# Resolution of the Bas-Relief Ambiguity in Structure-from-Motion under Orthographic Projection

**Chris Harris**

Roke Manor Research Ltd.
Roke Manor, Romsey, Hants SO51 0ZN, U K

*Under orthographic projection, two views of a rigid moving object are insufficient to uniquely determine the structure and motion of the object, due to the existence of the bas-relief ambiguity. Resolution of this ambiguity relies upon either prior information about the motion of the object, or else further views of the object. Two novel algorithms to resolve the bas-relief ambiguity are developed; they are well-formulated in that they minimise image-plane error, and are shown working on sequences of real images.*

Structure-from-motion (SFM) algorithms are used in the analysis of image motion caused by relative three-dimensional (3D) movement between the camera and the (unknown) imaged objects, which are assumed to be rigid. These algorithms attempt to recover both the 3D structure of the image objects and the 3D motion of each object with respect to the camera (or *vice versa*). The SFM algorithms explored in this paper use point image features, extracted independently from each image in the sequence by use of a 'corner' detection algorithm [1], and matched between images forming the sequence [2].

As the imaging mechanism of conventional cameras is perspective projection (ie. cameras behave as if they were 'pin-hole' cameras), most SFM algorithms have been based on perspective projection [3,4]. These algorithms have been found to provide acceptable solutions to the 'ego-motion' problem, where a camera (of relatively wide field-of-view) moves through an otherwise static environment. However, for the perspective SFM algorithms to be well-conditioned, the angle subtended by the viewed object (in the ego-motion problem, the viewed scene) must be large, and the viewed object must span a relatively large range of depths. Thus the perspective SFM algorithms are of little or no practical use for analysing everyday imagery of independently moving distant objects, such as driven cars and flying aircraft. It is algorithms for the analysis of such imagery that is the concern of this paper.

The ill-conditioning of the perspective SFM algorithms that we wish to circumvent occurs for objects subtending a small range of depths, and generally subtending a small angle. In these circumstances, a good approximation to the imaging process is orthographic projection, in which the variation in object depth is assumed to be negligible with respect to the distance of the object from the camera. A further reason for using orthographic projection is that it is mathematically tractable for the analysis of the motion of point-like image features between two images of a sequence [5,6]. Unfortunately, the SFM analysis of a pair of images suffers from an ambiguity in interpretation, the bas-relief ambiguity [6], which both distorts the structure and re-orientates the axis of rotation (unlike the speed-scale ambiguity, which leaves the structure undistorted).

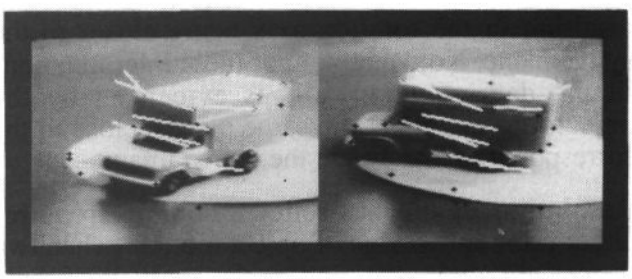

*Figure 1. Disparity vectors on two matched images*

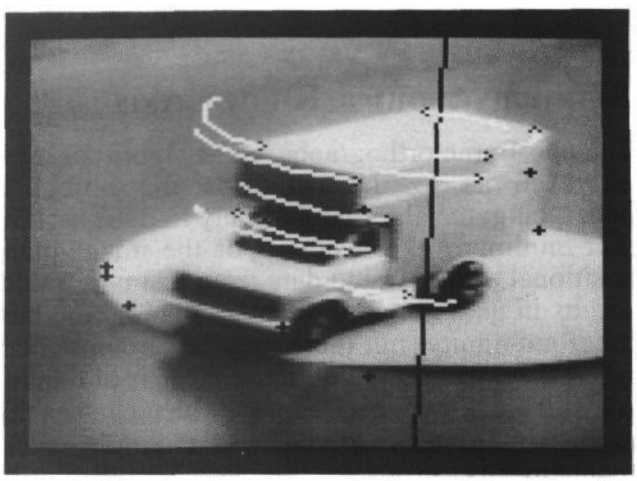

*Figure 2. The 60° rotation interpretation*

The author gratefully acknowledges the European Commission who together with the Plessey Company supported this work under the Esprit Initiative project P2502.



