International Journal of Computer Vision 33(1), 51–72 (1999)

Generalised Epipolar Constraints

KALLE ÅSTRÖM
Department of Mathematics, Lund University, Lund, Sweden

ROBERTO CIPOLLA
Department of Engineering, University of Cambridge, Cambridge, UK

PETER GIBLIN
Department of Pure Mathematics, University of Liverpool, Liverpool, UK

Received October 14, 1996; Revised December 28, 1998; Accepted February 4, 1999

Abstract. In this paper we will discuss structure and motion problems for curved surfaces. These will be studied using the *silhouettes* or *apparent contours* in the images. The problem of determining camera motion from the apparent contours of curved three-dimensional surfaces, is studied. It will be shown how special points, called epipolar tangency points or frontier points, can be used to solve this problem. A generalised epipolar constraint is introduced, which applies to points, curves, as well as to apparent contours of surfaces. The theory is developed for both continuous and discrete motion, known and unknown orientation, calibrated and uncalibrated, perspective, weak perspective and orthographic cameras. Results of an iterative scheme to recover the epipolar line structure from real image sequences using only the outlines of curved surfaces, is presented. A statistical evaluation is performed to estimate the stability of the solution. It is also shown how the motion of the camera from a sequence of images can be obtained from the relative motion between image pairs.

Keywords: curved surface, epipolar geometry, frontier point, uncalibrated camera, apparent contour, silhouette, motion extraction

1. Introduction

Structure and motion from the images of point features has attracted considerable attention and a large number of algorithms exist to recover both the spatial configuration of the points and the motion compatible with the views. The problem to recover structure and motion from the silhouettes or apparent contours of curved surfaces is more difficult, mainly because of the so called *aperture problem*, i.e., it is not possible to get the correspondence of points between two images of the same curve.

This paper is concerned with the problem of recovering viewer motion from the deformations of apparent contours. It is shown how special points on the apparent contour, called *frontier points*, can be detected in image sequences and used to recover viewer motion. The special case of frontier points under orthographic projection and object rotation around a single axis was considered in (Rieger, 1986; Giblin et al., 1994). In (Porrill and Pollard, 1991), although primarily concerned with stereo calibration from 3D space curves, it was noted that the intersection of the two contour generators from two discrete viewpoints generated a point, visible in both images. This constraint was exploited in (Carlsson, 1994) in the analysis of the visual motion of space-curves. An approach for parallel projection has been presented in (Vijaykumar et al., 1995, 1996). Another approach using trinocular stereo has been presented in (Joshi et al., 1995).

In this paper we introduce the general epipolar constraint. We derive the constraint for several camera models both in discrete and continuous time. We discuss the singularities of the contour generator and show that the general epipolar constraint can be derived from one of these singular cases. And finally we show how the constraint can be used to calculate the relative viewer motion between two cameras.

The viewing geometry of surfaces is discussed in Section 2. Then follows a brief discussion about the problem of extracting surface structure from the deformation of silhouettes with *known* camera motion. This has been treated in (Cipolla, 1991; Cipolla and Blake, 1992). The failure modes of these structure from motion algorithms can be understood by analysis of some singular cases of the surface and viewing geometry. The rest of the paper is devoted to one of these singular cases, the epipolar tangency point. This is exploited in Section 3 where a *generalised epipolar constraint* is derived. This constraint is similar to the bilinear constraint. There are many variations of the problem: The motion can be continuous or discrete, the cameras may be calibrated and uncalibrated, parallel and central projection may be considered. Section 3 is of necessity elaborate because all these cases are treated. The generalised epipolar constraint can be used to estimate the relative motion between two images. Implementational aspects on how to use the generalised epipolar constraint are discussed in Section 4. A statistical evaluation in Section 5 gives an estimate of the stability of the solution. Preliminary experimental results obtained from real image sequences of curved surfaces from unknown viewpoints are given in Section 6. Section 7 contains a discussion on how to use the relative motion between image pairs to calculate camera motion in a longer sequence of images. This is also demonstrated in an example. Some conclusions are given in Section 8.

2. Surface and Viewing Geometry

In this section we will provide notations and background material for the camera and surface geometry.

Definition 1. Let B be an open bounded subset of $\mathbb{R}^3$ with C^1 boundary. The boundary U is called a *curved surface*

This paper deals with the *structure and motion* problem from the images of curved surfaces.

Definition 2. For every camera position $\mathbf{c} \notin \bar{B}$, define the **contour generator** as

$$\Gamma_{\mathbf{c}} = \{\mathbf{r} \in U \mid \mathbf{c} \in \text{ tangency plane of } U \text{ at } \mathbf{r}\}. \quad (1)$$

Denote by $\mathbf{n} = \mathbf{n}(\mathbf{r})$ the normal to the surface U at a point $\mathbf{r} \in U$. The contour generator can then be defined as

$$\Gamma_{\mathbf{c}} = \{\mathbf{r} \in U \mid (\mathbf{r} - \mathbf{c}) \cdot \mathbf{n}(\mathbf{r}) = 0\}. \quad (2)$$

The contour generator can also be described as the locus of points Γ on the surface which separates the visible from the occluded parts. This is illustrated in Fig. 1. When the viewpoint $\mathbf{c}$ changes, i.e., $\mathbf{c} = \mathbf{c}(t)$, the contour generator moves over the surface U. This is illustrated in Figs. 2 and 3. In the sequel $\mathbf{c}$ will sometimes be omitted and the notation $\Gamma(t) = \Gamma_{\mathbf{c}(t)}$ will be used for the contour generator at time t.

Definition 3. The image γ of the contour generator Γ_c is called the **apparent contour**.

The *apparent contour* is usually, but not always, a smooth curve. Although the image is considered to be the viewing sphere, it is sometimes convenient to illustrate images and features in the image plane. Notice

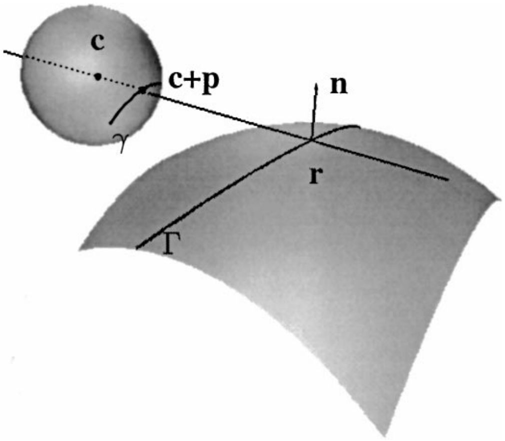

Figure 1. Perspective projection: the contour generator Γ with a typical point $\mathbf{r}$, the image sphere with centre $\mathbf{c}$ and the corresponding apparent contour point $\mathbf{c} + \mathbf{p}$. Thus $\mathbf{p}$ is the unit vector joining the centre $\mathbf{c}$ to the apparent contour point. Also $\mathbf{n}$ is normal to the surface at $\mathbf{r}$.

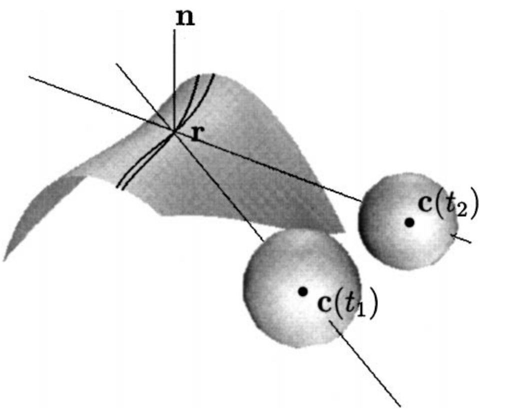

Figure 2. Degenerate case of epipolar parameterisation. The epipolar plane is a tangent plane of the surface at a frontier point. Movement of the viewpoint causes the contour generators to sweep over the surface. At a frontier point the contour generators from consecutive viewpoints intersect.

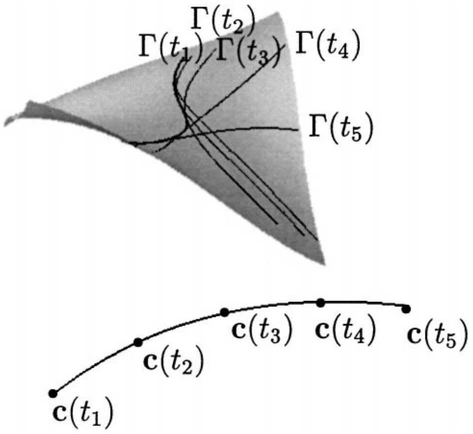

Figure 3. The figure shows a surface U and camera positions $\mathbf{c}(t)$. Five camera positions and the corresponding contour generator is shown. The frontier is the envelope of the contour generators corresponding to the camera positions $\mathbf{c}(t)$.

that lines in the image plane correspond to great circles on the viewing sphere.

Definition 4. If every apparent contour $\gamma_{\mathbf{c}(t)}$, $t_0 < t < t_1$, admits a parametrisation, $\gamma_{\mathbf{c}(t)}(s)$, such that

$$u(s,t) = \gamma_{\mathbf{c}(t)}(s)$$

is continuous, then u is called a **spatio-temporal parametrisation**.

In practice the whole apparent contour might not be visible e.g., due to occlusion. By abuse of language, we often use the term apparent contour and spatio-temporal parametrisation for those parts that have been detected.

The projection from contour generators on U to apparent contours in the image sphere will now be analysed in a little more detail. In the sequel different notations for the image point u will be used for different camera cases.

- Known internal calibration and camera orientation: $u = \mathbf{p}$.
- Calibrated camera: $u = \mathbf{q}$.
- Uncalibrated camera: $u = \mathbf{w}$.

Note that in each of these three cases the image point is represented by a 3×1 vector u. In the case of calibrated cameras this vector can be interpreted as a direction in three-dimensional Euclidean space. In the case of uncalibrated cameras this vector can be interpreted as the image point (x, y) in homogeneous or extended coordinates $u = [x \quad y \quad 1]^{\mathrm{T}}$.

For simplicity assume that the orthonormal camera and object coordinate systems coincide. Let all points be identified with their coordinate vectors. The image $\mathbf{p}$ on the viewing sphere of a point $\mathbf{r} \in U$ fulfills

$$\lambda\mathbf{p} = \mathbf{r} - \mathbf{c},$$

where λ is the depth of $\mathbf{r}$ and $|\mathbf{p}| = 1$. Thus the apparent contour is

$$\gamma_{\mathbf{c}} = \{\mathbf{p} \in \mathbb{S}^2 \mid \exists\lambda, \lambda\mathbf{p} = \mathbf{r} - \mathbf{c}, \mathbf{r} \in \Gamma_{\mathbf{c}}, \}. \quad (3)$$

Note that $\mathbf{p}$ is the orientation of the ray in the fixed reference/world frame for Euclidean 3-space. Now introduce an orthonormal camera coordinate system, where coordinates are denoted $\mathbf{q}$. For a moving observer the camera coordinate system is continuously changing with respect to the object coordinate system. The relationship between $\mathbf{p}$ and $\mathbf{q}$ can be conveniently expressed in terms of a rotation operator $\mathbf{R}(t)$,

$$\mathbf{p} = \mathbf{R}(t)\mathbf{q}. \quad (4)$$

The measurements in an **uncalibrated** camera, described by the coordinate vector $\mathbf{w}$ in some affine camera coordinate system, is related to the spherical image position $\mathbf{q}$ by an intrinsic calibration matrix $\mathbf{A}$,

$$\mathbf{q} \simeq \mathbf{A}(t)\mathbf{w}, \quad (5)$$

where $\simeq$ denotes equality up to positive scale. Only the direction of $\mathbf{q}$ is of interest. For simplicity the relationship between $\mathbf{w}$ and $\mathbf{p}$ will be expressed by a single matrix $\mathbf{S}$ representing both intrinsic calibration and orientation of the camera

$$\mathbf{p} \simeq \mathbf{R}(t)\mathbf{A}(t)\mathbf{w} \simeq \mathbf{S}(t)\mathbf{w}. \qquad (6)$$

By normalisation, the matrix $\mathbf{S}$ can be assumed to be an arbitrary real 3×3 matrix whose determinant is 1.

2.1. Structure from Motion

Under *known* viewer motion, the deformation of apparent contours can be used to recover the surface geometry, or structure, cf. (Giblin and Weiss, 1987; Cipolla and Blake, 1992; Vaillant and Faugeras, 1992). These algorithms work well at points where two conditions are fulfilled:

- The contour generators $\Gamma_{\mathbf{c}(t)}$ can be used as coordinate curves for a local coordinate system on the surface U.
- The apparent contours $\gamma_{\mathbf{c}(t)}$ are nonsingular curves.

These two conditions are studied in the next two sections. It is shown that they fail to be fulfilled in three singular cases, two of which refer to the contour generators and one to the apparent contour. It is remarkable that one of these cases can be used to derive the generalised epipolar constraint in Section 3.

