{23}$, i.e., when the two projections of the axis in I' coincide (cf. Fig. 17(b)). This occurs *whenever the 3D axis of rotation is fixed over the three views.* Since fixed-axis rotation occurs frequently in practice, a motion estimation algorithm should be able to cope with this.

A novel batch-solution approach was proposed by Tomasi and Kanade (1992). This involved minimising a three-view analogue of the cost function in Eq. (28),

$$\begin{aligned} E^3_{\mathrm{TK}} &= \sum_{i=0}^{n-1} |\mathbf{x}_i - \mathbf{M}\mathbf{X}_i - \mathbf{t}|^2 + \sum_{i=0}^{n-1} |\mathbf{x}'_i - \mathbf{M}'\mathbf{X}_i - \mathbf{t}'|^2 \\ &\quad + \sum_{i=0}^{n-1} |\mathbf{x}''_i - \mathbf{M}''\mathbf{X}_i - \mathbf{t}''|^2 \\ &= \sum_{i=0}^{n-1} |\Delta\mathbf{x}_i - \mathbf{M}\Delta\mathbf{X}_i|^2 + \sum_{i=0}^{n-1} |\Delta\mathbf{x}'_i - \mathbf{M}'\Delta\mathbf{X}_i|^2 \\ &\quad + \sum_{i=0}^{n-1} |\Delta\mathbf{x}''_i - \mathbf{M}''\Delta\mathbf{X}_i|^2 \end{aligned} \tag{47}$$

where Δ indicates registration with respect to the data centroids $\{\bar{\mathbf{x}}, \bar{\mathbf{x}}', \bar{\mathbf{x}}'', \bar{\mathbf{X}}\}$ and double-primes indicate quantities in I'', e.g. image coordinates in I'' are $\mathbf{x}'' = (x'', y'')^\top$. Employing the notation

$$\mathbf{V} = [\mathbf{v}_0\ \mathbf{v}_1\ \cdots\ \mathbf{v}_{n-1}] = \begin{bmatrix} \Delta\mathbf{x}_0 & \Delta\mathbf{x}_1 & \cdots & \Delta\mathbf{x}_{n-1} \\ \Delta\mathbf{x}'_0 & \Delta\mathbf{x}'_1 & \cdots & \Delta\mathbf{x}'_{n-1} \\ \Delta\mathbf{x}''_0 & \Delta\mathbf{x}''_1 & \cdots & \Delta\mathbf{x}''_{n-1} \end{bmatrix},$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{M} \\ \mathbf{M}' \\ \mathbf{M}'' \end{bmatrix} \quad \text{and} \quad \mathbf{S} = [\Delta\mathbf{X}_0\ \Delta\mathbf{X}_1\ \cdots\ \Delta\mathbf{X}_{n-1}],$$

with $\mathbf{V}$ the noisy $6 \times n$ 'measurement matrix', $\mathbf{L}$ the 6×3 'affine camera' matrix, and $\mathbf{S}$ the $3 \times n$'affine

structure' matrix, it follows that

$$E^3_{\mathrm{TK}} = \sum_{i=0}^{n-1} |\mathbf{v}_i - \mathbf{L}\Delta\mathbf{X}_i|^2 = \mathrm{Trace}\,[(\mathbf{V}-\mathbf{LS})^\top(\mathbf{V}-\mathbf{LS})]. \tag{48}$$

The explicit use of affine cameras and affine structure generalises the original formulation of Tomasi and Kanade, who computed the rigid motion parameters in two stages: first, they employed singular value decomposition to factor $\mathbf{V}$ into the rank-three product $\mathbf{LS}$; and second, they imposed the rigidity constraint to orthogonalise $\mathbf{L}$ appropriately. Since the decomposition into $\mathbf{LS}$ is not unique (any non-singular 3×3 matrix $\mathbf{A}$ satisfies $\mathbf{LS} = (\mathbf{LA})(\mathbf{A}^{-1}\mathbf{S})$), it was necessary to solve for $\mathbf{A}$. This involved non-linear "metric constraints" on $\mathbf{LA}$, requiring each successive pair of rows to belong to a rotation matrix[14], for instance $\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2 = 0$ and $\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_1 = \mathbf{M}_2^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2 = 1$. Having determined $\mathbf{A}$, the rotation matrices can be computed, and hence the rotation angles and scales recovered. Ullman and Basri (1991) used a similar approach. The general disadvantage of this "metric constraint" approach is that a system of simultaneous quadratic equations must be solved.

Weinshall and Tomasi (1993) improved the orthogonalisation stage, avoiding the aforementioned non-linear equations by solving for the Gramian matrix $\mathbf{A}\mathbf{A}^\top$ (a linear solution). This symmetric matrix was then decomposed into $\mathbf{A}$ using Cholesky decomposition. However, their algorithm still relied on the explicit choice of a local coordinate frame, which is unnecessary; for instance, McLauchlan et al. (1994) solved for the Gramian using *all* available points.

5.2 *Partial Solution Approach*

In contrast to the above methods, Koenderink and van Doorn (1991) used the *partial* solutions obtained from two view-pairs to compute the remaining unknown parameters in the three-view case. Specifically, they used three points from views I and I' to give two constraints of the form $A_i = B_i \cos\rho_{12} + \Delta z_i \sin\rho_{12}$ $(i = 1, 2)$, with $\{A_i, B_i\}$ known and $\{\Delta z_i, \rho_{12}\}$ unknown, and then eliminated ρ_{12} to give a single quadratic equation in Δz_1 and Δz_2. A similar process for views I and I'' yielded a second quadratic equation in Δz_1 and Δz_2. Intersecting these quadratics gave four possible solutions for the two depths, from which ρ_{12}, ρ_{13} and the remaining depths Δz_i $(i = 3 \cdots n-1)$ were determined. This algorithm has the drawbacks of using a minimal point-set and of computing the rotation angles via the structure Δz_i. The remainder of this section extends their approach to n points and contrasts performance with that of the direct three-view method. Moreover, the rotation angles $\{\rho_{12}, \rho_{13}\}$ are isolated directly, without having to first compute depth.

Consider minimising E^3_{TK} over structure, $\Delta\mathbf{X}_i$, giving $\Delta\mathbf{X}_i = (\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{v}_i$ (cf. Appendix C). Then

$$E^3_{\mathrm{TK}} = \sum_{i=0}^{n-1} |\mathbf{v}_i - \mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{L}^\top\mathbf{v}_i|^2, \tag{49}$$

leaving E^3_{TK} a function only of the motion parameters $\{s_{12}, \theta_{12}, \phi_{12}, \rho_{12}, s_{13}, \theta_{13}, \phi_{13}, \rho_{13}\}$. (The three-view motion parameters are again appropriately subscripted; for instance, $s_{13} = Z^c_{\mathrm{ave}}/Z^{c''}_{\mathrm{ave}}$. As in Section 4.2, the necessary ratios of focal lengths and aspect ratios are assumed known.) Performing the two-view analysis of Section 4 on views $\{I, I'\}$ and $\{I, I''\}$ yields the respective motion parameters $\{s, \theta, \phi\}$, and it then remains only to determine the two turn angles ρ_{12} and ρ_{13}, whence structure can be computed.