BMVC 1990 doi:10.5244/C.4.14

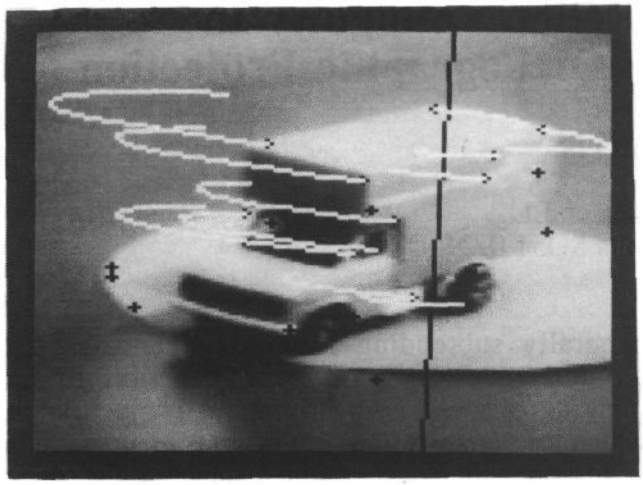

*Figure 3. The 120° rotation interpretation*

The bas-relief ambiguity is illustrated on the two images shown in Figure 1, of a toy truck which has undergone a rotation of 60°. Detected feature-points are indicated by black crosses, and the motion of matched feature-points by the white disparity vectors. Two (of the infinity) of motion interpretations are shown in Figures 2 and 3. In Figure 2, the correct (ie. 60°) rotation is selected, and the trajectories of the matched feature-points as the truck rotates shown as white curves. An incorrect rotation of 120° is shown in Figure 3, resulting in the rotation axis being closer to the image plane. Note that both of these interpretations fit the data equally well, as evidenced by the trajectories ending at the same places, close to the corresponding matches in the paired image.

Resolution of the bas-relief ambiguity requires either prior knowledge of object motion (when two frames are analysed), or else the analysis of three or more frames. Two new algorithms are presented below which use the aforementioned two approaches of breaking the bas-relief ambiguity, and they are applied to sequences of real images.

## Rotation About a Known Axis

The two-frame orthographic SFM algorithm [6] is unbiased as to the rotations the object is undergoing: the algorithms simply find the rotation that minimises the residuals of the image-plane positional errors. Using data from just two frames results in an ambiguity of interpretation, the bas-relief ambiguity, but there may be circumstances where *a priori* knowledge about the axis of rotation is available or may be assumed, which enables this ambiguity to be resolved. For example, a car manoeuvring on a flat, horizontal road may be assumed to be rotating about a vertical axis. In this section we shall assume that the object is rotating an unknown amount about an axis that appears vertical in the image (if the axis were at some other angle, then the image could be rotated until the axis was vertical in the image), and that the projection is orthographic, with the projection of the vertical axis aligned along the vertical (y) image axis (see Figure 4).

Let there be n matches between the two frames, at image locations $\{x_i,y_i\}$ on the first frame, and at $\{x'_i,y'_i\}$ on the second frame. Define a coordinate system with the z-axis aligned along the optical axis, the x and y axes aligned with the image coordinate axes, and the origin at a distance L in front of the centre of projection (the camera pin-hole), so placing the centre of projection at z = -L (see Figure 1). Let the i'th point on the moving object be located in 3D at $\mathbf{S}_i = (X_i,Y_i,Z_i)$ at the time of the first frame, and at $\mathbf{S}'_i = (X'_i,Y'_i,Z'_i)$ on the second frame. Below, the object will be assumed to be situated close to the coordinate origin, and be small compared to L. Without loss of generality, decompose the object motion between the two frames as a rotation about the origin, specified by the orthogonal rotation matrix R, followed by a translation **t**. Hence