When the two conditions hold, a spatio-temporal parametrisation of the image-curve motion can be chosen. The choice of spatial curve parameter s is of course underconstrained. A special parametrisation, called the **epipolar parametrisation**, is naturally matched to the recovery of surface curvature, as was shown in (Cipolla and Blake, 1992). Using the epipolar parametrisation points on consecutive contour generators are matched along the epipolar lines. This is briefly discussed below, but the question of determining structure from known motion will not be continued in this paper. Instead we will focus on the question of determining motion.

2.2. Singularities of the Contour Generator

In this subsection we will investigate the singular cases for the contour generators $\Gamma(t) = \Gamma_{\mathbf{c}(t)}$.

Let $\mathbf{r}_0 \in \Gamma(t_0)$ be a point on the surface U. We want to find conditions that make it possible to describe U locally in a neighbourhood of $\mathbf{r}_0$ by a function $U(s, t)$ in such a way that $U(\cdot, t) = \Gamma(t)$. In other words, we want (s, t) to define a local coordinate system with $\Gamma(t)$ as one of the coordinate curves. In (Giblin and Weiss, 1995, Prop. 3) it is shown that this is possible except in the following singular situations:

- *Singular case A*: If $\mathbf{r}_0$ is a **parabolic point** on U and the view-line $\mathbf{r} - \mathbf{c}$ is in an asymptotic direction, then the contour generator $\Gamma(t_0)$ is *singular*, in fact generally an isolated point or a crossing of two curves.
- *Singular case B*: The tangent plane to U at $\mathbf{r}_0$, is parallel not only to the viewing direction $\mathbf{r}_0 - \mathbf{c}(t_0)$, but also to the velocity of the camera, $\mathbf{c}_t(t_0)$, where the suffix t denotes differentiation. This is the same as requiring that $\mathbf{c}_t(t_0)$ is perpendicular to the normal $\mathbf{n}(\mathbf{r}_0)$ to the surface, i.e., $\mathbf{c}_t(t_0) \cdot \mathbf{n} = 0$. The point $\mathbf{r}_0$ is then called an **epipolar tangency point** and the tangent plane is called an **epipolar tangency plane**. Geometrically, the contour generators on U form an **envelope** at epipolar tangency points—see below—and this prevents their forming part of a coordinate grid since they intersect each other.

The epipolar tangency points, where the plane spanned by the view-line and the camera velocity vector is tangent to U, in general form a curve on U as the camera moves relative to the scene. This curve is called the **frontier** of U relative to the camera motion $\mathbf{c}(t)$. By the above, the **frontier condition** is $\mathbf{c}_t \cdot \mathbf{n} = 0$. Locally, the frontier is the boundary of the 'visible region' swept out by the contour generators. (Only locally because possibly there may come a later time when the contour generators will encroach on the 'far side' of the frontier.)

It is worth looking at this more closely. Consider a surface $U \in \mathbb{R}^3$ parametrised locally by u and v, and a camera motion with projection centre $\mathbf{c}(t)$ parametrised by time t. The condition that the point $\mathbf{r}(u, v) \in U$ lies on the contour generator at time t is simply

$$(\mathbf{r}(u, v) - \mathbf{c}(t)) \cdot \mathbf{n}(u, v) = 0, \qquad (7)$$

where $\mathbf{n}(u, v)$ is the normal to the surface at $\mathbf{r}(u, v)$. Equation (7) can be thought of as defining a family of curves in the u, v parameter plane. This curve is in fact the contour generator on U. The *envelope* of the family of curves, that is the ‘locus of intersections of consecutive curves of the family’, is given by differentiating (7) with respect to t (compare (Bruce and Giblin, 1992, Section 5.3)). This gives the condition

$$\mathbf{c}_t \cdot \mathbf{n} = 0. \tag{8}$$

The points $\mathbf{r}(u, v)$ obtained by eliminating t between (7) and (8) are precisely the points of the *envelope of contour generators* on the surface. This envelope is therefore exactly the frontier as defined above. Over a short period of time, the part of the surface covered by the contour generators is *on one side* of this frontier. This is illustrated in Fig. 3. The epipolar parametrisation breaks down at frontier points as pointed out above, because the contour generators cannot form part of a coordinate system on the surface since they do not cross each other transversally.

Note that if the motion is *linear*, then the frontier degenerates, since $\mathbf{c}_t$ is then a *constant* vector, and the condition (8) does not depend on t. If a point $\mathbf{r}(u, v)$ lies on the frontier at some time t, then this point continues to satisfy the frontier condition at subsequent times and the velocity vector $\mathbf{c}_t$ lies *in* the tangent plane at the isolated frontier point. This is illustrated in Fig. 4. The same holds for any motion which is entirely in the tangent plane to a surface at a particular point on the surface.

The surface cannot be reconstructed by the epipolar parameterisation at these points since the contour generator is locally stationary. However, because frontier points correspond to real, fixed feature points on the surface, which are visible in two ‘consecutive’ views, they can be used to provide a constraint on viewer motion.

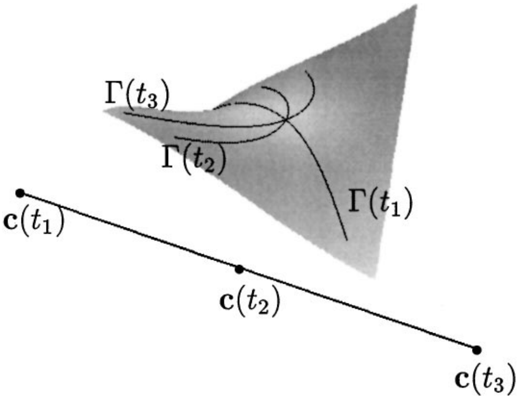

Figure 4. For linear translational motion the frontier degenerates to a point through which all the contour generators pass. The motion $\mathbf{c}(t)$ takes place in the tangent plane at this isolated frontier point.

2.3. Singularities of the Apparent Contours

So far we have discussed degeneracies A and B that are related to the contour generator Γ. Other degeneracies are related to the apparent contour γ, i.e., the image of the contour generator. Let us suppose that the singular cases A and B of the previous section do not hold and therefore that the contour generators form one family of a coordinate grid. The apparent contours in the image sphere will be a family of curves $\gamma(t)$, one for each value of t. If we are to parametrise each apparent contour with a curve parameter s, then we need to avoid a third situation:

- *Singular case C*: The apparent contour $\gamma(t)$ is a *singular* curve (generally with a cusp) when the viewline $\mathbf{r} - \mathbf{c}$ is in an asymptotic direction at $\mathbf{r}$ on U (see (Koenderink, 1990, p. 422) and (Cipolla et al., 1996) where surface geometry is obtained by tracking cusps of apparent contours under known motion). (Note that here it is $\gamma(t)$ that is singular, while in cases A and B it was the contour generator $\Gamma(t)$, that was singular. If the latter case then the apparent contour is automatically singular but with a more degenerate singularity than a cusp, cf. (Koenderink, 1990, p. 458).)

2.4. Epipolar Parametrisation

Assume that neither of the above singular situations hold. Then a spatio-temporal parametrisation $\mathbf{p}(s, t)$ can be chosen. Here s is any *regular* parameter on the apparent contour $\gamma(t) = \{\mathbf{p}(s, t)\colon t = \text{constant}\}$, i.e., $\mathbf{p}_s \neq 0, \forall s$. This induces a parametrisation on the surface, as in (Cipolla and Blake, 1992):

$$\mathbf{r}(s, t) = \mathbf{c}(t) + \lambda \mathbf{p}(s, t).$$

Differentiation with respect to t and scalar multiplication with $\mathbf{n}$ gives

$$\mathbf{r}_t \cdot \mathbf{n} = \mathbf{c}_t \cdot \mathbf{n} + \lambda \mathbf{p}_t \cdot \mathbf{n} + \lambda_t \mathbf{p} \cdot \mathbf{n}.$$

Using $\mathbf{r}_t \cdot \mathbf{n} = 0$ and $\mathbf{p} \cdot \mathbf{n} = 0$ we obtain

$$0 = \mathbf{c}_t \cdot \mathbf{n} + \lambda \mathbf{p}_t \cdot \mathbf{n}. \tag{9}$$

Since $0 < \lambda < \infty$, it follows that $\mathbf{c}_t \cdot \mathbf{n}$ and $\mathbf{p}_t \cdot \mathbf{n}$ are either both zero or both non-zero. Assuming that $\mathbf{c}_t \cdot \mathbf{n} \neq 0$, i.e., assuming that $\mathbf{r}$ is not a frontier point, the depth λ is given by, cf. (Cipolla and Blake, 1992, p. 91)

$$\lambda = -\frac{\mathbf{c}_t \cdot \mathbf{n}}{\mathbf{p}_t \cdot \mathbf{n}}.$$

Again following (Cipolla and Blake, 1992), we can use the *epipolar matching* on the apparent contours to provide a choice of parameter s, by requiring that

$$\mathbf{r}_t \times \mathbf{p} = 0.$$

Introducing the notation $|A|$ for the determinant of the matrix A, the condition can be written

$$|\mathbf{p} \quad \mathbf{c}_t \quad \mathbf{p}_t| = 0.$$

The condition says that the three vectors are coplanar. The epipolar matching thus breaks down at frontier points and whenever the apparent contour $\gamma(t)$ is singular.

Remark. The result for surfaces can also be applied for the case when the objects are *curves* in $\mathbb{R}^3$. A curve may be regarded as a limit case of surfaces with high curvature.

Under viewer motion the contour generators will normally slip over the surface. For any given contact point $\mathbf{r}$ the motion along the ray $\mathbf{p}$ is given by $\mathbf{r}_t$, which depends on the distance and surface curvature (Cipolla and Blake, 1992),

$$\mathbf{r}_t = -\left(\frac{\mathbf{c}_t \cdot \mathbf{n}}{\lambda\kappa^t}\right)\mathbf{p}, \qquad (10)$$

where κ^t is the normal curvature of the surface in the direction of the ray. The speed of the contact point is therefore inversely proportional to the surface curvature. Notice that the velocity $\mathbf{r}(t)$ is large when the curvature is small and vice versa. The velocity is zero when the curvature is infinite. It then follows from (10) that $\mathbf{r}_t = 0$ for all points along a space curve.

3. Generalised Epipolar Constraints

After discussing the principles we will now show how the viewer motion can be calculated from the constraints on the camera motion and the frontier points, described in the previous sections. Remember that frontier points are projections of points on the surface that are visible in both views, i.e., the same point can be identified in both views. However, the frontier points are defined by the epipolar tangency constraint.

In this section we will go through the details and show how frontier points and camera motion can be computed. The presentation will by necessity be elaborate because several different cases have to be considered. The final results can, however, be expressed in a compact form, see Tables 1 and 2.

3.1. *Formulating the Generalised Epipolar Constraints*

Consider the camera centers at two time instants, $\mathbf{c}_1 = \mathbf{c}(t_1)$ and $\mathbf{c}_2 = \mathbf{c}(t_2)$, and consider all tangent planes of the surface U that go through these two camera centers. This will be called the pencil of **epipolar tangency planes** with respect to $\mathbf{c}_1$ and $\mathbf{c}_2$. In each image, the image of the epipolar tangency planes is a pencil of lines, the **epipolar tangency lines**. They all go through a point, the **epipole e**, each line being tangent to an apparent contour. The tangent points on the apparent contours are called the **epipolar tangency points**. Here, lines in the projective image plane $\mathbb{P}^2$ and on the oriented projective plane, or viewing sphere $\mathbb{S}^2$ are planar subspaces defined up to scale or positive scale. Notice in particular that a line in $\mathbb{S}^2$ is a great circle. From the construction the following theorem is obtained, see (Porrill and Pollard, 1991).

Formulation 1 (Coordinate Free). *Given two images, and the epipoles* $\mathbf{e}_1$ *and* $\mathbf{e}_2$*, the pencil of lines through* $\mathbf{e}_1$ *in image one, which are tangent to an apparent contour, and the corresponding pencil of epipolar tangency lines in image two are projectively related.*

This is the generalised epipolar constraint. It can also be expressed in **dual form**, cf. (Coxeter, 1993, p. 15), where the dual of a point $\mathbf{x} \in \mathbb{P}^2$ is the set of all lines that pass through the point. After introduction of coordinates, lines can be represented using homogeneous coordinates as

$$\mathbf{l} = [a \quad b \quad c] \in \mathbb{P}^2.$$

if the equation of the line is

$$\mathbf{l} \cdot \mathbf{x} = 0.$$

Table 1. Summary of relevant motion parameters, number of observable degrees of freedom and generalised epipolar constraints in the case of infinitesimal motion. Five camera models are considered: PT – pure translation, C – Calibrated camera, U – uncalibrated camera, O – orthographic camera, WP – weak perspective. The motion parameters and the observable degrees of freedom *n* are presented as well as the generalised epipolar constraints.