Now, in the case of rigid motion (cf. Section 4.2), each $\mathbf{M}$ matrix comprises the first two rows $\{\mathbf{R}_1^\top, \mathbf{R}_2^\top\}$ of a scaled rotation matrix $\mathbf{R}$. Using the *KvD* form of a rotation matrix (cf. Appendix D), with s, θ and ϕ known from the two-view computations, $\mathbf{M}$ is only a function of the unknown parameter ρ,

$$\mathbf{M}(\rho) = s\begin{bmatrix}\mathbf{R}_1^\top(\rho)\\ \mathbf{R}_2^\top(\rho)\end{bmatrix} = s[\,\mathbf{D} + \cos\rho\,\mathbf{E} \mid \sin\rho\,\Phi^\perp\,],$$

where the 2×2 matrices $\mathbf{D}$ and $\mathbf{E}$ are known:

$$\mathbf{D}(\theta, \phi) = \begin{bmatrix}\cos\phi\\ \sin\phi\end{bmatrix}[\cos(\phi-\theta)\ \ \sin(\phi-\theta)]$$

$$\text{and}\quad \mathbf{E}(\theta, \phi) = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} - \mathbf{D}.$$

Similar expressions obtain for $\mathbf{M}'$ and $\mathbf{M}''$. Without loss of generality, the rotation matrix defining $\mathbf{M}$ can be set to $\mathbf{I}$ and its associated scale factor s can be set to unity[15], giving

$$\mathbf{L}(\rho_{12}, \rho_{13}) = \begin{bmatrix}\mathbf{M}\\ \mathbf{M}'\\ \mathbf{M}''\end{bmatrix}$$

$$= \left[\begin{array}{cc|c} 1 & 0 & 0\\ 0 & 1 & 0\\ s_{12}(\mathbf{D}_{12} + \cos\rho_{12}\,\mathbf{E}_{12}) & & s_{12}\sin\rho_{12}\,\Phi^\perp_{12}\\ s_{13}(\mathbf{D}_{13} + \cos\rho_{13}\,\mathbf{E}_{13}) & & s_{13}\sin\rho_{13}\,\Phi^\perp_{13}\end{array}\right], \tag{50}$$

which incorporates the partial solutions into $\mathbf{L}$ in Eq. (49). Then $E^3_{\mathrm{TK}}|_{s_{12},\theta_{12},\phi_{12},s_{13},\theta_{13},\phi_{13}}$ is only a function of ρ_{12} and ρ_{13}. Thus, rather than performing

unconstrained minimisation on E^3_{TK} (Eq. (48)) and later imposing the rigidity constraint (the Tomasi and Kanade approach), *constrained* minimisation is used to determine ρ directly given the partial solutions. The resulting expression is non-linear and has no apparent closed-form solution; a numerical minimisation technique (the Nelder-Mead downhill simplex method) is therefore employed to determine ρ_{12} and ρ_{13}.

It can be shown from Eq. (49) that $E^3_{\mathrm{TK}}(\rho_{12}, \rho_{23}) = E^3_{\mathrm{TK}}(-\rho_{12}, -\rho_{23})$; this is the well-known *Necker reversal* discussed in Section 4.1. There is also naturally a "depth-scale" ambiguity in the recovered structure $\Delta\mathbf{X}_i$. The final algorithm is summarised below:

Task

Given three distinct views of 4 or more points (with at least 4 non-coplanar points common to all views) and the partial solutions $\{s, \theta, \phi\}$ for the view-pairs $\{I, I'\}$ and $\{I, I''\}$, determine the turn angles ρ_{12} and ρ_{13}.

Algorithm

1. Form the matrices and vectors

$$\mathbf{D}_{12} = \begin{bmatrix} \cos\phi_{12} \\ \sin\phi_{12} \end{bmatrix} [\cos(\phi_{12} - \theta_{12}) \ \sin(\phi_{12} - \theta_{12})],$$

$$\mathbf{E}_{12} = \begin{bmatrix} \cos\theta_{12} & -\sin\theta_{12} \\ \sin\theta_{12} & \cos\theta_{12} \end{bmatrix} - \mathbf{D}_{12} \quad \text{and}$$

$$\Phi^{\perp}_{12} = \begin{bmatrix} \sin\phi_{12} \\ -\cos\phi_{12} \end{bmatrix},$$

and similarly $\mathbf{D}_{13}$, $\mathbf{E}_{13}$ and $\Phi^{\perp}_{13}$.

2. Minimise the expression

$$E^3_{\mathrm{TK}} = \sum_{i=0}^{n-1} \left| \mathbf{v}_i - \mathbf{L}(\mathbf{L}^{\top}\mathbf{L})^{-1}\mathbf{L}^{\top}\mathbf{v}_i \right|^2$$

over ρ_{12} and ρ_{13}, where n is the number of points common to all three views and

$$\mathbf{L}(\rho_{12}, \rho_{13}) = \left[\begin{array}{ccc|c} 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ s_{12}(\mathbf{D}_{12} + \cos\rho_{12}\,\mathbf{E}_{12}) & & & s_{12}\sin\rho_{12}\,\Phi^{\perp}_{12} \\ s_{13}(\mathbf{D}_{13} + \cos\rho_{13}\,\mathbf{E}_{13}) & & & s_{13}\sin\rho_{13}\,\Phi^{\perp}_{13} \end{array}\right].$$

3. Recover the structure (up to an overall scale factor and Necker reversal):

$$\Delta\mathbf{X}_i = (\mathbf{L}^{\top}\mathbf{L})^{-1}\mathbf{L}^{\top}\mathbf{v}_i.$$

Three-view depth-turn algorithm.

5.3 Results

We now contrast the motion estimation performance of the partial solution approach with that of Tomasi and Kanade. Our implementation of the latter used an n-point weak-perspective Gramian matrix approach, combining the 4-point weak-perspective algorithm of Weinshall (1993) with the n-point pure orthographic algorithm of (McLauchlan et al. 1994). A three-view moving window is used over the sequence of frames and the quantity ρ_{23} (rather than ρ_{13})[16] displayed since it is easier to interpret in the following examples, which have constant turn angle.