$$\mathbf{S}'_i = R\,\mathbf{S}_i + \mathbf{t}$$

Perspective projection onto a forward image plane a unit distance from the camera pin-hole gives

$$(x_i, y_i) = (X_i, Y_i) / (L + Z_i)$$

$$(x'_i, y'_i) = (X'_i, Y'_i) / (L + Z'_i)$$

Substituting gives

$$x'_i \approx R_{11}x_i + R_{12}y_i + R_{13}z_i + t_x / L + O(L^{-2})$$

and similarly for $y'_i$. Dropping the $O(L^{-2})$ terms for large L (this is the orthographic limit), and without loss of generality setting L=1, gives

$$x'_i = R_{11}x_i + R_{12}y_i + R_{13}z_i + t_x$$

$$y'_i = R_{21}x_i + R_{22}y_i + R_{23}z_i + t_y$$

Now, for real data, the positions $\{x_i, y_i, x'_i, y'_i\}$ will be contaminated by measurement noise, so that the above equations will not hold true exactly. Assuming isotropic Gaussian noise on the observed image-plane locations, the maximum likelihood solution is found by minimising, $\varepsilon$, the sum of the squares of the residuals of the above equations

$$\varepsilon(R,t_x,t_y,\{z_i\}) = \sum_{i=1}^{n} [\,(R_{11}x_i + R_{12}y_i + R_{13}z_i + t_x - x'_i)^2 + (R_{21}x_i + R_{22}y_i + R_{23}z_i + t_y - y'_i)^2\,]$$

Note that it is the actual residuals of the image-plane locations that are being minimised, and not

some other mathematically convenient but less meaningful formulation (for example, eliminating $z_i$ between the two equations). To minimise $\varepsilon$ with respect to $t_x$, set $\partial\varepsilon/\partial t_x = 0$, giving

$$t_x = -\sum_{i=1}^{n} (R_{11}x_i + R_{12}y_i + R_{13}z_i - x_i') / n$$

$$= -(R_{11}\bar{x} - R_{12}\bar{y} - R_{13}\bar{z}) + \bar{x}'$$

where bars indicate means over all matched points. Minimising similarly with respect to $t_y$, and substituting back into the equation for $\varepsilon$ gives

$$\varepsilon(R,\{z_i\}) = \sum_{i=1}^{n} [\ (R_{11}x_i+R_{12}y_i+R_{13}(z_i-\bar{z})-x_i')^2 + (R_{21}x_i+R_{22}y_i+R_{23}(z_i-\bar{z})-y_i')^2\ ]$$

where, for conciseness, the means have been subtracted from all of the image observations, $\{x_i,y_i,x_i',y_i'\}$. To perform the minimisation over the object rotations, write the rotation matrix, R, as

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

$$= \begin{pmatrix} \cos\theta & \sin\alpha\sin\theta & \cos\alpha\sin\theta \\ -\sin\alpha\sin\theta & \cos^2\alpha+\sin^2\alpha\cos\theta & -\sin\alpha\cos\alpha(1-\cos\theta) \\ -\cos\alpha\sin\theta & -\sin\alpha\cos\alpha(1-\cos\theta) & \sin^2\alpha+\cos^2\alpha\cos\theta \end{pmatrix}$$

where $\theta$ is the unknown angle of rotation, and $\alpha$ is the angle the 3D rotation axis makes to the vertical (y) axis of the image (see Figure 4).