Camera model	Motion params	Number of tangency points needed	Combined tangency and motion constraints
PT	$\mathbf{c}_t$	2	$\operatorname{rank}[\mathbf{c}_t \quad \mathbf{p} \quad \mathbf{p}_s \quad \mathbf{p}_t] = 2$
C	$\mathbf{c}_t, \mathbf{R}_t$	5	$\operatorname{rank}[\mathbf{c}_t \quad \mathbf{q} \quad \mathbf{q}_s \quad \mathbf{R}_t\mathbf{q} + \mathbf{q}_t] = 2$
U	$\mathbf{c}_t, \mathbf{S}_t$	7	$\operatorname{rank}[\mathbf{c}_t \quad \mathbf{w} \quad \mathbf{w}_s \quad \mathbf{S}_t\mathbf{w} + \mathbf{w}_t] = 2$
O	$\mathbf{k}_t, \mathbf{B}_t$	3	$\operatorname{rank}[\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{u}_s \quad \mathbf{B}_t\mathbf{u} + \mathbf{u}_t] = 2$
WP	$\mathbf{k}_t, \mathbf{C}_t$	4	$\operatorname{rank}[\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{u}_s \quad \mathbf{C}_t\mathbf{u} + \mathbf{u}_t] = 2$

Table 2. Summary of relevant motion parameters, number of observable degrees of freedom and generalised epipolar constraints in the discrete case. Five camera models are considered: PT – pure translation, C – Calibrated camera, U – uncalibrated camera, O – orthographic camera, WP – weak perspective. The motion parameters and the observable degrees of freedom *n* are presented as well as the generalised epipolar constraints.

Camera model	Motion params	Number of tangency points needed	Combined tangency and motion constraints
PT	$\Delta\mathbf{c}$	2	$\operatorname{rank}[\Delta\mathbf{c} \quad \mathbf{p}_1 \quad (\mathbf{p}_1)_s \quad \mathbf{p}_2 \quad (\mathbf{p}_2)_s] = 2$
C	$\Delta\mathbf{c}, \Delta\mathbf{R}$	5	$\operatorname{rank}[\Delta\mathbf{c} \quad \mathbf{q}_1 \quad (\mathbf{q}_1)_s \quad \Delta\mathbf{R}\mathbf{q}_2 \quad \Delta\mathbf{R}(\mathbf{q}_2)_s] = 2$
U	$\Delta\mathbf{c}, \Delta\mathbf{S}$	7	$\operatorname{rank}[\Delta\mathbf{c} \quad \mathbf{w}_1 \quad (\mathbf{w}_1)_s \quad \Delta\mathbf{S}\mathbf{w}_2 \quad \Delta\mathbf{S}(\mathbf{w}_2)_s] = 2$
O	$\Delta\mathbf{k}, \Delta\mathbf{B}$	3	$\operatorname{rank}[\Delta\mathbf{k} \quad \mathbf{k} \quad (\mathbf{u}_1)_s \quad (\Delta\mathbf{B}\mathbf{u}_2 - \mathbf{u}_1) \quad \Delta\mathbf{B}(\mathbf{u}_2)_s] = 2$
WP	$\Delta\mathbf{k}, \Delta\mathbf{C}$	4	$\operatorname{rank}[\Delta\mathbf{k} \quad \mathbf{k} \quad (\mathbf{u}_1)_s \quad (\Delta\mathbf{C}\mathbf{u}_2 - \mathbf{u}_1) \quad \Delta\mathbf{C}(\mathbf{u}_2)_s] = 2$

The dual of a point $\mathbf{x}$ is thus a 'line' in parameter space (a, b, c). The dual of a line $\mathbf{l}$ is a point. The dual of a curve is defined as the set of tangent lines to the curve.

Formulation 2 (Dual). *Given two sets of duals of apparent contours and the dual lines* $\mathbf{l}_1$ *and* $\mathbf{l}_2$ *of the epipoles,* $\mathbf{e}_1$ *and* $\mathbf{e}_2$*, in image one and two respectively. The intersection of the line* $\mathbf{l}_1$ *with the dual of the apparent contours in the first image is projectively related to the corresponding intersection of line* $\mathbf{l}_2$ *with the dual of the apparent contours in the second image.*

Using the fundamental matrix $\mathbf{F}$, see (Thompson, 1959; Stefanovic, 1973; Faugeras et al., 1992), and introducing coordinate system in both images, the generalised epipolar constraint can be expressed by the fundamental matrix. Notice that the two epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$ can be obtained as the left and right null-space of the fundamental matrix, i.e., $\mathbf{e}_1^{\mathrm{T}}\mathbf{F} = 0$ and $\mathbf{F}\mathbf{e}_2 = 0$.

Formulation 3 (Fundamental Matrix). *Given two images, a coordinate system in each image, and the fundamental matrix* $\mathbf{F}$. *The corresponding epipolar tangency points* $\mathbf{w}_1$ *and* $\mathbf{w}_2$*, fulfill*

$$\begin{aligned} \mathbf{w}_1^{\mathrm{T}}\mathbf{F}\mathbf{w}_2 &= 0, && \textit{motion constraint}, \\ \det[\mathbf{e}_1 \quad \mathbf{w}_1 \quad (\mathbf{w}_1)_s] &= 0, && \textit{tangency constraint}, \\ \det[\mathbf{e}_2 \quad \mathbf{w}_2 \quad (\mathbf{w}_2)_s] &= 0, && \textit{tangency constraint}, \end{aligned} \tag{11}$$

where subscript s denotes differentiation with respect to a parametrisation of the apparent contour.

Note that the tangency constraint also can be written $(\mathbf{w}_1)_s^{\mathrm{T}}\mathbf{F}\mathbf{w}_2 = 0$ and $\mathbf{w}_1^{\mathrm{T}}\mathbf{F}(\mathbf{w}_2)_s = 0$ respectively. The constraint can also be formulated using projection matrices. This requires the introduction of coordinates in both images and also for the object.

Formulation 4 (Projection Matrix). *Given two images, formed by projection matrices* $P_1 = \mathbf{S}_1^{-1}[I \;\; -\mathbf{c}_1]$,

$P_2 = \mathbf{S}_2^{-1}[I - \mathbf{c}_2]$ *respectively, where* $\mathbf{S}_1$, $\mathbf{S}_2$ *and* $\mathbf{I}$ *are* 3×3 *matrices and* $\mathbf{c}_1$ *and* $\mathbf{c}_2$ *are* 3×1 *vectors. The corresponding epipolar tangency points* $\mathbf{w}_1$ *and* $\mathbf{w}_2$ *fulfill*

$$\operatorname{rank}\,[\mathbf{c}_2 - \mathbf{c}_1 \quad \mathbf{S}_1\mathbf{w}_1 \quad \mathbf{S}_1(\mathbf{w}_1)_s \quad \mathbf{S}_2\mathbf{w}_2 \quad \mathbf{S}_2(\mathbf{w}_2)_s] = 2. \tag{12}$$

Formulations 1 and 2, have the advantage of being elegant and short. However, it is not apparent how to generalise them to other camera cases. Formulation 3, also has its advantages. The image coordinate system is often given a priori, or can be chosen at will. Once this is done the whole epipolar geometry is described by the fundamental matrix **F**. This matrix can thus be used to parametrise the problem uniquely.

Formulation 5 depends not only on the choice of coordinate system in each image, but also on the object coordinate system. A change in object coordinates changes the projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, while the epipolar constraint still holds. The matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ which have 22 degrees of freedom can only be determined up to an unknown projective transformation (15 degrees of freedom), leaving at least 7 degrees of freedom. The fundamental matrix has the same degrees of freedom, since **F** is a 3×3 matrix defined up to scale with determinant zero. In the sequel, the term **observable degrees of freedom** will be used to denote this number, i.e., the number of degrees of freedom in the projection matrices minus the number of degrees of freedom in the unknown object coordinate system.

Despite this difficulty in defining observable degrees of freedom, the coordinate dependent formulation has the advantage of being compact. Another advantage is that all camera cases can be expressed in a uniform way. Notice that we are interested in computing two quantities, the camera positions and the frontier points. All constraints are expressed by (12). If the camera positions are known, the frontier points can be calculated, and vice versa. It is useful to think of the constraints in (12) as being of two types:

1. **Epipolar tangency constraints.** These link motion parameters to the position of the frontier point in an image. The constraints allow us to select a discrete number of points in each image contour.
2. **Motion constraints.** The motion constraints, on the other hand, link viewer motion to the image motion of epipolar tangency points. This is possible because the epipolar tangency points correspond to fixed features visible in both views. The motion constraint involves the epipole (direction of translation) and changes in orientation.

Remark. The actual tangency points may not be well defined at contour points of small curvature. This does not affect the stability of the motion estimate as they rely on the tangency planes only. These are determined by the normal of the apparent contour γ and the epipole. Even at points of high curvature the tangency plane is well defined.

Although the motion constraints can be formulated using the set of tangency planes, it is easier to express them using the set of frontier points. This will be done in the following. This also highlights the similarities to the well known epipolar constraints for points, cf. (Faugeras, 1992). Keep in mind that each tangency point defines a tangency plane.

3.2. Central Projection Models

We will structure the analysis as follows:

- Known rotation and internal calibration.
- Calibrated camera. Unknown rotation but known internal calibration.
- Uncalibrated camera.

If the rotation of the camera is known, then the image point $\mathbf{p} \in \mathbb{S}^2$ gives the direction from the camera center to the apparent contour. In the calibrated camera case, $\mathbf{q} \in \mathbb{S}^2$ will be used to denote image points in the camera coordinate system. It is related to the absolute direction $\mathbf{p}$ as $\mathbf{p} = \mathbf{R}\mathbf{q}$. In the uncalibrated camera case, the image point in homogeneous coordinates $\mathbf{w}$ has to be corrected with an unknown internal calibration matrix $\mathbf{A}$ in order to get the direction $\mathbf{q}$.

Oriented homogeneous coordinates for vectors and matrices will be used. Two vectors or matrices will be considered equal if they are a positive multiple of each other. This simplifies the notation considerably. Remember that $\mathbf{p} = \mathbf{R}\mathbf{q}$ and also $\mathbf{p} \simeq \mathbf{R}\mathbf{A}\mathbf{w} \simeq \mathbf{S}\mathbf{w}$. Sometimes a specific scaled representations of vectors and matrices is needed. In this case all vectors are normalised to have unit length and all matrices have unit determinant.

3.2.1. Discrete Motion with Known Rotation. Let $\mathbf{p}_1(s)$ denote the apparent contour in image 1 with curve

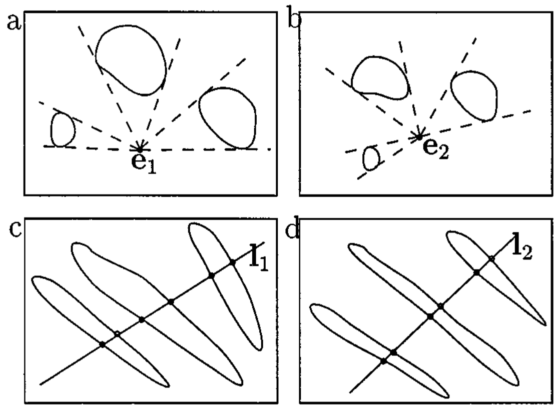

Figure 5. (a) and (b): The two figures illustrate the configuration of apparent contours, the epipole and the epipolar tangency lines in two images. The two pencils of epipolar tangency lines are projectively equivalent. (c) and (d): The figure illustrate the dual of (a) and (b). The dual of an apparent contour is a curve, the dual of the epipole is a line and the dual of each epipolar tangency line is a point. The duals of the set of epipolar tangency lines are projectively equivalent. The figure is only schematic.

parameter s and let $\mathbf{p}_2(s)$ be the corresponding apparent contour in image 2. The point $\mathbf{r}$ where the epipolar tangency point is tangent to the surface U belongs to both contour generators, as illustrated in Fig. 2. The normal $\mathbf{n}$ to the surface at this point is orthogonal not only to $\mathbf{p}_1$ and $\mathbf{p}_2$ but also to their image tangents, $(\mathbf{p}_1)_s$ and $(\mathbf{p}_2)_s$, and to the direction of motion $\mathbf{c}_2 - \mathbf{c}_1$. This can be written as the **generalised epipolar constraints**,

$$\operatorname{rank}\,[\mathbf{c}_2 - \mathbf{c}_1 \quad \mathbf{p}_1 \quad (\mathbf{p}_1)_s \quad \mathbf{p}_2 \quad (\mathbf{p}_2)_s] = 2. \qquad (13)$$

The five column vectors in this matrix are perpendicular to the normal $\mathbf{n}$. Notice that the rank constraints involve both curve parameters s_1 and s_2 and the motion parameters $(\mathbf{c}_1, \mathbf{c}_2)$.