First, a 20-frame synthetic example is shown with turn angle $\rho = 8°$ about a fixed axis ($\theta = 0°$ and $\phi = 25°$). Figure 25 graphs ρ_{12} and ρ_{23} computed from successive triplets of frames, where $\{I_{k-2}, I_{k-1}, I_k\}$ yields $\rho_{12}(k-1)$ and $\rho_{23}(k)$. *Two* estimates are thus obtained for each turn angle (excluding $\rho_{12}(1)$ and $\rho_{23}(20)$), since each turn angle appears in two successive frame triplets (i.e., $\rho_{12}(k) = \rho_{23}(k-1)$). The results in Figs. 25(a)(b) and (c)(d) were obtained with Gaussian noise of $\sigma = 0.3$ pixels and $\sigma = 0.6$ pixels respectively (on 200×200 images). It emerges that the direct 3-view method (b)(d) is superior to the constrained solution (a)(c), because the former can better distribute the errors amongst *all* the variables, while the latter has $\{s, \theta, \phi\}$ *fixed* by the two-view computations, so that all the remaining error must be absorbed in ρ alone. A similar effect is observed on real data sequences. For instance, Figs. 25(e)(f) shows the results for the Harris truck sequence from Fig. 23.

Figures 25(d) and (f) illustrate a potential failure-mode of the Gramian matrix approach: the algorithm fails if the computed Gramian $\mathbf{G}$ is not positive-definite. That is, after minimising

$$\min_{\{\mathbf{G}\}} E_G(\mathbf{M}) + E_G(\mathbf{M}') + E_G(\mathbf{M}'') \text{ subject to } \mathbf{G} = \mathbf{G}^{\top},$$

where $E_G(\mathbf{M}+) = (\mathbf{M}_1^{\top}\mathbf{G}\mathbf{M}_2)^2 + (\mathbf{M}_1^{\top}\mathbf{G}\mathbf{M}_1 - 1)^2 + (\mathbf{M}_2^{\top}\mathbf{G}\mathbf{M}_2 - 1)^2$, the symmetric solution $\mathbf{G}$ cannot necessarily be decomposed as $\mathbf{A}\mathbf{A}^{\top}$. This problem was noted by Weinshall and Tomasi (1993), whose explanation was that "the computation of depth is more sensitive to errors than the computation of the similarity invariant representation" (p. 682). In fact, the problem lies in their linear solution method, which exploits the property of positive-definiteness; this property holds in the noise-free case, but not necessarily in presence of noise. The method of Tomasi and Kanade does not suffer from this problem because, having computed

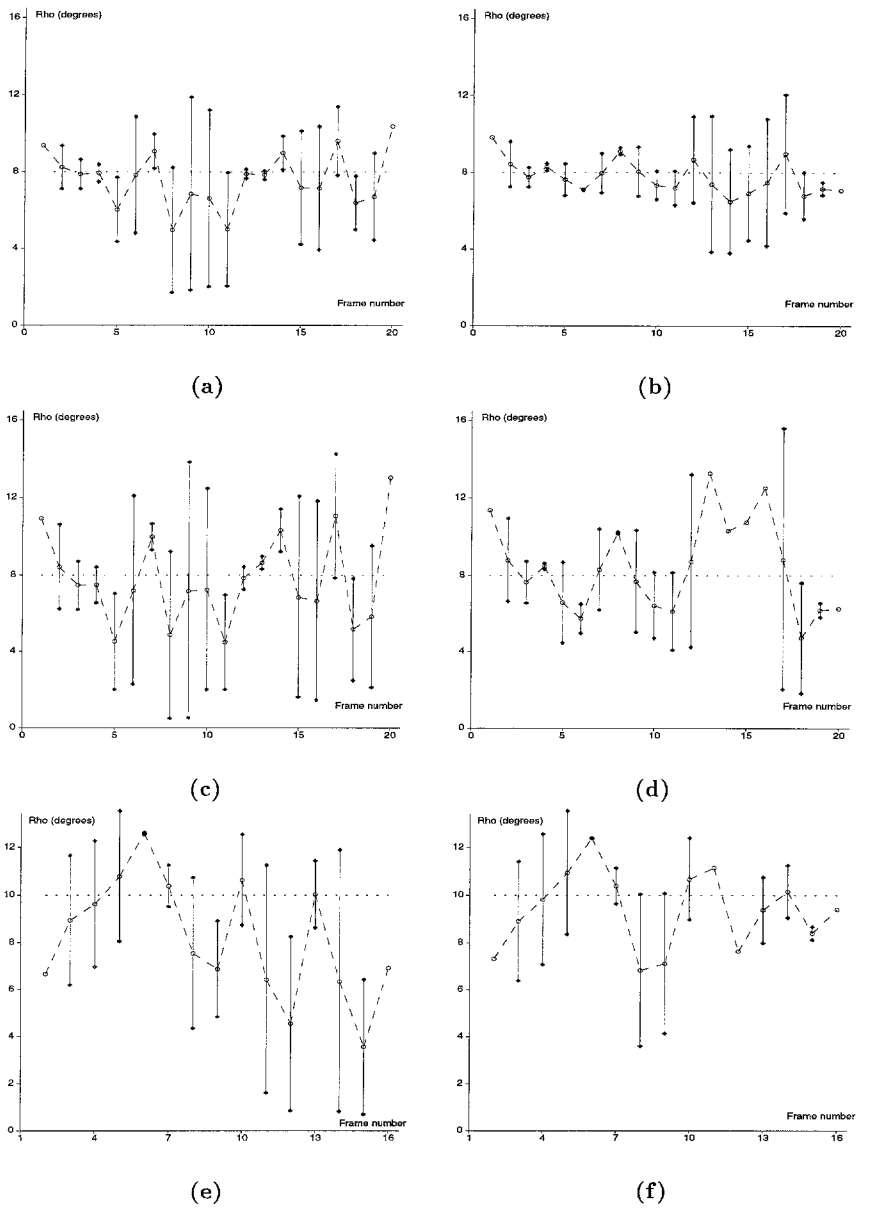

Fig. 25. Comparison between the ρ angles computed using our partial solution method [left] and the direct 3-view method of Tomasi and Kanade [right]. Where two estimates for each ρ angle are available (crosses), they are connected by a vertical line, and the average marked by a circle. The dotted line gives the correct solution and the dashed line connects the averages. Three-frame moving windows are used (see text for further details and interpretation): (a), (b) synthetic example (20 frames, $\sigma = 0.3$ pixels); (c), (d) same synthetic example ($\sigma = 0.6$ pixels); (e), (f) Harris truck sequence (16 frames, cf. Fig. 23).

$\{\mathbf{M}, \mathbf{M}', \mathbf{M}''\}$, the set of quadratic "metric constraints" on $\mathbf{A}$ are solved by minimising the cost function *directly* over $\mathbf{A}$. For instance, in the pure orthographic case[17],

$$\min_{\{\mathbf{G}\}} E_A(\mathbf{M}) + E_A(\mathbf{M}') + E_A(\mathbf{M}''),$$

where $E_A(\mathbf{M}) = (\mathbf{M}_1^\top \mathbf{A}\mathbf{A}^\top \mathbf{M}_2)^2 + (\mathbf{M}_1^\top \mathbf{A}\mathbf{A}^\top \mathbf{M}_1 - 1)^2 + (\mathbf{M}_2^\top \mathbf{A}\mathbf{A}^\top \mathbf{M}_2 - 1)^2$. This solution makes the extra decomposition unnecessary and a solution for $\mathbf{A}$ can be found, even in the presence of noise.