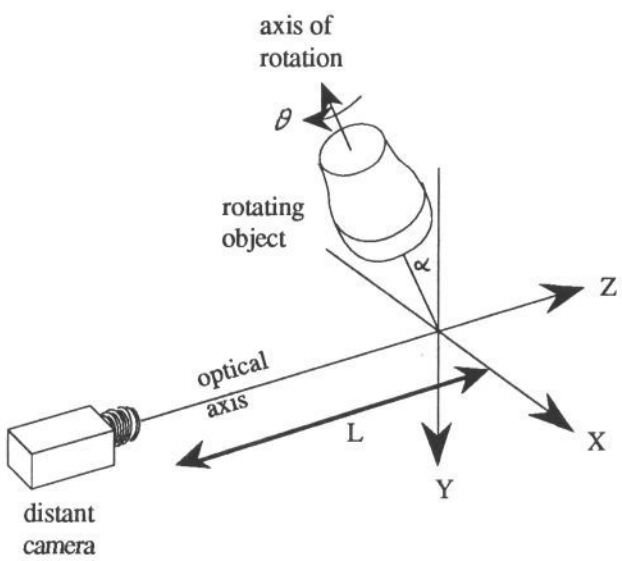


*Figure 4. Coordinate system.*

Substituting for elements of the rotation matrix enables $\varepsilon$ to be written as

$$\varepsilon(\theta) = \sum_{i=1}^{n} \{\ [\ x_i\cos\theta + y_i\sin\theta\sin\alpha - x_i' + (z_i - \bar{z})\sin\theta\cos\alpha\ ]^2 + [\ x_i\sin\alpha\sin\theta - y_i(\cos^2\alpha+\sin^2\alpha\cos\theta) + y_i' + (z_i - \bar{z})\sin\alpha\cos\alpha(1-\cos\theta)\ ]^2\ \}$$

Minimising $\varepsilon$ with respect to $z_i$ by setting $\partial\varepsilon/\partial z_i = 0$, and substituting back into the equation for $\varepsilon$ gives

$$\varepsilon(\theta) = \sum_{i=1}^{n} [\ (x_i'+x_i)\sin\alpha(1-\cos\theta) - (y_i'-y_i)\sin\theta\ ]^2 / [\ \sin^2\theta + \sin^2\alpha(1-\cos\theta)\ ]$$

Letting $s = \sin\alpha \tan(\theta/2)$ gives

$$\varepsilon(s) = \sum_{i=1}^{n} [\ s(x_i' + x_i) - (y_i' - y_i)\ ]^2 / [\ s^2 + 1\ ]$$

Finally, setting $\partial\varepsilon/\partial s = 0$ results in the following quadratic equation

$$s^2 - s\,\frac{\sum_{i=1}^{n}[(x_i' + x_i)^2 - (y_i' - y_i)^2]}{\sum_{i=1}^{n}(x_i' + x_i)(y_i' - y_i)} - 1 = 0$$

Of the two solutions to this equation, choose the one which minimises the value of $\varepsilon$. It is interesting to note that the single variable s contains all the dependence on both $\alpha$ and $\theta$, so that if the rotation angle were $\theta$ known, but the angle $\alpha$ unknown, it too could be determined by solving the above equation.

## 3-Frame Algorithm

The most general (ie. least biased) approach to resolving the bas-relief ambiguity under orthographic projection is to use three or more views of the object (provided in practice that they span a great enough range of viewing directions). In keeping with our previous approach, we shall consider only algorithms formulated to minimis image-plane residuals, as these should show the greatest stability and range of applicability. However, the minimum residuals formulation generally appears to be intractable, except in the case of three views and constant angular

displacement between views, and this is the case considered below.

Let there be n matches between the three frames, at image locations $\{x_i,y_i\}$ on the first frame, at $\{X_i, Y_i\}$ on the second frame, and at $\{x'_i,y'_i\}$ on the third frame. Performing the orthographic projection as before, let the i'th point have (unknown) depth $Z_i$ on the second frame, and so be located in 3D at $\mathbf{S}_i = (X_i,Y_i,Z_i)$. Let the object motion between frames two and one be decomposed into a rotation about the origin, as specified by the orthogonal rotation matrix R, followed by a translation $\mathbf{t}$

$\mathbf{r}_i = R\,\mathbf{S}_i + \mathbf{t}$ (x and y components)

Similarly, let the object motion between frames two and three be given by

$\mathbf{r}'_i = R^T\,\mathbf{S}_i + \mathbf{t}'$ (x and y components)

Note that the choice of the rotation matrix to be the transpose of that used before means that the angular displacements between frames one and two and between frames two and three have been chosen to be equal. This is done principally on grounds of mathematical tractability, but, if a correct assumption, will aid the stability of the algorithm. This assumption will often be appropriate for three equally spaced framed taken in rapid succession of an object with relatively large moment of inertia.