The constraints (13) can be used in several ways. Firstly, once the direction of motion $\Delta\mathbf{c} = \mathbf{c}_2 - \mathbf{c}_1$ is known the first image of the epipolar tangency point can be found by searching for the curve parameter s_1 which gives $\det[\Delta\mathbf{c} \quad \mathbf{p}_1(s_1) \quad (\mathbf{p}_1)_s(s_1)] = 0$, and similarly for the second image, see (Carlsson, 1994). Secondly, when the image of the epipolar tangency points is known, the condition $\det[\Delta\mathbf{c} \quad \mathbf{p}_1 \quad \mathbf{p}_2] = 0$ can be checked.

The motion parameters $(\mathbf{c}_1, \mathbf{c}_2)$ have six degrees of freedom. They can, however, only be found up to an unknown coordinate transformation consisting of translation and change of scale which gives 4 degrees of freedom. This leaves 2 observable degrees of freedom. A canonical parametrisation can be chosen with $|\Delta\mathbf{c}| = 1$, so that the condition (13) becomes

$$\operatorname{rank}\,[\Delta\mathbf{c} \quad \mathbf{p}_1 \quad (\mathbf{p}_1)_s \quad \mathbf{p}_2 \quad (\mathbf{p}_2)_s] = 2. \qquad (14)$$

3.2.2. Discrete Motion with Calibrated Camera.

The constraints in the calibrated camera case can be derived by introducing $\mathbf{p} = \mathbf{Rq}$ in (13). This gives:

$$\operatorname{rank}\,[\mathbf{c}_2 - \mathbf{c}_1 \quad \mathbf{R}_1\mathbf{q}_1 \quad \mathbf{R}_1(\mathbf{q}_1)_s \quad \mathbf{R}_2\mathbf{q}_2 \quad \mathbf{R}_2(\mathbf{q}_2)_s] = 2. \qquad (15)$$

The motion parameters $(\mathbf{R}_1, \mathbf{c}_1, \mathbf{R}_2, \mathbf{c}_2)$ have 12 degrees of freedom, but they can only be determined up to a similarity transformation, which has 7 degrees of freedom. This leaves 5 observable degrees of freedom. A canonical parametrisation is obtained by choosing $\mathbf{R}_1 = I$, $\mathbf{c}_1 = 0$ and $|\Delta\mathbf{c}| = |\mathbf{c}_2 - \mathbf{c}_1| = 1$. Using $\Delta\mathbf{R} = \mathbf{R}_2$ we get

$$\operatorname{rank}\,[\Delta\mathbf{c} \quad \mathbf{q}_1 \quad (\mathbf{q}_1)_s \quad \Delta\mathbf{R}\mathbf{q}_2 \quad \Delta\mathbf{R}(\mathbf{q}_2)_s] = 2. \qquad (16)$$

3.2.3. Discrete Motion with Uncalibrated Camera.

The constraints in the uncalibrated camera case can be derived by introducing $\mathbf{p} \simeq \mathbf{Sw}$ in (13). This gives:

$$\operatorname{rank}\,[\mathbf{c}_2 - \mathbf{c}_1 \quad \mathbf{S}_1\mathbf{w}_1 \quad \mathbf{S}_1(\mathbf{w}_1)_s \quad \mathbf{S}_2\mathbf{w}_2 \quad \mathbf{S}_2(\mathbf{w}_2)_s] = 2. \qquad (17)$$

The motion parameters $(\mathbf{S}_1, \mathbf{c}_1, \mathbf{S}_2, \mathbf{c}_2)$ have 22 degrees of freedom, but they can only be determined up to a projective transformation, which has 15 degrees of freedom. This leaves 7 observable degrees of freedom. Using $\mathbf{S}_1 = I$, $\mathbf{c}_1 = 0$ and $|\Delta\mathbf{c}| = |\mathbf{c}_2 - \mathbf{c}_1| = 1$, we remove some of the arbitrariness, but three degrees of freedom are left. Using $\Delta\mathbf{S} = \mathbf{S}_2$ we get

$$\operatorname{rank}\,[\Delta\mathbf{c} \quad \mathbf{w}_1 \quad (\mathbf{w}_1)_s \quad \Delta\mathbf{S}\mathbf{w}_2 \quad \Delta\mathbf{S}(\mathbf{w}_2)_s] = 2. \qquad (18)$$

The parameter ambiguity in $\Delta\mathbf{S}$ can be understood by doing a projective transformation by the matrix $\Delta\mathbf{S}$ of the viewing sphere of image 2 and then by choosing a coordinate system so that the direction of translation is along the x-axis. Compare this with the standard rectification in Fig. 7. The set of epipolar tangency planes is now invariant under the projective transformations

$$S = e^{x_1 s_1 + x_2 s_2 + x_3 s_3},$$

where e is the matrix exponential and

$$s_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad s_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$s_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

are a basis for corresponding Lie-algebra. The parameter ambiguity in $\Delta\mathbf{S}$ corresponds to the three parameter ambiguity in choosing the plane at infinity, (see (Luong and Vieville, 1994)).

3.2.4. Infinitesimal Motion with Known Rotation. In some cases it is of interest to compute the infinitesimal motion of the camera. The constraints can be derived from the previous results by a limiting procedure. It follows from Eq. (13) that

$$\begin{aligned} &\text{rank}\,[\Delta\mathbf{c} \quad \mathbf{p}_1 \quad (\mathbf{p}_1)_s \quad \mathbf{p}_2 \quad (\mathbf{p}_2)_s] \\ &= \text{rank}\,[\Delta\mathbf{c}/\Delta t \quad (\mathbf{p}_1)_s \quad (\mathbf{p}_2 - \mathbf{p}_1)/\Delta t \quad (\mathbf{p}_2)_s] \\ &\to \text{rank}\,[\mathbf{c}_t \quad \mathbf{p} \quad (\mathbf{p})_s \quad \mathbf{p}_t \quad (\mathbf{p})_s] = 2, \quad \text{as } t \to 0. \end{aligned} \tag{19}$$

Thus the infinitesimal generalised epipolar constraint is

$$\text{rank}\,[\mathbf{c}_t \quad \mathbf{p} \quad (\mathbf{p})_s \quad \mathbf{p}_t] = 2. \tag{20}$$

Remark. Note that the frontier condition $\mathbf{c}_t \cdot \mathbf{n} = 0$ is equivalent to $\mathbf{p}_t \cdot \mathbf{n} = 0$, see (9), which is the condition for the apparent contours to form an *envelope in the image*.

As in the discrete time case there are 2 observable degrees of freedom. A unique parametrisation is obtained by requiring that $|\mathbf{c}_t| = 1$.

3.2.5. Infinitesimal Motion with Calibrated Camera. The constraints in the calibrated camera case are obtained by introducing $\mathbf{p} = \mathbf{Rq}$ in (20). This gives

$$\begin{aligned} &\text{rank}\,[\mathbf{c}_t \quad \mathbf{Rq} \quad (\mathbf{Rq})_s \quad (\mathbf{Rq})_t] \\ &= \text{rank}\,[\mathbf{c}_t \quad \mathbf{Rq} \quad \mathbf{R}(\mathbf{q})_s \quad \mathbf{R}_t\mathbf{q} + \mathbf{Rq}_t] = 2. \end{aligned} \tag{21}$$

The motion parameters $(\mathbf{R}, \mathbf{c}, \mathbf{R}_t, \mathbf{c}_t)$ have 12 degrees of freedom, but they can only be determined up to a similarity transformation, which has 7 degrees of freedom. This leaves 5 observable degrees of freedom. A unique parametrisation is obtained by $\mathbf{R} = I$, $\mathbf{c} = 0$ and $|\mathbf{c}_t| = 1$. This gives,

$$\text{rank}\,[\mathbf{c}_t \quad \mathbf{q} \quad (\mathbf{q})_s \quad \mathbf{R}_t\mathbf{q} + \mathbf{q}_t] = 2. \tag{22}$$

3.2.6. Infinitesimal Motion with Uncalibrated Camera. The constraints in the uncalibrated camera case are found by introducing $\mathbf{p} = \mathbf{Sw}$ in (20). This gives,

$$\begin{aligned} &\text{rank}\,[\mathbf{c}_t \quad \mathbf{Sw} \quad (\mathbf{Sw})_s \quad (\mathbf{Sw})_t] \\ &\quad = \text{rank}\,[\mathbf{c}_t \quad \mathbf{Sw} \quad \mathbf{S}(\mathbf{w})_s \quad \mathbf{S}_t\mathbf{w} + \mathbf{Sw}_t] = 2. \end{aligned} \tag{23}$$

The motion parameters $(\mathbf{S}, \mathbf{c}, \mathbf{S}_t, \mathbf{c}_t)$ have 22 degrees of freedom, but they can only be determined up to a projective transformation, which has 15 degrees of freedom, leaving 7 observable degrees of freedom. By choosing $\mathbf{S} = I$, $\mathbf{c} = 0$ and $|\mathbf{c}_t| = 1$, we obtain

$$\text{rank}\,[\mathbf{c}_t \quad \mathbf{q} \quad (\mathbf{q})_s \quad \mathbf{S}_t\mathbf{q} + \mathbf{q}_t \quad (\mathbf{q})_s] = 2. \tag{24}$$

As in the discrete time case, three unobservable degrees of freedom remain. To understand this, choose coordinates so that $\mathbf{c}_t$ is parallel to the x-axis. The constraints then only involve the last two components of $\mathbf{S}_t\mathbf{w}$. The first row of $\mathbf{S}_t$ is thus not observable. The choice of the first row of $\mathbf{S}_t$ corresponds to the choice of the plane at infinity.

3.3. Parallel Projection Models

The same analysis can be made for the parallel projection camera model. In this case points at position $\mathbf{r}$ are projected onto the image plane along the same direction $\mathbf{k}$. A point $\mathbf{r}$ on the surface U lies on the contour generator if

$$\mathbf{k} \cdot \mathbf{n}(\mathbf{r}) = 0.$$

The normal direction is uniquely constrained by

$$\mathbf{k}(t_1) \cdot \mathbf{n}(\mathbf{r}) = 0, \quad \mathbf{k}(t_2) \cdot \mathbf{n}(\mathbf{r}) = 0,$$

if the two directions of projection $\mathbf{k}(t_1)$ and $\mathbf{k}(t_1)$ are different. The direction $\mathbf{k}$ of the projection plays a similar role in the parallel camera model as the focal point $\mathbf{c}$ does in the perspective camera models.

Introduce $\mathbf{k}_1 = \mathbf{k}(t_1)$, $\mathbf{k}_2 = \mathbf{k}(t_2)$, $\mathbf{u}_1 = \mathbf{u}(\cdot, t_1)$ and $\mathbf{u}_2 = \mathbf{u}(\cdot, t_2)$. The normal $\mathbf{n}(\mathbf{r})$ of the surface is orthogonal not only to $\mathbf{k}_1$ and $\mathbf{k}_2$ but also to the image tangents $(\mathbf{u}_1)_s$ and $(\mathbf{u}_2)_s$ and the vector $\mathbf{u}_2 - \mathbf{u}_1$. This can be written

$$\operatorname{rank}\,[\mathbf{k}_2 \quad \mathbf{k}_1 \quad \mathbf{u}_2 - \mathbf{u}_1 \quad (\mathbf{u}_1)_s \quad (\mathbf{u}_2)_s] = 2. \tag{25}$$

The infinitesimal time constraint is obtained by a limiting procedure

$$\operatorname{rank}\,[\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{u}_t \quad \mathbf{u}_s] = 2. \tag{26}$$

In the orthographic camera model the image coordinate system is known up to a Euclidean transformation. Thus we obtain

$$\operatorname{rank}\,[\mathbf{k}_2 \quad \mathbf{k}_1 \quad \mathbf{B}_2\mathbf{u}_2 - \mathbf{B}_1\mathbf{u}_1 \quad \mathbf{B}_1(\mathbf{u}_1)_s \quad \mathbf{B}_2(\mathbf{u}_2)_s] = 2 \tag{27}$$

for discrete motion and

$$\operatorname{rank}\,[\mathbf{k}_t \quad \mathbf{k} \quad (\mathbf{B}\mathbf{u})_t \quad \mathbf{B}\mathbf{u}_s] = 2 \tag{28}$$

for continuous motion, where $\mathbf{B}$ is a 3×3 matrix representing planar Euclidean transformations. In the weak perspective camera model the image coordinate systems are known up to a similarity transformation $\mathbf{C}$. By choosing coordinate system such that $\mathbf{B}_1 = I$ and $\mathbf{k}_1 = [0 \;\; 0 \;\; 1]$ and denoting $\Delta\mathbf{B} = \mathbf{B}_2$ and $\Delta\mathbf{k} = \mathbf{k}_2$, and similar for the other cases we obtain,

$$\operatorname{rank}\,[\Delta\mathbf{k} \quad \mathbf{k} \quad \Delta\mathbf{B}\mathbf{u}_2 - \mathbf{u}_1 \quad (\mathbf{u}_1)_s \quad \Delta\mathbf{B}(\mathbf{u}_2)_s] = 2,$$
$$\operatorname{rank}\,[\Delta\mathbf{k} \quad \mathbf{k} \quad \Delta\mathbf{C}\mathbf{u}_2 - \mathbf{u}_1 \quad (\mathbf{u}_1)_s \quad \Delta\mathbf{C}(\mathbf{u}_2)_s] = 2,$$

for discrete motion and

$$\operatorname{rank}\,[\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{B}_t\mathbf{u} + \mathbf{u}_t \quad \mathbf{u}_s] = 2, \tag{29}$$
$$\operatorname{rank}\,[\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{C}_t\mathbf{u} + \mathbf{u}_t \quad \mathbf{u}_s] = 2, \tag{30}$$

for infinitesimal motion, where

$$\mathbf{k} = [\,0 \quad 0 \quad 1\,],$$
$$\Delta\mathbf{k} = \simeq [\cos\theta \quad \sin\theta \quad 0]^{\mathrm{T}},$$
$$\Delta\mathbf{B} = \begin{bmatrix} \cos(b_1) & \sin(b_1) & b_2 \\ -\sin(b_1) & \cos(b_1) & b_3 \\ 0 & 0 & 1 \end{bmatrix},$$
$$\Delta\mathbf{C} = \begin{bmatrix} c_1\cos(c_2) & c_1\sin(c_2) & c_3 \\ -c_1\sin(c_2) & c_1\cos(c_2) & c_4 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{k}_t = \simeq [\,\cos\theta \quad \sin\theta \quad 0\,]^{\mathrm{T}},$$
$$\mathbf{B}_t = \begin{bmatrix} 0 & b_1 & b_2 \\ -b_1 & 0 & b_3 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\mathbf{C}_t = \begin{bmatrix} c_1 & c_2 & c_3 \\ -c_2 & c_1 & c_4 \\ 0 & 0 & 0 \end{bmatrix}.$$

In all of these cases all parameters but one are observable.