In summary, the direct three-view method returns better estimates for ρ than the partial-solution algorithm. Nonetheless, the latter approach does make two

important contributions. First, it provides sequential estimates of ρ, with the optional facility of incorporating a dynamic motion model (via filtering). This differs from the batch approach of Tomasi and Kanade, which cannot incorporate a mathematical model of the motion. Indeed, we expect the ρ estimates in Figs. 25(a), (c) and (e) to improve when filtered values of $\{s, \theta, \phi\}$ are used from the partial motion solutions. Second, it introduces the known rigidity constraints *directly* into the minimisation expression, rather than first performing unconstrained minimisation and imposing the rigidity constraints afterwards. This suggests an alternative direct three-view method, which imposes the rigidity constraints on E^3_{TK} from the start. That is, the $\mathbf{M}$ matrix in each view is constrained to have the weak perspective form $\mathbf{M}_{\mathrm{wp}}$. Although non-linear, this approach may well yield a more correct solution for the rotation parameters than that obtained via the affine cameras, an issue worthy of further theoretical and empirical research.

6 Conclusions

A new framework has been proposed for computing motion from point features viewed under parallel projection. Based on the affine camera and its epipolar geometry, this framework incorporates the major theoretical results pertaining to the problem (Ullman 1979; Bennett et al. 1989; Huang and Lee 1989; Harris 1990; Lee and Huang 90; Ullman and Basri 1991; Tomasi and Kanade 1992), including partial solutions, ambiguities and degeneracies. The necessary camera calibration parameters have been identified, and the facility to use all available points both ensures robustness to noise and obviates the need to choose a local coordinate frame. Error models provide confidence estimates in the computed parameters, and the processing of successive frame-pairs permits straight forward extension to long sequences in sequential mode. The rotation representation of Koenderink and van Doorn has been shown to be particularly apt since it makes explicit the ambiguities inherent in parallel projection (e.g., the bas-relief ambiguity).

Partial solutions are computed when complete ones are impossible; indeed, our algorithms *ascertain* when special cases have occurred (e.g., pure image motion) or when critical assumptions have been violated (e.g., perspective effects become significant), by considering matrix ranks. In short, we have shown that affine epipolar geometry provides a solid foundation for both understanding motion and devising reliable algorithms.

There are several interesting directions for future work. First, a framework for imposing a dynamic motion model on the rotation parameters would be useful. It would be advisable to pose this evolution model in a natural physical rotation frame, namely *angle-axis* form rather than *KvD* form, to enable more realistic inertia constraints (e.g., constant angular velocity about the space axis). Second, it would be interesting to study the transition between the affine fundamental matrix $\mathbf{F}_A$ and the projective fundamental matrix $\mathbf{F}$, to see how the results might extend to perspective projection (perhaps using similar "bootstrapping" techniques to those used in (Longuet-Higgins 1991; Faugeras et al. 1992; Luong et al. 1993)).

Appendix

A *The Parallel Epipolar Fundamental Matrix*

The form of a fundamental matrix with both epipoles at infinity was derived by Torr (1993); it is presented below in a slightly different form. The fundamental matrix $\mathbf{F}$ is a 3×3 homogeneous matrix obeying $\mathbf{Fe} = \mathbf{0}$ and $\mathbf{F}^\top\mathbf{e}' = \mathbf{0}$, where $\mathbf{e}$ and $\mathbf{e}'$ are the epipoles in I and I' respectively (in homogeneous coordinates). It is well-known that $\mathbf{F}$ has seven degrees of freedom, since it is homogeneous and its rank equals 2.

The requirement for the epipoles to lie at infinity ($\mathbf{e} = (e_x, e_y, 0)^\top$ and $\mathbf{e}' = (e'_x, e'_y, 0)^\top$) introduces two new constraints on $\mathbf{F}$, giving a five degree-of-freedom matrix which can be parameterised as

$$\mathbf{F}_{\|} = \begin{bmatrix} e_y e'_y \alpha & -e_x e'_y \alpha & e_y e'_y \gamma \\ -e_y e'_x \alpha & e_x e'_x \alpha & -e_y e'_x \gamma \\ e_y e'_y \beta & -e_x e'_y \beta & e_y e'_y \delta \end{bmatrix},$$

where α, β, γ and δ are scalar parameters. The five degrees of freedom are $e_x : e_y$, $e'_x : e'_y$ and the three ratios $\alpha : \beta : \gamma : \delta$. When $\alpha = 0$, $\mathbf{F}_{\|}$ becomes $\mathbf{F}_A$ (see Eq. (23)), which has four degrees of freedom. This parameter α can be shown to tend towards zero as the camera focal lengths tend towards infinity.

B *Epipolar Geometry Minimisation*

Consider the optimisation problem from Section 3.2:

$$E_3(\mathbf{n}, e) = \min_{\{\mathbf{n},e\}} \frac{1}{|\mathbf{n}|^2} \sum_{i=0}^{n-1} \left(\mathbf{n}^\top \mathbf{r}_i + e\right)^2.$$

Using a Lagrange multiplier λ, we get

$$E' = \sum_{i=0}^{n-1} \left(\mathbf{n}^\top \mathbf{r}_i + e\right)^2 - \lambda(\mathbf{n}^\top \mathbf{n} - 1).$$

We solve by setting the partial derivatives of the Lagrangian to zero. Differentiating with respect to e gives

$$\frac{\partial E'}{\partial e} = 0 = 2\sum_{i=0}^{n-1} \left(\mathbf{n}^\top \mathbf{r}_i + e\right)$$

$$\Rightarrow e = -\frac{1}{n}\sum_{i=0}^{n-1} \left(\mathbf{n}^\top \mathbf{r}_i\right) = -\mathbf{n}^\top \bar{\mathbf{r}}.$$

The optimum solution $\mathbf{n}$ thus passes through the data centroid $\bar{\mathbf{r}}$, and substituting gives

$$E' = \sum_{i=0}^{n-1} \left(\mathbf{n}^\top \mathbf{v}_i\right)^2 - \lambda(\mathbf{n}^\top \mathbf{n} - 1). \qquad (51)$$

Writing $\mathbf{W} = \sum_i \mathbf{v}_i \mathbf{v}_i^\top$ yields

$$E' = \mathbf{n}^\top \mathbf{W} \mathbf{n} - \lambda(\mathbf{n}^\top \mathbf{n} - 1),$$

and differentiating with respect to $\mathbf{n}$ gives

$$\frac{\partial E'}{\partial \mathbf{n}} = \mathbf{0} = 2\mathbf{W}\mathbf{n} - 2\lambda\mathbf{n} \Rightarrow \mathbf{W}\mathbf{n} = \lambda\mathbf{n} \qquad (52)$$

Thus, $\mathbf{n}$ is a unit eigenvector of $\mathbf{W}$ corresponding to the eigenvalue λ. To decide *which* eigenvalue, we substitute into E_3,

$$E_{3,\min} = \mathbf{n}^\top \mathbf{W} \mathbf{n} = \mathbf{n}^\top \lambda \mathbf{n} = \lambda |\mathbf{n}|^2 = \lambda,$$

showing that λ is the *minimum* eigenvalue of $\mathbf{W}$ (and $\mathbf{n}$ its associated eigenvector).