Explicitly, the above equations are

$$x_i = R_{11}X_i + R_{12}Y_i + R_{13}Z_i + t_x$$

$$y_i = R_{21}X_i + R_{22}Y_i + R_{23}Z_i + t_y$$

$$x'_i = R_{11}X_i + R_{21}Y_i + R_{31}Z_i + t'_x$$

$$y'_i = R_{12}X_i + R_{22}Y_i + R_{32}Z_i + t'_y$$

Now, for real data, the positions $\{x_i, y_i, x'_i, y'_i\}$ will be contaminated by measurement noise, so that the above equations will not hold true exactly. Assuming isotropic Gaussian noise on the observed image-plane locations, the maximum likelihood solution is found by minimising, $\varepsilon$, the sum of the squares of the residuals of the above equations

$$\varepsilon(R,t_x,t_y,t'_x,t'_y,\{Z_i\}) =$$

$$\sum_{i=1}^{n} [\ (R_{11}X_i+R_{12}Y_i+R_{13}Z_i+t_x-x_i)^2 + (R_{21}X_i+R_{22}Y_i+R_{23}Z_i+t_y-y_i)^2 + (R_{11}X_i+R_{21}Y_i+R_{31}Z_i+t'_x-x'_i)^2 + (R_{12}X_i+R_{22}Y_i+R_{32}Z_i+t'_y-y'_i)^2\ ]$$

To start on the minimisation of $\varepsilon$, first define

$$u_i = R_{11}X_i + R_{12}Y_i - x_i$$

$$v_i = R_{21}X_i + R_{22}Y_i - y_i$$

$$u'_i = R_{11}X_i + R_{21}Y_i - x'_i$$

$$v'_i = R_{12}X_i + R_{22}Y_i - y'_i$$

$$\varepsilon = \sum_{i=1}^{n} [\ (u_i + R_{13}Z_i + t_x)^2 + (v_i + R_{23}Z_i + t_y)^2 + (u'_i + R_{31}Z_i + t'_x)^2 + (v'_i + R_{32}Z_i + t'_y)^2\ ]$$

To minimise $\varepsilon$ with respect to $t_x$, set the differential to zero, resulting in

$$\sum_{i=1}^{n} [\ (u_i + R_{13}Z_i + t_x) = 0.$$

Hence $t_x = -\sum_{i=1}^{n} (u_i + R_{13}Z_i)\ /\ n\ =\ \bar{u} + R_{13}\bar{Z}$

where the bars indicate means over all the matched points. Minimising similarly with respect to the other components of translation gives

$$\varepsilon\,(R,\{Z_i\}) =$$

$$\sum_{i=1}^{n} [(u_i - \bar{u} + R_{13}(Z_i-\bar{Z}))^2 + (v_i - \bar{v} + R_{23}(Z_i-\bar{Z}))^2 + (u'_i - \bar{u}' + R_{31}(Z_i-\bar{Z}))^2 + (v'_i - \bar{v}' + R_{32}(Z_i-\bar{Z}))^2]$$

Henceforth, for conciseness, write $u_i - \bar{u}$ as $u_i$, etc. This is accomplished simply by removing the means from the all the observations, $\{x_i,y_i,X_i,Y_i,x'_i,y'_i\}$. Minimising now with respect to the $Z_i$'s gives

$$Z_i - \bar{Z} = -\ [\ R_{13}u_i + R_{23}v_i + R_{31}u'_i + R_{32}v'_i\ ]\ /\ [\ R_{13}^2 + R_{23}^2 + R_{31}^2 + R_{32}^2\ ]$$