4. Implementation

In this section we will discuss some details of the computation of camera motion. This involves detection of apparent contours, determination of an initial estimate of motion parameters and refinements of the estimates.

Notice that there is not a closed form solution to the problem. The epipoles are needed to find the frontier points. The frontier points are needed to find the epipoles.

We have developed algorithms for several different camera models. The idea is to obtain an initial estimate of motion and then use optimisation techniques to obtain the final solution. The algorithms can be divided into two groups: continuous versus discrete time.

Algorithm for continuous time parameters

1. Track the contours.
2. Tesselate each focus of expansion (infinitesimal epipole).
3. For each focus of expansion calculate optimal motion parameters and measure residual.
4. Select the best focus of expansion as initial motion estimate.
5. Calculate scaled residuals, likelihood and their derivatives with respect to motion parameters.
6. Quit if residuals are small.
7. Otherwise update motion parameters and goto 5.

Algorithm for discrete time parameters

1. Track the contours.
2. Get initial motion estimate, for example using a continuous time approximation.
3. Calculate scaled residuals, likelihood and their derivatives with respect to motion parameters.

4. Quit if residuals are small.
5. Otherwise update motion parameters and goto 3.

These steps will be commented upon here and illustrated in Section 6.

4.1. Extraction and Tracking of Apparent Contours

An important part of the calculation of motion from the deformation of apparent contour is the extraction and tracking of the contour. This is a difficult practical problem which has received considerable attention, see (Blake and Yuille, 1992). The notion of **B-spline snake** has been used for this purpose, see (Cipolla and Blake, 1990). Roughly speaking, a snake see (Kass et al., 1987) is a parametrised curve, in this case as B-splines, whose parameters are changed dynamically to fit the contour. The spline curve wriggles to adapt the image, thus resembling a snake. The curve is represented as a collection of B-spline segments, where each segment is represented by four control points. These points generate a segment of the contour, see (Foley et al., 1990, p. 493). This representation has several nice properties. The contour obtained by joining the segment generated by control points (1,2,3,4) and the segment generated by control points (2,3,4,5) is automatically C^2, unless some of the control points coincide. Closed contours are easily represented using the control points cyclically.

The B-spline snake is matched to the contour in two steps. Euclidean transformations are first used. This ensures a fast, robust, but rough positioning of the snake in the new image, cf. Fig. 6(a) and (b). The snake is then deformed to match the new image. Figure 6 illustrate this for one of the contours. The procedure is explained in more detail in (Curwen and Blake, 1992).

To deform the B-spline snakes, a subpixel edge detectors is used, that not only give the location of the contour but also a confidence interval in the normal direction of the curve. This is done with the technique described in (Åström and Heyden, 1999). For clear, well defined edges, like the ones in Fig. 6, the individual edge positions can be found with a standard deviation of about on tenth of a pixel. This uncertainty measure is important in estimating motion parameters. Different frontier points are weighted according to the uncertainty in their positioning.

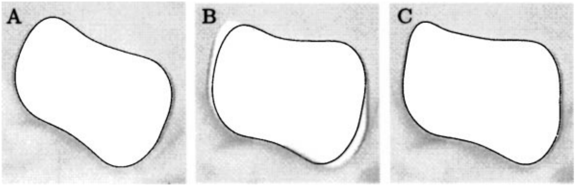

Figure 6. The B-spline snake (a) is used as a template to track the contour in the next image. A rough positioning is found by allowing rigid motion of the snake (b). The new snake is then found by allowing the snake to deform (c).

A rough estimate of point correspondences are obtained as a by-product of the snake type tracking. These correspondences can be used to calculate an initial estimate of motion parameters as described in the next section.

4.2. Initial Hypothesis of Motion

An initial estimate of the motion parameters is needed in order to use the generalised epipolar constraints. There are a number of different ways to obtain these:

1. *Point matches*: In most cases it is useful to match points as well as contours. The points can be used to estimate motion parameters with conventional methods, e.g., the linear eight point method (Longuet-Higgins, 1981) or non-linear methods (Luong et al., 1993).
   Approximate point matches can also be obtained by matching points with high curvature in the image or by using the centroid of the matched contours.
   The B-spline snake tracker can also be used to obtain approximate point correspondences. Individual points on the apparent contour are first identified through the rigid motion of the template as illustrated in Fig. 6, and then through the deformation of the contour in the normal direction.
2. *Motion sensors*: In some situations, partial knowledge of the motion can be obtained by other means. The camera might be mounted on a robot with sensors that give approximate and/or partial knowledge of viewer motion.
3. *Prediction*: If viewer motion is smooth it might be possible to predict motion parameters from motion history.

The problem of finding good initial estimates is an important and difficult one. The above suggestions indicate some possible techniques. In the experiements, additional suggestions are given and tested. These methods work reasonably well for the type of data and motion present in the experiments. A more thorough testing of methods for finding the initial estimates is needed, but is outside the scope of this paper.

4.3. Maximum Likelihood Estimate

The maximum likelihood method is a natural way to estimate the motion parameters given noisy input data. It has several advantages, and is relatively easy to apply. In the sequel $\mathbf{m}$ will be used as an abstract variable for the motion parameter and $\mathcal{M}$ will be used for the motion parameter manifold. The general principle is first described.

1. Create a residual function $\alpha_i = \alpha_i(\mathbf{m})$. This is described below.
2. Calculate the joint conditional distribution $f_n(\alpha \mid \mathbf{m})$ of the residuals given the motion parameter $\mathbf{m}$.
3. Define the likelihood function $L(\mathbf{m}) = f_n(\alpha \mid \mathbf{m})$ as a function of $\mathbf{m} \in \mathcal{M}$.
4. The maximum likelihood estimate $\hat{\mathbf{m}}$ is the parameter which maximises the likelihood $L(\mathbf{m})$ over $\mathcal{M}$.

To simplify the minimisation it is often assumed that the residuals α_i are independent and Gaussian with zero mean and standard deviation σ_i. This is a reasonable assumption if the images of the frontier points are not too close to each other. The likelihood function is then

$$L = \prod \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\alpha_i^2/2\sigma_i^2}.$$

Maximising the likelihood L is then almost the same as minimising

$$g(\mathbf{m}) = \sum \frac{\alpha_i^2(\mathbf{m})}{\sigma_i^2(\mathbf{m})}. \tag{31}$$

The estimate $\hat{\mathbf{m}}$ is the motion parameters that minimise this weighted sum of squared residuals, i.e.,

$$\hat{\mathbf{m}} = \operatorname{argmin} g(\mathbf{m}). \tag{32}$$

Although the method is straightforward, some points need careful consideration. The functions $\alpha_i(\mathbf{m})$ and $\sigma_i(\mathbf{m})$, must be determined and suitable optimisation methods must be found. To do this we will consider three cases: central projection and orthographic projection in the discrete time case and the continuous time case.

4.3.1. Residuals for Discrete Time Central Projection. We will first define the residuals $\alpha_i(\mathbf{m})$ for the discrete time case with central projection and an uncalibrated camera. Consider two images. Let $\Delta\mathbf{S}$ and $\Delta\mathbf{c}$ be the incremental motion parameters. Recall that $\Delta\mathbf{S}$ is the change in generalised orientation of the camera and $\Delta\mathbf{c}$ is the change in position. These motion parameters can be used to rectify the camera, as illustrated in Fig. 7. Two images are shown in Fig. 7(a) and (b). Each of them is mapped onto the viewing sphere in Fig. 7(c) and (d). The image c remains unchanged in e but the image d is projectively transformed with $\Delta\mathbf{S}$, to compensate

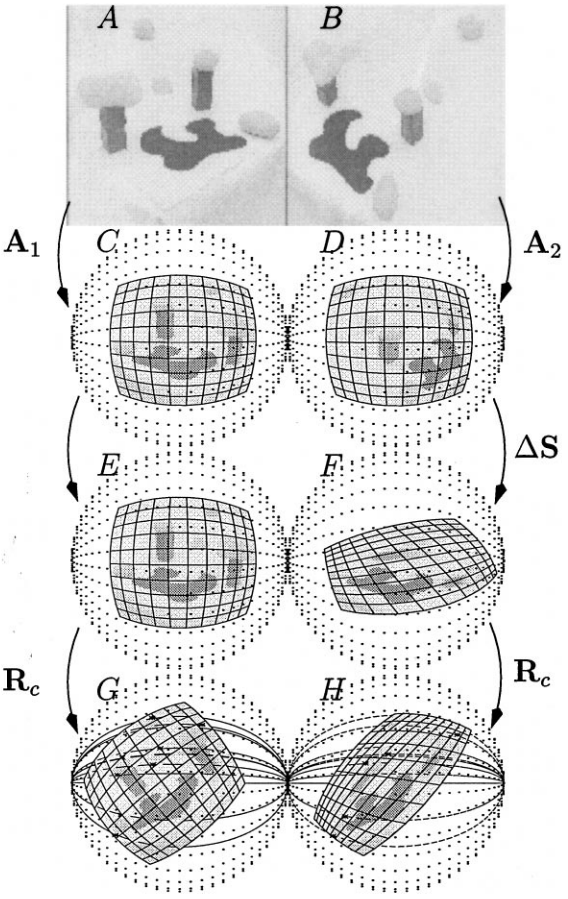

Figure 7. Rectification of uncalibrated images. (a) and (b): The figures show two images. (c) and (d): These are projected onto the viewing sphere using approximate intrinsic calibration matrices. (e) and (f): Image (d) is then projectively transformed by the matrix $\Delta\mathbf{S}$. (g) and (h): Both images are rotated with matrix $\mathbf{R}_c$, so that the direction of motion $\Delta\mathbf{c}$ is along the x-axis. After rectification the epipolar tangency planes all intersect at the x-axis. The two sets of epipolar tangency planes should be equal. The angular difference is used as a residual.

for changes in orientation and internal calibration. The images e and f are then transformed with the same rotation matrix $\mathbf{R}_c$, so that the direction of motion is along the x-axis. Having made the transformation it now remains to find the frontier points in the images. According to the generalised epipolar constraint (18), the frontier points in image c are given by

$$|\Delta\mathbf{c} \quad \mathbf{w}_1 \quad (\mathbf{w}_1)_s| = 0,$$

and those in image d by

$$|\Delta\mathbf{c} \quad \Delta\mathbf{S}\mathbf{w}_2 \quad \Delta\mathbf{S}(\mathbf{w}_2)_s| = 0.$$

After rectification this simplifies to

$$|\mathbf{e}_1 \quad \mathbf{w}_1 \quad (\mathbf{w}_1)_s| = 0,$$

and

$$|\mathbf{e}_1 \quad \mathbf{w}_2 \quad (\mathbf{w}_2)_s| = 0,$$

where $\mathbf{e}_1 = [1 \quad 0 \quad 0]^{\mathrm{T}}$. When the corresponding frontier points have been found, the epipolar tangency planes through the x-axis and the frontier points in g and h should be identical. The residual α_i is then defined as the angular difference between the planes (Fig. 8). The standard deviation σ_i of each residual is estimated using the edge detector. The transformation from contour errors to angular errors in the epipolar tangency planes is straightforward. It will not be described explicitly here, see (Faugeras, 1993). It has now been described how to compute α_i and σ_i. Summing over all frontier points gives the loss function (31). However, efficient minimisation of the loss function requires the derivatives

$$\frac{\partial\alpha_i}{\partial\mathbf{m}} \quad \text{and} \quad \frac{\partial\sigma_i}{\partial\mathbf{m}}.$$

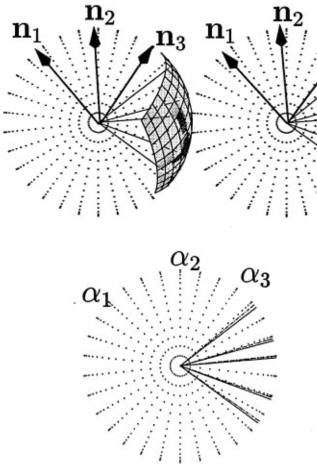

Figure 8. Two sets of epipolar tangency planes are calculated from two images. These two sets should ideally be identical. The residual is defined as the angular difference α_i between corresponding epipolar tangency planes after rectification.