C *Relation to the Tomasi & Kanade Cost Function*

This section establishes that the point-to-point cost function from Section 3.4.3,

$$E_{\mathrm{TK}}(\mathbf{M}, \mathbf{M}', \mathbf{t}, \mathbf{t}', \mathbf{X}_i) = \sum_{i=0}^{n-1} |\mathbf{x}_i - \mathbf{M}\mathbf{X}_i - \mathbf{t}|^2 + \sum_{i=0}^{n-1} |\mathbf{x}'_i - \mathbf{M}'\mathbf{X}_i - \mathbf{t}'|^2,$$

reduces to the cost function in Section 3.4,

$$E_3(\mathbf{n}, e) = \frac{1}{|\mathbf{n}|^2} \sum_{i=0}^{n-1} \left(\mathbf{r}_i^\top \mathbf{n} + e\right)^2,$$

after minimisation with respect to structure $\mathbf{X}_i$. From Eq. (51), and using the $\mathbf{v}$ notation of Section 3.4.1, the above form of E_3 is equivalent to

$$E_3(\mathbf{n}) = \frac{1}{|\mathbf{n}|^2} \sum_{i=0}^{n-1} \left(\mathbf{v}_i^\top \mathbf{n}\right)^2,$$

which is the form which will be derived here.

The equations $\partial E_{\mathrm{TK}}/\partial \mathbf{t} = \partial E_{\mathrm{TK}}/\partial \mathbf{t}' = \mathbf{0}$ give

$$\mathbf{t} = \bar{\mathbf{x}} - \mathbf{M}\bar{\mathbf{X}} \quad \text{and} \quad \mathbf{t}' = \bar{\mathbf{x}}' - \mathbf{M}'\bar{\mathbf{X}},$$

where $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}'$ and $\bar{\mathbf{X}}$ are the centroids respectively in I, I' and the scene. Substitution back into E_{TK} yields the expression

$$E_{\mathrm{TK}}(\mathbf{M}, \mathbf{M}', \mathbf{X}_i) = \sum_{i=0}^{n-1} |\Delta\mathbf{x}_i - \mathbf{M}\Delta\mathbf{X}_i|^2 + \sum_{i=0}^{n-1} |\Delta\mathbf{x}'_i - \mathbf{M}'\Delta\mathbf{X}_i|^2,$$

where $\Delta\mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, $\Delta\mathbf{x}'_i = \mathbf{x}'_i - \bar{\mathbf{x}}'$ and $\Delta\mathbf{X}_i = \mathbf{X}_i - \bar{\mathbf{X}}$. Thus, E_{TK} becomes

$$E_{\mathrm{TK}}(\mathbf{L}, \Delta\mathbf{X}_i) = \sum_{i=0}^{n-1} |\mathbf{v}_i - \mathbf{L}\Delta\mathbf{X}_i|^2, \quad \text{where} \quad \mathbf{L} = \begin{bmatrix} \mathbf{M}' \\ \mathbf{M} \end{bmatrix}.$$

Differentiating E_{TK} with respect to $\Delta\mathbf{X}_i$ yields

$$\partial E_{\mathrm{TK}}/\partial \Delta\mathbf{X}_i = -2\mathbf{L}^\top \mathbf{v}_i + 2\mathbf{L}^\top \mathbf{L}\Delta\mathbf{X}_i = \mathbf{0}$$
$$\Rightarrow \Delta\mathbf{X}_i = (\mathbf{L}^\top \mathbf{L})^{-1}\mathbf{L}^\top \mathbf{v}_i,$$

and substituting back into E_{TK} gives

$$E_{\mathrm{TK}}(\mathbf{L}) = \sum_{i=0}^{n-1} \left|\mathbf{v}_i - \mathbf{L}(\mathbf{L}^\top \mathbf{L})^{-1}\mathbf{L}^\top \mathbf{v}_i\right|^2 = \sum_{i=0}^{n-1} |(\mathbf{I} - \mathbf{P}_L)\mathbf{v}_i|^2.$$

Here, $\mathbf{I}$ is the 4×4 identity matrix and $\mathbf{P}_L$ the projection matrix $\mathbf{L}(\mathbf{L}^\top \mathbf{L})^{-1}\mathbf{L}^\top$, with $\mathbf{P}_L \mathbf{v}_i$ giving the component of $\mathbf{v}_i$ in the column space of $\mathbf{L}$. (A projection matrix satisfies $\mathbf{P}_L^2 = \mathbf{P}_L$ and $\mathbf{P}_L^\top = \mathbf{P}_L$.) Similarly, $(\mathbf{I} - \mathbf{P}_L)$ is also a projection matrix (Strang 1988), projecting $\mathbf{v}_i$ onto the orthogonal complement, namely the left null space of $\mathbf{L}$ (which is orthogonal to its column space).

Now, the vector $\mathbf{n} = (a, b, c, d)^\top$ spans the one-dimensional left null-space of $\mathbf{L}$, i.e. $\mathbf{L}^\top \mathbf{n} = \mathbf{0}$. This

fact follows from the definitions of Sections 3.2 and 3.3 (and the relations $\Gamma = \mathbf{B}'\mathbf{B}^{-1}$, $\mathbf{d} = \mathbf{b}' - \Gamma\mathbf{b}$ and $\mathbf{d}^\top\mathbf{d}^\perp = 0$), which give

$$\mathbf{L}^\top\mathbf{n} = [\mathbf{M}'^\top\ \mathbf{M}^\top]\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \mathbf{B}'^\top & \mathbf{B}^\top \\ \mathbf{b}'^\top & \mathbf{b}^\top \end{bmatrix}\begin{bmatrix} \mathbf{d}^\perp \\ -\Gamma^\top\mathbf{d}^\perp \end{bmatrix} = \mathbf{0}.$$

Combining the above two results gives the relation

$$\mathbf{I} - \mathbf{P}_L = \frac{\mathbf{n}\mathbf{n}^\top}{|\mathbf{n}|^2},$$

whence

$$E_{\mathrm{TK}}(\mathbf{n}) = \sum_{i=0}^{n-1}\left(\frac{\mathbf{v}_i^\top\mathbf{n}}{|\mathbf{n}|}\right)^2,$$

which is E_3 □.