Substituting back for the $Z_i$'s leaves the following residuals term to be minimised

$$\varepsilon(R) = \sum_{i=1}^{n} \{\ [\ u_i^2 + v_i^2 + u_i^2 + v_i^2\ ]\ -\ [\ R_{13}u_i + R_{23}v_i + R_{31}u'_i + R_{32}v'_i\ ]^2\ /\ [\ R_{13}^2 + R_{23}^2 + R_{31}^2 + R_{32}^2\ ]\ \}$$

Without loss of generality, write the rotation matrix, R, as

$$R = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\eta & -\sin\eta \\ 0 & \sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \cos\phi\cos\theta + \sin\phi\sin\theta\cos\eta & \cos\phi\sin\theta - \sin\phi\cos\theta\cos\eta & \sin\phi\sin\eta \\ \sin\phi\cos\theta - \cos\phi\sin\theta\cos\eta & \sin\phi\sin\theta + \cos\phi\cos\theta\cos\eta & -\cos\phi\sin\eta \\ -\sin\theta\sin\eta & \cos\theta\sin\eta & \cos\eta \end{pmatrix}$$

Setting $\lambda = \cos\eta$ , and substituting the above elements of the rotation matrix gives

$$\varepsilon(\theta,\phi,\lambda) =$$

$$\sum_{i=1}^{n} \{ \; (X_i\cos\theta + Y_i\sin\theta - x_i\cos\phi - y_i\sin\phi)^2 +$$

$$(X_i\cos\phi + Y_i\sin\phi - x_i'\cos\theta - y_i'\sin\theta)^2 +$$

$$[\,(\lambda Y_i - y_i')\cos\theta - (\lambda X_i - x_i')\cos\theta +$$

$$(\lambda Y_i - y_i))\cos\phi - (\lambda X_i - x_i)\cos\phi\,]^2 / 2\,] \}$$

Keeping for the moment $\lambda$ constant, the above equation can be transforming into finding the zeros of an 8'th order polynomial. Zeros of this polynomial are found by using a standard numerical algorithm (such as NAG), and the solution generated by each real root compared numerically to see which provides the minimum value of $\varepsilon$. $\varepsilon(\lambda)$ turns out in practice to be a well-behaved function with a single minimum in the range (-1,1) in all cases investigated, and the approximate location of the minimum value of $\lambda$ we determine numerically by a binary chop method starting at $\lambda$=-1 and $\lambda$=1.

## Results

The SFM algorithms were applied to a sequence of real images of a toy truck on a turntable, shown in Figure 5. The images are 128 pixels square, and the truck subtends an angle of about 5° from the camera. Between each frame of the sequence, the truck was rotated by 10° about an axis passing through the centre of the turntable, and oriented some 3° clockwise of the vertical. Thus the true projection of the axis of rotation is a nearly vertical line in the image, about one third of the image width from the right-hand edge of the image. From each image between 20 and 30 feature-points were extracted using a corner detector [1], and these are indicated by the black crosses in the later Figures. The feature-points were matched by hand for expediency, though work is currently underway to automate the matching procedure.