The expressions for these derivatives are complicated. The analytical calculations have been checked numerically and with computer algebra. An additional complication arises because $\mathcal{M}$ is a manifold. It is thus necessary to introduce new local coordinates at each iteration. Second derivatives have also been used to implement the Newton-Raphson iteration for some of the camera cases.

4.3.2. Residuals for Discrete Time Parallel Projection. The parallel projection case is similar to the central projection case. The motion parameters are used to rectify the image pair. The epipolar tangency planes are calculated through the epipolar tangency constraints (Fig. 9).

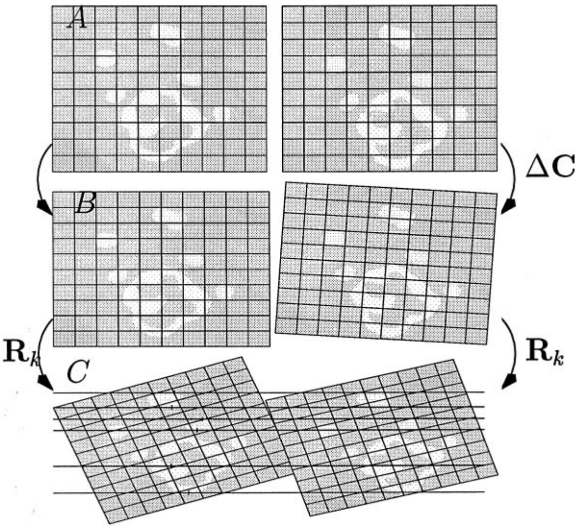

Figure 9. The case of discrete motion with weak perspective cameras. (a) and (b): The figures shows two images. (c) and (d): Image (b) is transformed with similarity transformation $\Delta\mathbf{C}$. (e) and (f): Both images are rotated so that the direction of motion is along the x-axis. The two sets of epipolar tangency planes should be equal. The difference (in y-direction) is used as residual.

The distance between the parallel epipolar tangency planes is used as a residual α_i. The residual is scaled with respect to its standard deviation σ_i. The residual variance, due to edge localisation error, is changed in these transformation. These effects must be taken into account.

4.3.3. Residuals for Infinitesimal Time. In the infinitesimal case, the direction of viewer motion $\mathbf{c}_t$ is used as an infinitesimal epipole, or the focus of expansion. The tangency constraint is then used to find the epipolar tangency planes and the corresponding frontier points. For example, in the calibrated case we have

$$|\mathbf{c}_t \quad \mathbf{q} \quad \mathbf{q}_s| = 0.$$

Each plane defines a normal $\mathbf{n}$. The motion constraint is then simply

$$\mathbf{n} \cdot (\mathbf{R}_t \mathbf{q} + \mathbf{q}_t) = 0.$$

It is reasonably to use

$$\alpha = \mathbf{n} \cdot (\mathbf{R}_t \mathbf{q} + \mathbf{q}_t)$$

as residual. The maximum likelihood estimate is obtained by minimising (31), i.e.,

$$g(\mathbf{m}) = \sum_i \frac{\alpha_i^2(\mathbf{m})}{\sigma_i^2(\mathbf{m})}. \tag{33}$$

It seams reasonable to assume that errors in α_i are mostly due to the errors in $\mathbf{q}_t$. If so the following approximation holds

$$\sigma_i = \sigma[\alpha_i] \approx \sigma[\mathbf{n}_i \cdot \mathbf{q}_t].$$

These standard deviations are obtained from the subpixel edge detector routines. Furthermore, it seams reasonable to assume that this standard deviation is approximately constant around each frontier points. Using these approximations g is in fact quadratic in $\mathbf{R}_t$ so that the minimisation with respect to $\mathbf{R}_t$ can be found by linear methods.

The implementation of the infinitesimal case is simpler than the discrete time case. The major reasons for this are the following

- The derivative of the matrix rotation, $\mathbf{R}_t$, and similarly for $\mathbf{S}_t$, $\mathbf{B}_t$ and $\mathbf{C}_t$, is an element of a linear space, a Lie algebra, as opposed to the rotation operator $\mathbf{R}$. The same holds for $\mathbf{S}$, $\mathbf{B}$ or $\mathbf{C}$, which are elements of a non-linear manifold, a Lie group.
- For a fixed choice of $\mathbf{c}_t$ the weighted residual α_i/σ_i is linear in $\mathbf{R}_t$.

4.4. Finding the Tangency Points

Determination of the epipolar tangency planes is an important part of the calculations. The motion parameters give the position of the epipoles, or the focus of expansion in the infinitesimal case. The B-spline representation is very useful for computing the tangents to the apparent contours that go through a given epipole. There is a way to check each segment and to determine if it contains an epipolar tangency. The tangent can then easily be found with Newton–Raphson iterations. The solution is typically found within a few (3–6) iterations.

4.5. Optimisation Techniques

Computing the maximum likelihood estimate is in essence an optimisation problem. The motion parameters $\mathbf{m}$ are found by minimising (31), i.e.,

$$g(\mathbf{m}) = \sum_i \left(\frac{\alpha_i(\mathbf{m})}{\sigma_i(\mathbf{m})}\right)^2 = \sum Y_i(\mathbf{m})^2,$$

where $Y_i(\mathbf{m})$ is the normalised residual $Y_i(\mathbf{m}) = \alpha_i(\mathbf{m})/\sigma_i(\mathbf{m})$.

The first practical difficulty lies in the non-linear nature of the motion parameter manifold. Therefore a new parametrisation is chosen at each iteration around the current motion parameter $\mathbf{m}_k \in \mathcal{M}$. This parametrisation is a local mapping between $\mathbb{R}^d$ and a neighbourhood of $\mathbf{m}_k$ in $\mathcal{M}$,

$$\mathbb{R}^d \ni x \to \mathbf{m}_k(x) \in \mathcal{M}.$$

The scaled residual Y at $\mathbf{m}_k$ and its derivative $\frac{\partial Y}{\partial x}$ at $\mathbf{m}_k$ are calculated as described on Pages 14. Both Newton-Raphson and Gauss–Newton methods have been used to find the minimum of

$$g(x) = Y(m_k(x))^{\mathrm{T}} Y(m_k(x)),$$

A couple of difficulties have to be solved. The minimisation routine has to be modified so that a descent in the error function g is guaranteed. This involves line-search methods if the error function does not decrease, and checking whether the second order derivative matrix is positive definite, as it should be to guarantee a descent direction. The technique is standard, see (Luenberger, 1984).

Another difficulty is the appearance of new and disappearance of old tangency planes as the motion parameters are changed. In our implementation we have solved this by not allowing new tangencies in the error function until the minimum for the other ones is reached.

5. Statistical Evaluation

The maximum likelihood estimate has several good properties. One is that it is guaranteed to be asymptotically unbiased. Another is that it is asymptotically efficient under reasonable conditions.

The residuals at the minimum can be used to estimate empirically the magnitude of edge localisation error. This can then be compared to the estimates obtained directly in the edge detectors. The residuals can thus be used to automatically verify whether an acceptable estimate of the minimum has been found.

The second order derivative matrix $\frac{\partial^2 g}{\partial x^2}$ or its approximation $\frac{\partial Y}{\partial x}^{\mathrm{T}} \frac{\partial Y}{\partial x}$ together with the variance of the scaled residuals give an estimate of the covariance of the estimated motion parameters,

$$C[\mathbf{x}] = 2\hat{\sigma}^2 \left(\frac{\partial^2 g}{\partial x^2} \right)^{-1} \approx \hat{\sigma}^2 \left(\frac{\partial Y}{\partial x}^{\mathrm{T}} \frac{\partial Y}{\partial x} \right)^{-1}.$$

Note that this covariance matrix is expressed with respect to the particular local parametrisation of the motion parameter manifold.

6. Examples

In the previous sections we have discussed principles for motion determination. In this section we will present results of practical experiments. There are many details that have to be considered when implementing the algorithms. Some of these will be discussed. In the experiments we have illustrated several of the principles discussed previously. Several different camera models are discussed with both continuous- and discrete-time approaches. We will start with the case that gives the simplest calculations, but not so accurate results. The results will then be refined to give more accurate results at the cost of more complicated algorithms. The result from the simpler approaches is used as initial estimates in the more complicated procedures. The first four examples thus form a unit. All procedures have been applied to the same image sequence.

The first four examples are based on scenes in a laboratory. In the fifth example we have used an outdoor scene of Henry Moore sculptures, where the images were taken with a hand held video camera.

A final example illustrates how additional information can be used, e.g., that one of the objects in the scene is a planar. This information admits a drastic simplification as was discussed in (Heyden and Åström, 1997). The experiment shows that this idea can indeed be implemented effectively.

6.1. Infinitesimal Motion, Weak Perspective

In this experiment we use a scene consisting of five stones and a piece of paper on a black cloth. A sequence of pictures have been taken from different view-points with a camera mounted on a tripod. The scene is approximately 40 cm wide and the camera is roughly half a meter from the scene. One image is shown in Fig. 10. The baseline of the camera movement is fairly small, about 5–10 cm, making it reasonable to use an infinitesimal approximation. The apparent contours of the images were extracted and tracked automatically, using the B-spline method described in Section 4.1. The apparent contours in the first image, the normal velocity and its standard deviation were represented with B-splines.

In the weak perspective case the motion parameters are $\mathbf{k}_t$ and $\mathbf{B}_t$, cf. Section 3.3. The focus of expansion $\mathbf{k}_t$ can be represented as

$$\mathbf{k}_t = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{bmatrix},$$

with $0 \leq \theta < \pi$. If $\mathbf{k}_t$ is given the epipolar tangency points can be found through the epipolar tangency constraint

$$|\mathbf{k}_t \quad \mathbf{k} \quad \mathbf{u}_s| = 0.$$

The function to be minimised is

$$g = \sum \frac{\alpha_i^2(\mathbf{k}_t, \mathbf{B}_t)}{\sigma_i^2(\mathbf{k}_t)},$$

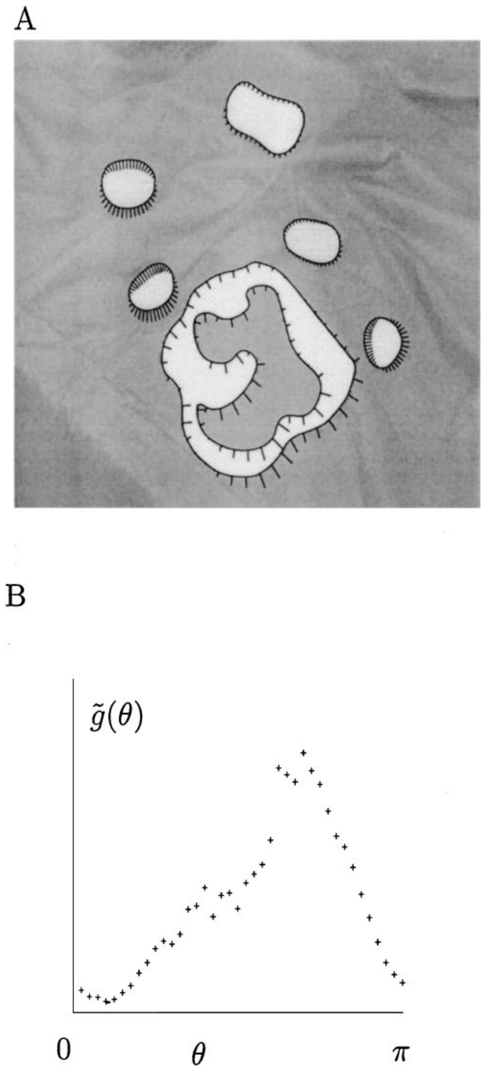

Figure 10. The case of infinitesimal motion with weak perspective camera. (a): One image with apparent contours and the estimated normal velocity is used as input. (b): The minimal error function $\tilde{g}$ is calculated for each of forty choices of focus of expansion $\mathbf{k}_t = (\cos(\theta), \sin(\theta), 0)$ and plotted against θ. The best choice is used as input to a Gauss–Newton minimisation routine. The minimum is found within a few iterations. The logarithm in base 10 of the step lengths is plotted as a function of iteration number.

where $\alpha = \mathbf{n} \cdot (\mathbf{B}_t \mathbf{u} + \mathbf{u}_t)$. The other motion parameter $\mathbf{B}_t$ can be represented as a linear combination of matrices

$$s_1 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad s_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$s_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

If $\mathbf{k}_t$ is fixed, the matrix $\mathbf{B}_t$ belongs to a linear manifold and the loss function g is quadratic in $\mathbf{B}_t$. The optimisation with respect to $\mathbf{B}_t$ can thus be done analytically, and the global minimum is estimated by sweeping over all $\mathbf{k}_t$. In practice we have done this by quantising θ in 40 equidistant parts. The result is illustrated in Fig. 10, which shows the image in a. The value of

$$\tilde{g}(\theta) = \min_{\mathbf{B}_t} g(\mathbf{k}_t(\theta), \mathbf{B}_t), \tag{34}$$

is shown in b. Notice that $\tilde{g}(\theta)$ is π-periodic. A crude value of θ is found as the argmin $\tilde{g}(\theta)$ of the discretised function. In our particular example this gives $\tilde{\theta}$ with a resolution corresponding to the quantisation. An improved estimate is obtained by Gauss–Newton iteration, as illustrated in Fig. 10(c), which shows that a numerical accuracy of 10^{-6} is obtained after 10 iterations.