D Rotation Matrices

This appendix describes three rotation representations and demonstrates the relationships between them.

D.1 Angle-Axis Form. Consider a rotation about the unit axis $\mathbf{a} = (a_x, a_y, a_z)^\top$ through angle χ. The Rodrigues equation gives the resulting rotation matrix (e.g., Kanatani 1993):

$$\begin{bmatrix} a_x^2(1-\cos\chi)+\cos\chi & a_xa_y(1-\cos\chi)-a_z\sin\chi & a_xa_z(1-\cos\chi)+a_y\sin\chi \\ a_ya_x(1-\cos\chi)+a_z\sin\chi & a_y^2(1-\cos\chi)+\cos\chi & a_ya_z(1-\cos\chi)-a_x\sin\chi \\ a_za_x(1-\cos\chi)-a_y\sin\chi & a_za_y(1-\cos\chi)+a_x\sin\chi & a_z^2(1-\cos\chi)+\cos\chi \end{bmatrix} \quad (53)$$

The angle-axis formulation is identical to the *quaternion* representation. The unit quaternion $\mathbf{u} = (s_u, \mathbf{v}_u)$ comprises a scalar component $s_u = \cos(\chi/2)$ and a vector component $\mathbf{v}_u = (x_u, y_u, z_u) = \sin(\chi/2)\mathbf{a}$. The corresponding rotation matrix is (Kanatani 1993)

$$\mathbf{R} = \begin{bmatrix} s_u^2+x_u^2-y_u^2-z_u^2 & 2(x_uy_u - s_uz_u) & 2(x_uz_u + s_uy_u) \\ 2(y_ux_u + s_uz_u) & s_u^2-x_u^2+y_u^2-z_u^2 & 2(y_uz_u - s_ux_u) \\ 2(z_ux_u - s_uy_u) & 2(z_uy_u + s_ux_u) & s_u^2-x_u^2-y_u^2+z_u^2 \end{bmatrix}.$$

This is the same as the previous matrix after substituting the half-angle formulae ($1 + \cos\chi = 2\cos^2(\chi/2)$ and $1 - \cos\chi = 2\sin^2(\chi/2)$).

D.2 Euler-Angle Form. The rotation is decomposed into three simpler rotations, first about the Z axis (by φ), then about the X axis (by η), and finally about the Z axis again (by μ):

$$\mathbf{R} = \begin{bmatrix} \cos\mu & -\sin\mu & 0 \\ \sin\mu & \cos\mu & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\eta & -\sin\eta \\ 0 & \sin\eta & \cos\eta \end{bmatrix} \times \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (54)$$

D.3 KvD Form. Koenderink and van Doorn (1991) decomposed the rotation into two parts: a rotation *in* the image plane followed by a rotation *out of* the image plane. This is expressed as $\mathbf{R} = \mathbf{R}_\rho\mathbf{R}_\theta$ (cf. Eq. (34)), where $\mathbf{R}_\theta$ is a rotation by θ about the Z-axis and $\mathbf{R}_\rho$ is a rotation by ρ about the unit axis Φ (which lies parallel to the X-Y plane, angled at ϕ to the positive X axis). These matrices are obtained from Eq. (53) with $\mathbf{a}_\rho = (\cos\phi, \sin\phi, 0)^\top$ and $\mathbf{a}_\theta = (0, 0, 1)^\top$ respectively:

$$\mathbf{R} = \begin{bmatrix} (1-\cos\rho)\cos\phi\cos(\phi-\theta) + \cos\rho\cos\theta & (1-\cos\rho)\cos\phi\sin(\phi-\theta) - \cos\rho\sin\theta & \sin\phi\sin\rho \\ (1-\cos\rho)\sin\phi\cos(\phi-\theta) + \cos\rho\sin\theta & (1-\cos\rho)\sin\phi\sin(\phi-\theta) + \cos\rho\cos\theta & -\cos\phi\sin\rho \\ -\sin\rho\sin(\phi-\theta) & \sin\rho\cos(\phi-\theta) & \cos\rho \end{bmatrix}$$

The rotation parameters can be obtained from the matrix elements using the equations

$$\cos\rho = R_{33}, \quad \tan\phi = -\frac{R_{13}}{R_{23}}$$

and

$$\tan(\phi-\theta) = -\frac{R_{31}}{R_{32}} \quad (55)$$

To establish a direct relationship between ρ, ϕ and θ and the Euler angles φ, η and μ, recall Eq. (54),

$$\mathbf{R} = \mathbf{R}^Z(\mu)\mathbf{R}^X(\eta)\mathbf{R}^Z(\varphi)$$

where $\mathbf{R}^Z(\mu)$ denotes rotation about the Z axis by angle μ, and so on. Since $\mathbf{R}(\mu)^{-1} = \mathbf{R}(\mu)^\top = \mathbf{R}(-\mu)$, is follows that

$$\begin{aligned}\mathbf{R} &= \mathbf{R}^Z(\mu)\mathbf{R}^X(\eta)[\mathbf{R}^Z(-\mu)\mathbf{R}^Z(\mu)]\mathbf{R}^Z(\varphi) \\ &= [\mathbf{R}^Z(\mu)\mathbf{R}^X(\eta)\mathbf{R}^Z(-\mu)]\mathbf{R}^Z(\mu+\varphi).\end{aligned}$$

The rotation $\mathbf{R}^Z(\mu)\mathbf{R}^X(\eta)\mathbf{R}^Z(-\mu)$ represents a rotation out of the plane by η about a line lying in the plane at angle μ: it first rotates this line to the position of

the X axis, then rotates about this new X axis by μ, and then rotates the line back to its former position. This is equivalent to $\mathbf{R}_\rho$, with $\rho = \eta$ and $\phi = \mu$. The rotation $\mathbf{R}^Z(\mu + \varphi)$ is a rotation in the image plane through $\mu + \varphi$, equivalent to $\mathbf{R}_\theta$ with $\theta = \mu + \varphi$. Thus,

$$\rho = \eta, \quad \phi = \mu \quad \text{and} \quad \theta - \phi = \varphi.$$

Acknowledgments

We are grateful for financial support from the ORS Award, the Foundation for Research Development (RSA), Esprit BRA 6448 'VIVA' and SERC Grant GR/J44049. We have had many fruitful discussions with Paul Beardsley, Andrew Blake, Steve Maybank, Phil McLauchlan, David Murray, Ian Reid, Phil Torr and Bill Triggs (all of the RRG) along with Richard Hartley and Joe Mundy (of GE). We thank Chris Harris for kindly providing his data (Fig. 23), Phil McLauchlan for providing his implementation of the Tomasi and Kanade algorithm, and the referees for their helpful comments.