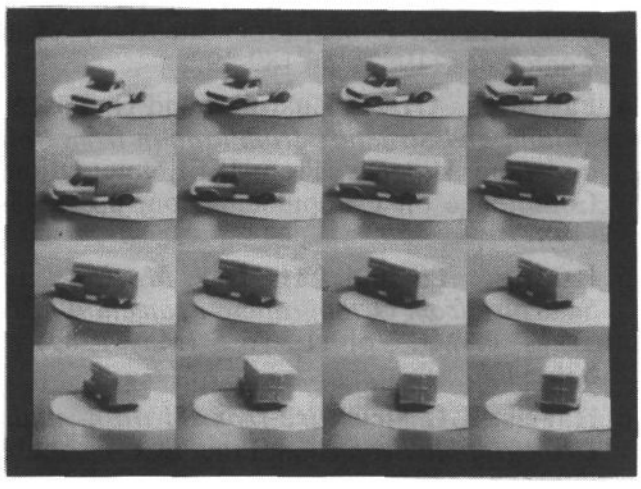

*Figure 5. Sequence of 16 images.*

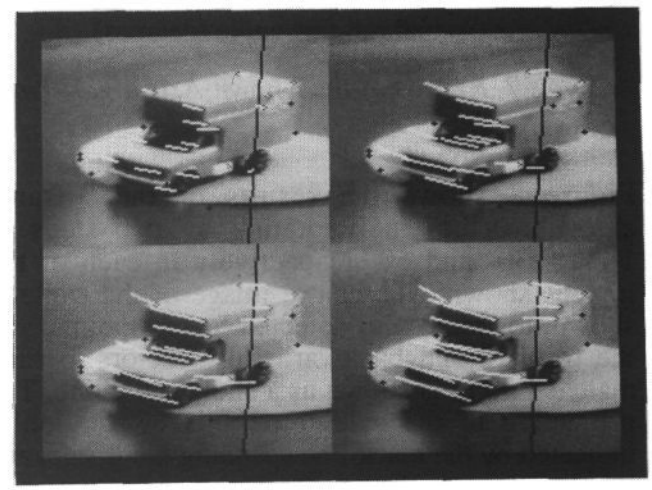

*Figure 6. Results of the Known-Axis algorithm for rotations of 10°, 20°, 30° and 40°.*

In Figure 6 are shown results of the Known-Axis algorithm as applied to pairs of images, the truck rotating by angles of 10°, 20°, 30° and 40° respectively. The first image of each pair analysed is shown in the Figure. The flow-vector of each matched point is shown as a short white line, which terminates at the location of the feature-point in the later of the image pairs. The projection of the calculated axis of rotation is indicated by the black line spanning the image, and it is seen to pass correctly through the centre of the turntable. The calculated rotation angles for Figure 6 are 9.4°, 19.1°, 30.7° and 44.9°. Results of all pair-wise analyses of the 16 images are shown in Figure 7, and a good correspondence between actual and calculated rotations is obtained for all angles of rotation.

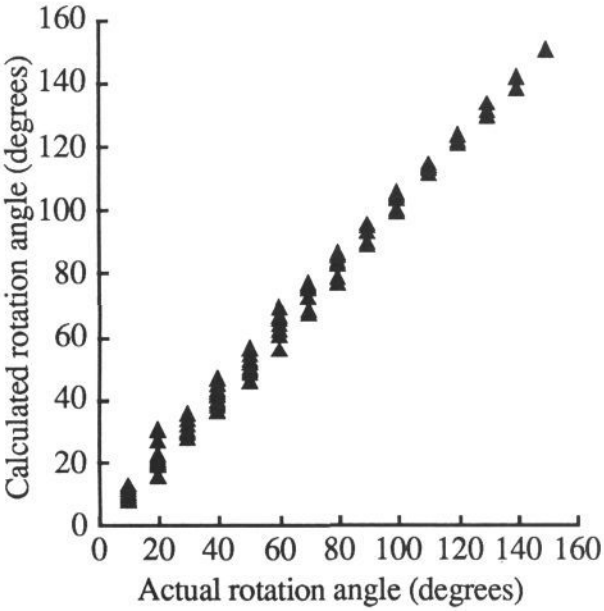


*Figure 7. Results of the Known-Axis algorithm.*

Figure 8 shows the results of the 3-frame algorithm, the truck rotating by two steps of 10°,20°,30° and 50° respectively in each of the images shown. The centre image of the triple is shown, together with detected feature-points, and the forward and backward flow-vectors to each matched point is shown as a short white line. The projection of the calculated axis of rotation is indicated by the black line spanning the image, and it is seen to pass close to the centre of the turntable. The lay-over of the rotation axis is illustrated by the black ellipse, which is the projection of a circle lying in the plane of rotation, and centred on the axis of rotation (its radius was chosen to be equal to that of the turntable for ease of interpreting the results).