6.2. *Infinitesimal Motion, Uncalibrated Camera*

The image sequence of Example 6.1 will now be analysed using the uncalibrated camera model. In this case the motion parameters are $\mathbf{c}_t \in \mathbb{S}^2$ and $\mathbf{S}_t$ is a general 3×3 matrix. Given the focus of expansion $\mathbf{c}_t$ the epipolar tangency points can be found through the epipolar tangency constraint

$$|\mathbf{c}_t \quad \mathbf{w} \quad \mathbf{w}_s| = 0.$$

The function to be minimised is

$$g = \sum \frac{\alpha_i^2(\mathbf{c}_t, \mathbf{S}_t)}{\sigma_i^2(\mathbf{c}_t)},$$

where $\alpha = \mathbf{n} \cdot (\mathbf{S}_t \mathbf{w} + \mathbf{w}_t)$. Analogously to the previous example, if $\mathbf{c}_t$ is fixed the matrix $\mathbf{S}_t$ belongs to a linear manifold and the loss function g is quadratic in $\mathbf{S}_t$. The optimisation with respect to $\mathbf{S}_t$ can thus be done analytically, and the global minimum is obtained by sweeping over all $\mathbf{c}_t$. In practice we have done this by tessellation of the sphere. In our particular example, this is done by quantising the latitude and longitude angles in 20 steps, as illustrated in Fig. 11. The value of

$$\tilde{g}(\mathbf{c}_t) = \min_{\mathbf{S}_t} g(\mathbf{c}_t, \mathbf{S}_t) \tag{35}$$

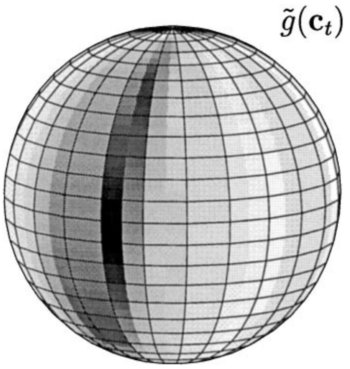

Figure 11. The case of infinitesimal motion with uncalibrated camera. The sphere of directional velocities is tessellated. For each direction $\mathbf{c}_t$, the minimal error function $\tilde{g}$ is found and plotted against $\mathbf{c}_t$. Dark regions correspond to low values of $\tilde{g}$. Notice the long dark valley indicating good choices of $\mathbf{c}_t$. The rough estimate of the minimum is used as an input to a Gauss–Newton minimisation routine. The minimum is found within a few iterations. The shape of this long valley indicates that $\mathbf{c}_t$ will be poorly located along the length of the valley. This is confirmed by the statistical validation.

is shown. A crude value of $\mathbf{c}_t$ is found as argmin $\tilde{g}(\mathbf{c}_t)$ of the discretised function. Improved values is obtained by Gauss–Newton iterations.

Notice that the low values of $\tilde{g}$ form a long valley on the sphere. We expect the direction of motion to be poorly located along that valley. This is confirmed by the statistical evaluation. Also notice that choosing the weak perspective model corresponds to searching this sphere along the equator only.

The minimum obtained from tessellating the sphere or the minimum obtained from the weak perspective case above can both be used as initial estimate in a Gauss–Newton search of the minima. This was done and ten iterations were needed to find the minima.

6.3. Discrete Motion, Weak Perspective Camera

In some of the previous examples the infinitesimal motion parameters were determined. We will now determine the discrete motion parameters. The same images as in the previous example will be used. With the weak perspective camera model the motion parameters are $\Delta\mathbf{k}$ and $\Delta\mathbf{C}$. The loss function $g(\Delta\mathbf{k}, \Delta\mathbf{C})$ was derived on page 15–15. The problem no longer has the nice structure of infinitesimal motions, where we could optimise with respect to one motion parameter analytically. Attempting to discretise all variables gives a very unwieldy optimisation problem. Therefore, we will use the result of Example 6.1 to obtain an initial estimate of the discrete motion parameters:

$$\Delta\mathbf{C} = e^{\mathbf{C}_t \Delta t}, \quad \Delta\mathbf{k} = \mathbf{k}_t \Delta t.$$

The estimate is then refined using Gauss–Newton optimisation.

6.4. Discrete Motion, Uncalibrated Camera

This example is similar to the Example 6.3. The result from the infinitesimal case in Example 6.2 are used as an initial estimates of the discrete motion parameters. A standard extrapolation is used:

$$\Delta\mathbf{S} = e^{\mathbf{S}_t \Delta t}, \quad \Delta\mathbf{c} = \mathbf{c}_t \Delta t.$$

This initial estimate is used as input in a Gauss–Newton search. This will be described in a little more detail here, for one iteration.

The current estimates of motion parameters are used to calculate the epipoles and the epipolar tangencies. This is illustrated in Fig. 12, where (a) and (b) show the images of the epipolar tangency planes. These images are then rectified as described on pages 14–15. The result of the rectification is shown in Fig. 12(c). The epipolar tangency planes should coincide after rectification. In practice there will be deviations, due to edge localisation errors. The difference, represented by the angle α, is calculated together with its standard deviation, as explained on page 14–15. The weighted residual is then computed as

$$Y = \frac{\alpha}{\sigma}.$$

The gradient of Y with respect to the motion parameters is also calculated. The motion parameters are then adjusted using the Gauss–Newton method. Figure 12 shows the result at iteration 4 and 12 in the second and third row respectively.

The figure gives an indication of the epipolar geometry. Comparing Fig. 12(c) and (i), we can also see that the iteration decreases the angle residuals, particularly the residual represented by the lowest lines.

The effectiveness of the optimisation routine is illustrated in Table 3. Notice the rapid reduction of the loss function and the norm of the gradient. A few iterations (6 in this case) are typically required to get close to the minimum. A few more iterations may be required to localise the minimum within machine accuracy.

Table 3. The table illustrate the decrease in loss function g and the gradient magnitude $|\nabla g|$ using the Gauss–Newton optimisation.

Iteration	1	2	3	4	5	6	7	8	9	10
g	110	66	34	16	7.7	6.5	6.5	6.5	6.5	6.5
$\log_{10}(\|\nabla g\|)$	−1.0	−1.1	−1.3	−1.5	−2.0	−3.6	−5.5	−7.6	−9.7	−9.7

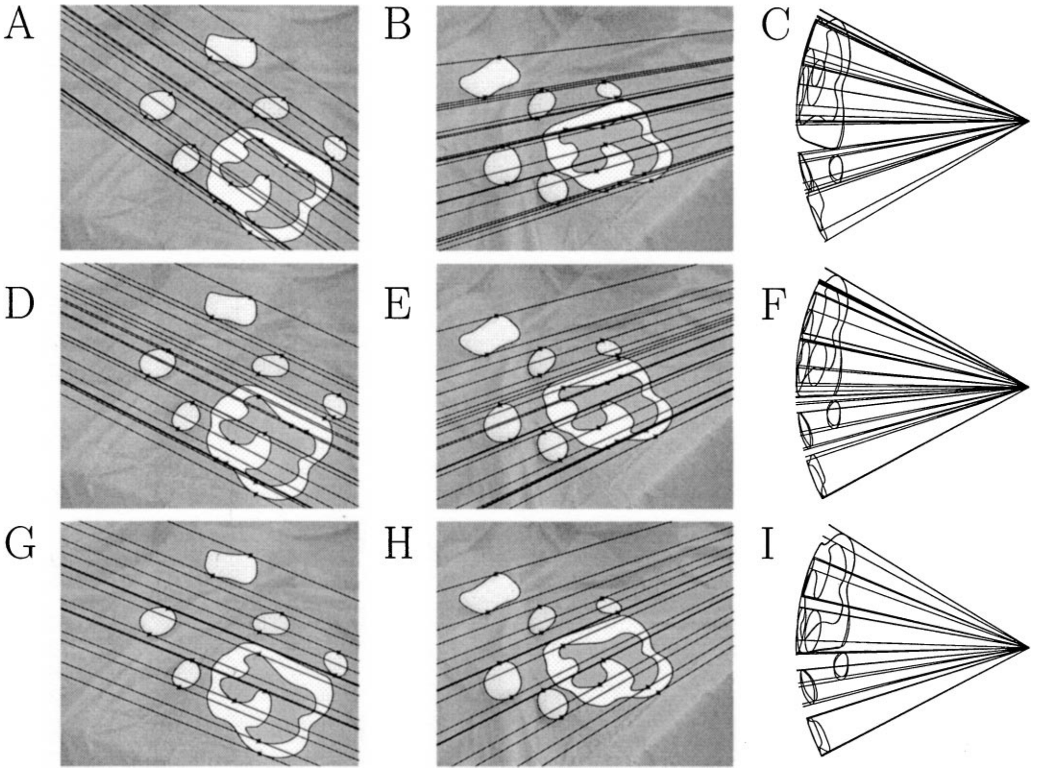

Figure 12. Finding the uncalibrated motion parameters using the generalised epipolar constraints. Optimisation of the likelihood function. Iteration number 1 (first row), 4 (second row) and 10 (third row). The first and second columns illustrate the epipolar tangencies in the first and second image. The third column illustrates the rectified epipolar tangency planes, projected on the view-sphere and viewed along the direction of motion.

6.5. Discrete Motion, Calibrated Camera

Experiments have also been made to determine motion parameters based on real outdoor scenes. We have taken video sequences of a Henry Moore statue in Yorkshire Sculpture Park, UK, see Fig. 13. The images were taken using a hand-held video camera. The figure illustrates two frames of a longer video sequence of the scene. In this example the apparent contours were detected and tracked manually. Two plausible local minima were found, one with eight tangencies, Fig. 13, and one with six tangencies, Fig. 14. Both epipolar line structures agree with the image data. However, they yield different solutions. The covariance matrices of the motion parameters are quite large in both cases. This is probably caused by the small number of tangencies, the small baseline and perhaps a poorly calibrated camera. These are the same problems that plague all algorithms for determination of motion parameters from structure.

6.6. Discrete Motion, Known Rotation

The experiment illustrate the use of a priori information. In (Heyden and Åström, 1997) it was shown that the presence of a planar feature in the scene makes it possible to simplify the algorithms drastically. By

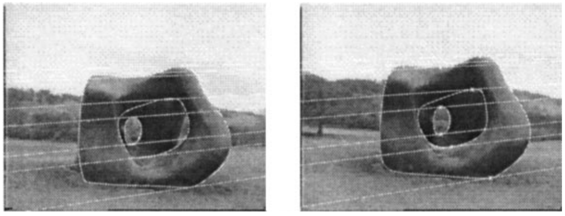

Figure 13. Henry Moore sculpture. 8 epipolar tangencies lead to convergence to different local minima. Due to the small field of view and because the direction of translation is outside the image frame the solution is very sensitive to image contour localisation errors.

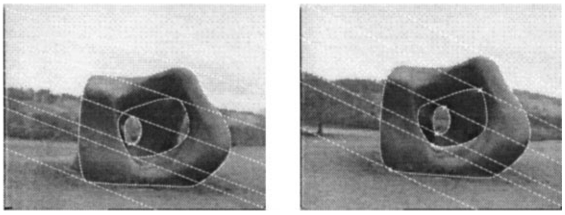

Figure 14. Henry Moore sculpture. 6 epipolar tangencies are used to estimate the motion between frames. Although they yield a consistent estimate of the motion between frames (judged by quality of epipolar line structure) the solution is ill-conditioned.

detecting and aligning a planar feature in a sequence of images, the analysis of the sequence can be reduced to the case of a purely translating camera. The planar curve is regarded as a curve on the plane at infinity. Thus it has no apparent image motion. This simplification has been used in (Heyden and Åström, 1997; Sinclair et al., 1995). It is known as *projective reduction*, which is a generalisation of the '*plane plus parallax*' method.