Notes

1. A pure *orthographic* camera arises when the magnification f/Z^c_{ave} is unity (cf. Figs. 2 and 3(b)).
2. This formulation ties in neatly with the scheme of Koenderink and van Doorn (1991). Three world points $\{\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2\}$ define the reference plane π_r, establishing the X-Y axes of the affine world frame (with $\mathbf{X}_0$ the reference point and Z the direction of viewing in the first frame). Since these points lie on π_r, $\Delta Z_1 = \Delta Z_2 = 0$. From Eq. (18), a 2D affine transform then maps the image I of π_r to its image I', that is, $\Delta\mathbf{x}'_i = \mathbf{\Gamma}\,\Delta\mathbf{x}_i$ with $\mathbf{\Gamma} = [\Delta\mathbf{x}'_1\ \Delta\mathbf{x}'_2][\Delta\mathbf{x}_1\ \Delta\mathbf{x}_2]^{-1}$. A point $\mathbf{X}_3$ lying *off* π then satisfies $\Delta\mathbf{x}'_3 = \mathbf{\Gamma}\Delta\mathbf{x}_3 + \Delta Z_3\mathbf{d}$, where $\mathbf{\Gamma}\,\Delta\mathbf{x}_3$ is the "piercing point" of $\mathbf{X}_3$, that is, the position it would project to in I' if it *did* lie on π. The difference vector $\Delta Z_3\mathbf{d}$ (known since both $\Delta\mathbf{x}_3$ and $\Delta\mathbf{x}'_3$ are known) yields the direction of $\mathbf{d}$, $\hat{\mathbf{d}}$, and the scaled depth coordinate $\Delta Z_3\,|\mathbf{d}|$. A second point lying off π ($\mathbf{X}_4$) gives a second scaled depth $\Delta Z_4\,|\mathbf{d}|$, and the third affine coordinate is then formed from the ratio $\Delta Z_4/\Delta Z_3$.
3. Orthogonality is not defined in the affine plane, but we could equally well have resolved *parallel* to $\mathbf{d}^\perp$.
4. Kanatani (1993b) discusses the case of anisotropic image noise.
5. Section 4 provides an alternative definition for s, namely the ratio of the average depth planes $Z_{\text{ave}}/Z'_{\text{ave}}$. Thus s effectively measures divergence ($s > 1$ when the object approaches the camera).
6. They also weighted each point by its inverse distance to the epipole; in the affine case, the epipole lies at infinity so all points are weighted equally.
7. The constraint $\text{rank}(\mathbf{F}) = 2$ (or $\det(\mathbf{F}) = 0$) ensures that all epipolar lines pass through the epipole.
8. When $s = 1$, $E_1 = 2\Sigma(\mathbf{n}^\top\mathbf{r}_i + e)^2/|\mathbf{n}_1|^2$, $E_2 = E_1/2$ and $E_3 = E_1/4$, so the solutions for $\mathbf{n}$ are identical (up to an arbitrary scale factor).
9. The cost function value that is smallest in a given column is obviously the one that was actually minimised.
10. There are four rotation matrices here: the pose of the two cameras ($\mathbf{R}_p$ and $\mathbf{R}'_p$), the motion of the object within the world coordinate system ($\mathbf{R}_m$), and the composite relative rotation between scene and camera ($\mathbf{R}$). The latter is the matrix of interest here.
11. We use this form, rather than $q(\mathbf{n}) = q(\bar{\mathbf{n}} + \delta\mathbf{n})$, in order that the expressions will be functions of the known vector $\mathbf{n}$, rather than the unknown $\bar{\mathbf{n}}$.
12. Lee and Huang (1990) termed this "degenerate motion", and noted that it also arises when the rotation angle about *any* axis is a multiple of 360° (equivalent to no motion) and when $\rho = 180°$. In all such cases, the projection plane remains the same, so the observer doesn't get a new viewing direction.
13. We could observe θ directly; however, $\phi - \theta$ is obtained naturally from $\arctan(d/c)$ in Eq. (36).
14. These constraints can in fact be relaxed to cope with *scaled* orthography (rather than pure orthography) by letting the rows of $\mathbf{LA}$ belong to *scaled* rotation matrices, with the equivalent constraints simply being $\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2 = 0$ and $\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_1 = \mathbf{M}_2^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2$ (Weinshall 1993).
15. $\mathbf{R}$ merely defines the initial coordinate frame (which is arbitrary) while s defines the overall scale of the recovered structure (which is anyhow ambiguous up to a scale factor due to the depth-scale ambiguity).
16. Transformation to the motion parameters for I', I'' is achieved by computing $\mathbf{R}_{23} = \mathbf{R}_{13}\mathbf{R}_{12}^{-1}$.
17. For the weak perspective case, the constraint of unit norm is removed (except for the first $\mathbf{M}$ matrix), and the expression becomes: $(\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2)^2 + (\mathbf{M}_1'^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2')^2 + (\mathbf{M}_1''^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2'')^2 + (\mathbf{M}_1^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_1 - 1)^2 + (\mathbf{M}_2^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2 - 1)^2 + (\mathbf{M}_1'^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_1' - \mathbf{M}_2'^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2')^2 + (\mathbf{M}_1''^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_1'' - \mathbf{M}_2''^\top\mathbf{A}\mathbf{A}^\top\mathbf{M}_2'')^2$.

References

Aloimonos, J. 1986. Detection of surface orientation from texture I: The case of planes. In *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR'86), Florida, pp. 584–593.

Aloimonos, J. and Bandyopadhyay, A. 1985. Perception of structure from motion: lower bound results. Tech. Report 158, Dept. Computer Science, University of Rochester.

Aloimonos, J.Y. 1992. Perspective approximations. *Image and Vision Computing*, 8(3):179–192.

Arnold, R.D. and Binford, T.O. 1980. Geometrical constraints in stereo vision. In *Proceedings S.P.I.E.*, 238:281–292.

Bar-Shalom, Y. and Fortmann, T.E. 1988. *Tracking and Data Association*, Academic Press Inc., USA.

Barnard, S.T. and Thompson, W.B. 1980. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4):333–340.

Beardsley, P.A., Zisserman, A., and Murray, D.W. 1993. Projective structure from image sequences. Tech. Report OUEL 1985/93, Dept. Engineering Science, University of Oxford.

Beardsley, P.A., Zisserman, A., and Murray, D.W. 1994. Navigation using affine structure from motion. In J.O. Eklundh (ed.), *Proceedings European Conference on Computer Vision* (ECCV-94), II:85–96.

Bennett, B.M., Hoffman, D.D., Nicola, J.E., and Prakash, C. 1989. Structure from two orthographic views of rigid motion. *Journal of Optical Society of America*, 6(7):1052–1069.