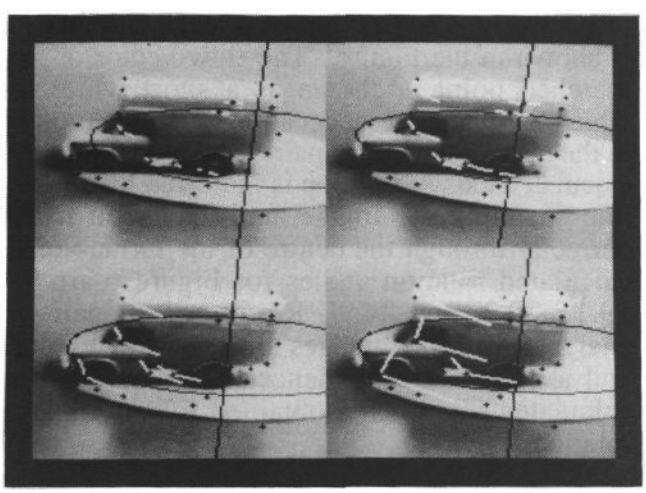

*Figure 8. Results of the 3-frame algorithm for rotations of 10°, 20°, 30° and 50°.*

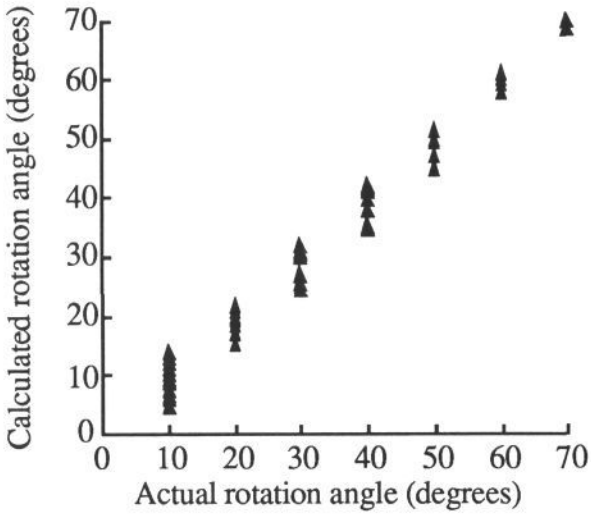


*Figure 9. Results of the 3-frame algorithm.*

The calculated rotation angles for the four images are respectively 10.9°, 21.6°, 32.1° and 51.6°.

Results of analysing all equally spaced triples of the 16 images are shown in Figure 9; here the truck undergoes two rotations of the indicated rotation angle. Note that the results show a larger spread for the smaller angles of rotation, because here the rotation is sufficiently small for the bas-relief ambiguity not to be well resolved. In general, a good correspondence between the calculated and actual angles of rotation is obtained.

## References


1 **Harris, CG** and **MJ Stephens**, *A Combined Corner and Edge Detector*, Proc. 4th Alvey Vision Conference (1988), pp.147-152.

2 **Harris, CG** and **JM Pike**, *3D Positional Integration from Image Sequences*, Proc. 3rd Alvey Vision Conference (1987), pp. 233-236.

3 **Harris, CG**, *Determination of Ego-Motion from Matched Points*, Proc. 3rd Alvey Vision Conference (1987), pp.189-192.

4 **Faugeras, OD, F Lustman** and **G Toscani** *Motion and Structure from Motion from Point and Line Matches*, Proc. IEEE International Conference on Computer Vision, pp. 25-34, (1987).

5 **Huang, TS** and **CH Lee**, *Motion and Structure from Orthographic Projection*, IEEE Trans. PAMI, Vol 11, No 5 (May 1989) pp. 536-540.

6 **Harris, CG** *Structure-from-Motion under Orthographic Projection*, Proc. ECCV, Antibes (1990) pp.118-123.