This is illustrated in Fig. 15. Only the direction of motion $\Delta\mathbf{c}$ needs to be estimated. The sphere of possible directions can then be tessellated and the error function g can be calculated for each direction. The minimum obtained after tessellation is improved by local Newton-Raphson search (6 iterations were needed). The scene used in the experiment consists of stones placed on building blocks. The camera was mounted on a robot and moved around the scene. Four images are shown in Fig. 15. The contours from all images are aligned so that the planar curve coincides.

7. Extension to Multiple Images

The techniques in this paper describe in detail how motion parameters between pairs of images can be estimated. General features such as points, planar curves, space curves and apparent contours of general surfaces can be used. The same kind of **generalised epipolar constraints** apply to all these features. One interesting application is to use estimates of motion parameters for pairs of images in a sequence, to obtain the full motion of the camera. In the calibrated case the motion can be represented by camera positions $\mathbf{c}_i = \mathbf{c}(t_i)$ and camera orientation $\mathbf{R}_i = \mathbf{R}(t_i)$. From the motion constraint

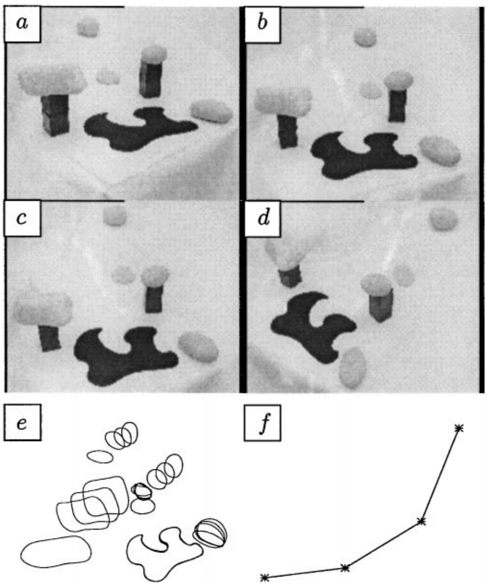

Figure 15. Projective reduction. (a)–(d): Four images out of a longer sequence. (e): By detecting and aligning the image of a planar feature the images can be thought of as coming from a purely translating camera. The apparent contours after alignment are shown. (f): This makes it relatively easy to extract motion parameters between each pair of images. These parameters can then be used to calculate the full motion of the camera.

$$|\mathbf{c}_j - \mathbf{c}_i \quad \mathbf{R}_i\mathbf{q}_i \quad \mathbf{R}_j\mathbf{q}_j|,$$

elementary determinant operations give

$$\begin{aligned} &|\mathbf{R}_i^{-1}(\mathbf{c}_j - \mathbf{c}_i) \quad \mathbf{q}_i \quad \mathbf{R}_i^{-1}\mathbf{R}_j\mathbf{q}_j| = \\ &\qquad |\Delta\mathbf{c}_{ij} \quad \mathbf{q}_i \quad \Delta\mathbf{R}_{ij}\mathbf{q}_j| = 0. \end{aligned} \tag{36}$$

Our algorithm was constructed so that (36) holds. It thus follows that

$$\begin{aligned} \mu_{ij}\Delta\mathbf{c}_{ij} &= \mathbf{R}_i^{-1}(\mathbf{c}_j - \mathbf{c}_i), \\ \gamma_{ij}\Delta\mathbf{R}_{ij} &= \mathbf{R}_i^{-1}\mathbf{R}_j, \end{aligned}$$

where $(\Delta\mathbf{c}_{ij}, \Delta\mathbf{R}_{ij})$ describes the incremental motion parameters from image i to image j. Since $\Delta\mathbf{R}_{ij}$ and

$\mathbf{R}_i^{-1}\mathbf{R}_j$ are rotation matrices, γ_{ij} must be one. The overall coordinate system must be chosen, e.g., by choosing $\mathbf{c}_0 = 0$, $\mathbf{R}_0 = I$ and $|\mathbf{c}_n| = 1$.

Similar equations apply to the other camera models. In the uncalibrated case the 3 parameter ambiguity in determining $\mathbf{S}_{ij}$ and the choice of plane at infinity must also be taken into account.

The idea is illustrated in Fig. 15. Four images of a short sequence is shown (a)–(d) and the camera motion is represented as the corresponding four camera positions (f).

8. Conclusions and Future Work

The apparent contour and its deformation under viewer motion is known to be a rich source of surface geometric information. This can be used in visual navigation and object manipulation. Here we have shown how so called frontier points of apparent contours can be used to recover the viewer motion from the deformation of apparent contours. The epipolar constraint for points is generalised to curves and apparent contours. The results hold for continuous and discrete motion cases, for uncalibrated and calibrated cameras and for perspective and parallel camera models. An iterative method to obtain the maximum likelihood estimate of the motion parameters is presented and the problem of finding initial estimates is discussed. Statistical evaluation of the results are presented. They can be used to evaluate the validity of the solution and also to obtain estimates of the covariance of the estimated motion parameters. The theory is applied to experiments with real image sequences. It is also indicated how motion between image pairs can be used to obtain full camera motion.

In this paper we have shown that we can estimate camera motion by measuring the deformation of the apparent contour. In the future we intend to evaluate the performance of this approach, for example by studying if the motion estimate is good enough to test whether an apparent contour is the silhouette of a curved surface or the image of a fixed curve. We are also going to study the problem of estimating surface structure using the motion estimate.

As with all methods that depend on optimisation by iterative techniques, there is the question of finding good initial estimates, so that the global optimum is found. In the paper some ideas for doing this is described, but more work is needed in order to establish the feasibility of the methods.

In the examples of this paper we have only used apparent contours in the motion estimates. In practice one would use a combination of image features to estimate motion. Note that the generalised epipolar constraint applies to both points, curves and curved surfaces.

It is unclear whether the extraction of motion leads to a unique solution. For circular motion and parallel projection it does, see (Giblin et al., 1994), but we intend to apply the new insights gained from the present work back to this case, generalising as far as perspective projection. Possibly other simple motion can be included too.

Another theoretical question of interest is to determine necessary and sufficient conditions for profiles $\mathbf{p}(s, t)$ and camera motion $\mathbf{c}(t)$ to arise from a surface in space with the contour generators forming an envelope on the surface. In the case of point reconstructions bilinear and trilinear constraints are enough but the situation is less clear for surfaces. Some progress is described in (Fletcher, 1996), where 'higher frontier conditions' are obtained which in principle provide constraints on camera motion additional to those described in the present paper.

Acknowledgments

Kalle Åström acknowledge support from ESPRIT Long term research project CUMULI. Kalle Åström and Peter Giblin acknowledge support from ESPRIT-BRA VIVA. Roberto Cipolla acknowledge support EPSRC grant GR/K08635. Peter Giblin acknowledge support EPSRC grants GR/F59855 and GR/F28162. The authors also thank Gordon Fletcher at Liverpool University for help in making some of the illustrations.

References

Åström, K. 1996. Invariancy methods for points, curves and surfaces in computational vision. Ph.D. thesis, Department of Mathematics, Lund University, Sweden.

Åström, K., Cipolla, R., and Giblin, P.J. 1996. Generalised epipolar constraints. In *Proc. 4th European Conf. on Computer Vision*, B. Buxton and R. Cipolla (Eds.), vol. 1065 of Lecture notes in Computer Science, Springer-Verlag: Cambridge, UK, pp. 97–108.

Åström, K. and Heyden, A. 1999. Stochastic analysis of scale space smoothing. *Advances of Applied Probability*.

Blake, A. and Yuille, A.L. (Eds.). 1992. *Active Vision*, MIT Press: Cambridge, Massachusetts, London, England.

Bruce, J.W. and Giblin, P.J. 1992. *Curves and Singularities*, 2nd edition, Cambridge University Press.

Carlsson, S. 1994. Sufficient image structure for 3D motion and shape estimation. In *Proc. 3rd European Conf. on Computer Vision*, J.-O. Eklundh (Ed.), Springer-Verlag: Stockholm, Sweden, vol. I, pp. 83–91.

Cipolla, R. 1991. Active visual inference of surface shape. Ph.D. thesis, University of Oxford. Also published in Springer-Verlag as LNCS 1016 (1995).

Cipolla, R., Åström, K., and Giblin, P.J. 1995. Motion from the frontier of curved surfaces. In *Proc. 5th Int. Conf. on Computer Vision*, MIT: Boston, MA, pp. 269–275.

Cipolla, R. and Blake, A. 1990. The dynamic analysis of apparent contours. In *Proc. 3rd Int. Conf. on Computer Vision*, Osaka, Japan, pp. 616–623.

Cipolla, R. and Blake, A. 1992. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112.

Cipolla, R., Fletcher, G., and Giblin, P. 1996. Following cusps. *Int. Journal of Computer Vision*.

Coxeter, H.S.M. 1993. *The Real Projective Plane*, 3rd edition, Springer-Verlag: New York, NY, USA.

Curwen, R. and Blake, A. 1992. Dynamic contours: Real-time active splines. In *Active Vision*, A. Blake and A. Yuille (Eds.), MIT Press, pp. 39–58.

Faugeras, O. 1993. *Three-Dimensional Computer Vision*. MIT Press: Cambridge, Mass.

Faugeras, O.D. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. 2nd European Conf. on Computer Vision*, G. Sandini (Ed.), Springer-Verlag: Santa Margherita Ligure, Italy, pp. 563–578.

Faugeras, O.D., Luong, Q.-T., and Maybank, S. 1992. Camera self-calibration: Theory and experiment. In *Proc. 2nd European Conf. on Computer Vision*, G. Sandini (Ed.), Springer-Verlag: Santa Margherita Ligure, Italy, pp. 321–334.

Fletcher, G. 1996. Curves and surfaces. Ph.D. thesis, Department of Pure Mathematics, Liverpool University, UK.

Foley, J.D., van Dam, A., Feiner, S.K., and Hughes, J.F. 1990. *Computer Graphics, Principles and Practice*. Addison-Wesley Publishing Company.

Giblin, P.J., Pollick, F.E., and Rycroft, J.E. 1994. Recovery of an unknown axis or rotation from the profiles of a rotating surface. *J. Opt. Soc. America*, 11(A):1976–1984.

Giblin, P.J. and Weiss, R. 1987. Reconstruction of surfaces from profiles. In *Proc. 1st Int. Conf. on Computer Vision*, London, pp. 136–144.

Giblin, P.J. and Weiss, R.S. 1995. Epipolar curves on surfaces. *Image and Vision Computing*, 13:33–44.

Heyden, A. and Åström, K. 1997. Simplifications of multilinear forms for sequences of images. *Image and Vision Computing*, 15(10):749–757.

Joshi, T., Ahuja, N., and Ponce, J. 1995. Structure and motion estimates from dynamic silhouettes under perspective projection. In *Proc. 5th Int. Conf. on Computer Vision*, MIT: Boston, MA, pp. 290–295.

Kass, M., Witkin, A., and Terzopoulos, D. 1987. Snakes: Active contour models. *Int. Journal of Computer Vision*, 1(4):321–331.

Koenderink, J.J. 1990. *Solid Shape*. MIT Press.

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

Luenberger, D.G. 1984. *Linear and Nonlinear Programming*. Addison-Wesley.

Luong, Q.T., Deriche, R., Faugeras, O.D., and Papadopoulo, T. 1993. On determining the fundamental matrix: Analysis of different methods and experimental results. *Technical Report RR-1894*, INRIA.

Luong, Q. and Vieville, T. 1994. Canonic representations for the geometries of multiple projective views. In *Proc. 3rd European Conf. on Computer Vision*, Stockholm, Sweden, pp. 589–599.

Porrill, J. and Pollard, S.B. 1991. Curve matching and stereo calibration. *Image and Vision Computing*, 9(1):45–50.

Rieger, J.H. 1986. Three dimensional motion from fixed points of a deforming profile curve. *Optics Letters*, 11:123–125.

Sinclair, D., Christiansenn, H., and Rothwell, C. 1995. Using the relation between a plane projectivity and the fundamental matrix. In *Proc. 9th Scandinavian Conference on Image Analysis*, pp. 181–188.

Stefanovic, P. 1973. Relative orientation—a new appraoch. *ITC Journal*, 3:417–448.

Thompson, E.H. 1959. A rational algebraic formulation of the problem of relative orientation. *Photogrammetric Record*, 14(3):152–159.

Vaillant, R. and Faugeras, O.D. 1992. Using extremal boundaries for 3D object modelling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):157–173.

Vijayakumar, B., Kriegman, D., and Ponce, J. 1995. Invariant-based recognition of complex 3D curved objects from image contours. In *Proc. 5th Int. Conf. on Computer Vision*, MIT: Boston, MA, pp. 508–514.

Vijayakumar, B., Kriegman, D., and Ponce, J. 1996. Structure and motion of curved 3D objects from monocular silhouettes. In *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 327–334.