Berger, M. 1980. *Geometry I*, Springer Verlag.

Bookstein, F.L. 1979. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9:56–71.

Charnley, D., Harris, C., Pike, M., Sparks, E., and Stephens, M. 1988. The DROID 3D vision system: algorithms for geometric integration. Plessey Research, Roke Manor, Technical Note 72/88/N488U.

Chen, H.H. and Huang, T.S. 1991. Using motion from orthographic views to verify 3D point matches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):872–878.

Demey, S., Zisserman, A., and Beardsley, P. 1992. Affine and projective structure from motion. *Proceedings British Machine Vision Conference* (BMVC'92), pp. 49–58.

Durrant–Whyte, H.F. 1993. *Methods and Systems in Data Fusion*, in press.

Faugeras, O.D., Luong, Q.-T., and Maybank, S.J. 1992. Camera self-calibration: theory and experiments. In G. Sandini (ed.), *Proceedings European Conference on Computer Vision* (ECCV-92), pp. 321–334.

Faugeras, O.D. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini (ed.), *Proceedings European Conference on Computer Vision* (ECCV-92), pp. 563–578.

Gelb, M. 1974. *Applied Optimal Estimation*, MIT Press.

Harris, C. 1990. Structure-from-motion under orthographic projection. In *Proceedings European Conference on Computer Vision* (ECCV-90), pp. 118–123.

Hartley, R.I. 1992. Estimation of relative camera positions for uncalibrated cameras. In *Proceedings European Conference on Computer Vision* (ECCV-92), pp. 579–587.

Hollinghurst, N. and Cipolla, R. 1993. Uncalibrated stereo hand-eye coordination. In *Proceedings British Machine Vision Conference* (BMVC'93), Surrey, pp. 389–398.

Hu, X. and Ahuja, N. 1991. Motion estimation under orthographic projection. In *IEEE Transactions on Robotics and Automation*, 7(6):848–853.

Huang, T.S. and Lee, C.H. 1989. Motion and structure from orthographic projections. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-11(5):536–540.

Kanatani, K. 1993. *Geometric Computation for Computer Vision*, Oxford University Press, UK.

Koenderink, J.J. and van Doorn, A.J. 1991. Affine structure from motion. *Journal of Optical Society of America*, 8(2):377–385.

Lee, C. and Huang, T. 1990. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *Computer Vision, Graphics and Image Processing*, 52: 309–327.

Longuet-Higgins, H.C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.

Longuet-Higgins, H.C. 1991. A method of determining the relation positions of 4 points from 3 perspective projections. In P. Mowforth (ed.), *Proceedings British Machine Vision Conference* (BMVC'91), pp. 86–94.

Luong, Q.-T., Deriche, R., Faugeras, O., and Papadopoulo, T. 1993. On determining the fundamental matrix: analysis of different methods and experimental results. Tech. Report 1894, INRIA (Sophia Antipolis).

Mundy, J.L. and Zisserman A. (eds). 1992. *Geometric Invariance in Computer Vision*, MIT Press, USA.

Olsen, S.I. 1992. Epipolar line estimation. In G. Sandini, (ed.), *Proceedings European Conference on Computer Vision* (ECCV-92), pp. 307–311.

Pollard, S.B., Mayhew, J.E.W., and Frisby, J.P. 1985. PMF: A Stereo Correspondence Algorithm using a Disparity Gradient Limit. *Perception*, 14:449–470.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. 1988. *Numerical Recipes in C*, Cambridge University Press, USA.

Quan, L. and Mohr, R. 1991. Towards structure from motion for linear features through reference points. *IEEE Workshop on Visual Motion*, New Jersey.

Reid, I.D. and Murray, D.W. 1993. Tracking foveated corner clusters using affine structure. In *Proceedings International Conference on Computer Vision (ICCV-4)*, Berlin, pp. 76–83.

Reid, I.D. and Murray, D.W. 1994. Active tracking of foveated feature clusters using affine structure. To appear IJCV.

Shapiro, L.S. 1993. *Affine Analysis of Image Sequences*, Ph.D. thesis, Dept. Engineering Science, Oxford University. Also 1995, Cambridge University Press, UK.

Shapiro, L.S. and Brady, J.M. 1993. Rejecting outliers and estimating errors in an orthogonal regression framework. Tech. Report OUEL 1974/93, Dept. Engineering Science, University of Oxford. Also in *Phil. Tran. R. Soc. Lond.*, A (1995) 350: 407–439.

Shapiro, L.S., Wang H., and Brady, J.M. 1992. A matching and tracking strategy for independently moving objects. In D. Hogg and R. Boyle (eds.), *Proceedings British Machine Vision Conference*, Leeds, Springer-Verlag, U.K., pp. 306–315.

Shapiro, L.S., Zisserman, A., and Brady, J.M. 1994. Motion from point matches using affine epipolar geometry. In J.O. Eklundh (ed.), *Proceedings European Conference on Computer Vision* (ECCV-94), II:73–84.

Strang, G. 1988. *Linear Algebra and its Applications*, 3rd ed., Harcourt Brace Jovanovich Inc., U.S.A.

Thompson, D.W. and Mundy, J.L. 1987. Three dimensional model matching from an unconstrained viewpoint. In *IEEE Conference on Robotics and Automation*, Raleigh, NC, pp. 208–220.

Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154.

Torr, P.H.S. 1993. Notes on epipole at infinity. Personal communication.

Torr, P.H.S. and Murray, D.W. 1993. Outlier detection and motion segmentation. In Schenker (ed.), *Sensor Fusion VI*, SPIE Vol. 2059, Boston, pp. 432–443.

Ullman, S. 1979. *The Interpretation of Visual Motion*, MIT Press, U.S.A.

Ullman, S. and Basri, R. 1991. Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):992–1006.

Wang, H. and Brady, J.M. 1992. Corner detection: some new results. *IEE Colloquium Digest of System Aspects of Machine Perception and Vision*, pp. 1.1–1.4. London: IEE.

Weinshall, D. and Tomasi, C. 1993. Linear and incremental acqui-

sition of invariant shape models from image sequences. In *Proceedings International Conference on Computer Vision (ICCV-4)*, pp. 675–682.

Weinshall, D. 1993. Model-based invariants for 3-D vision. *International Journal of Computer Vision*, 10(1):27–42.

Weng, J., Huang, T.S., and Ahuja, N. 1989. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-11(5):451–476.

Weng, J., Ahuja, N., and Huang, T.S. 1993. Optimal motion and structure estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-15(9):864–884.

Xu, G., Nishimura, E., and Tsuji, S. 1993. Image correspondence and segmentation by epipolar lines: theory, algorithm and applications. Technical Report, Dept. Systems Engineering, Osaka University.

Zisserman, A. 1992. *Notes on geometric invariance in vision: BMVC'92 tutorial*, Leeds University.