An analytical approach to determining the ego-motion of a camera having free intrinsic parameters

Michael J. Brooks
mjb@cs.adelaide.edu.au
Department of Computer Science
University of Adelaide
Adelaide, SA 5005, Australia

Luis Baumela
lbaumela@mayor.dia.fi.upm.es
Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Boadilla del Monte Madrid, Spain

Wojciech Chojnacki
wojtek@cs.adelaide.edu.au
Department of Computer Science
University of Adelaide
Adelaide, SA 5005, Australia

Synopsis

In this paper, we analyse the motion of a camera having free intrinsic parameters. We define a free parameter to be one that is unknown and may vary continuously. A time-dependent epipolar equation is presented, followed by a formal definition of the time-derivative of the fundamental matrix for the case of a mobile camera. Next, differential forms of the epipolar equation are obtained. This may be seen as a recasting of the recent work of Viéville and Faugeras [12] into an analytical framework. Critical to the approach is the determination, to within a common scalar factor, of two special matrices from optical flow data. The case of a camera with free focal length undergoing arbitrary motion is then considered in detail. Closed-form expressions are given, in terms of the entries of the two matrices, for the ego-motion parameters, as well as the focal length and its derivative.

Key words and phrases. Self-calibration, ego-motion, epipolar equation, fundamental matrix, intrinsic parameters.

1 Introduction

The *epipolar equation* in stereo vision takes the form

$$\mathbf{m}^T \mathbf{F} \mathbf{m}' = 0, \tag{1}$$

where $\mathbf{m}$ and $\mathbf{m}'$ represent corresponding points in the images obtained by left and right cameras, respectively, expressed in terms of homogeneous coordinates, and $\mathbf{F}$ is the *fundamental matrix* influenced by both extrinsic and intrinsic imaging factors, henceforth termed the *key parameters* [8]. (Note that a slightly non-standard notation is used here, as described in Appendix A.) Given sufficiently many corresponding points in a typical non-degenerate configuration, it is sometimes possible, via a process of *self-calibration,* to determine various of the key parameters [2, 9]. It is well known that using corresponding points extracted from a single pair of images, at most 7 key parameters may be determined. These might, for example, comprise 5 relative orientation parameters and 2 focal lengths (e.g. see [4, 10]).

Our aim in this work is to introduce into (1) a dependency on *time* and to derive corresponding differential equations. In this way, we hope to carry out self-calibration (determining camera motion and intrinsic parameter values) using only optical flow information. Part of our work may be seen as a recasting of the research of Viéville and Faugeras [12] into an analytical framework. For related work dealing with *ego-motion* of a *calibrated* camera, see for example [3, 5–7]. (We recall that the ego-motion refers to the motion of a camera described in terms of its own local coordinate system.)

Adding a dependency upon time to (1), we have

$$\mathbf{m}^T(t)\mathbf{F}(t)\mathbf{m}'(t) = 0, \tag{2}$$

which we may call *the time dependent epipolar equation for stereo cameras*. The questions now arise: What is being modelled by this equation, and to what use may the equation be put?

At a given time, (2) is simply an instance of (1), enabling recovery of key parameters given sufficiently many corresponding points. At a later time, (2) can once again be used to recompute the key parameters. If these are unchanged, then so too will be the fundamental matrix (although it can be estimated only up to a scalar factor). Therefore, for a pair of cameras in a fixed relationship, with unchanged relative orientation and intrinsic parameters, consideration of time will have no usefulness. This applies even if the stereo cameras are in motion relative to some global frame (since each camera remains stationary relative to the other).

Assume now that the cameras are not in a fixed relationship, but that they are free to move independently. Consideration of (2) then offers the possibility of computing the instantaneous changes in both relative orientation and intrinsic parameters of the cameras, expressed in terms of the location and movement of various image points. Differentiating (2) with respect to time, we have

$$\dot{\mathbf{m}}^T(t)\mathbf{F}(t)\mathbf{m}'(t) + \mathbf{m}^T(t)\dot{\mathbf{F}}(t)\mathbf{m}'(t) + \mathbf{m}^T(t)\mathbf{F}(t)\dot{\mathbf{m}}'(t) = 0.$$

If, say, the left camera remains stationary, then $\mathbf{m}(t) = \mathbf{m}$ and $\dot{\mathbf{m}}(t) = 0$, and so

$$\mathbf{m}^T\dot{\mathbf{F}}(t)\mathbf{m}'(t) + \mathbf{m}\mathbf{F}(t)\mathbf{m}'(t) = 0.$$

There is now the possibility of computing over time the relative orientation of the right camera with respect to a fixed left camera. Thus the motion of the right camera may be computed relative to a fixed left camera's frame of reference, instead of motion relative to the (possibly moving) left camera. However, in order to achieve this, corresponding points would have to be extracted from successive image pairs generated by the static and moving cameras. We shall not pursue this awkward approach here.

Of greater interest is to envisage the time-varying nature of the epipolar equation arising from views of a *single* mobile camera at successive time intervals. In order to use an adaptation of (2) in this situation, we have to be very clear about the nature of both the fundamental matrix and corresponding points. The discrete approach to motion presented in [13] and [1] involved stereo pairs generated by a pair of cameras at successive, discrete times. Of interest here is to contemplate the limiting case of the time difference between images tending to zero (as in [12]), thereby permitting computation of both the ego-motion and the intrinsic parameters of the camera.

Note, immediately, that the following equation holds little value:

$$\mathbf{m}^T(t)\mathbf{F}(t)\mathbf{m}(t) = 0.$$

This deals merely with identical left and right images and points. In this situation, we clearly have $\mathbf{F}(t) = \mathbf{0}$. Here there is a failure to recognise that a fundamental matrix dealing with single camera should relate a pair of images captured at different times.

In order to clarify matters, it proves useful to consider a relatively general formulation of the time-dependent epipolar equation. Consider the following:

$$\mathbf{m}^T\left(I(t_1)\right)\mathbf{F}\left(I(t_1), I'(t_2)\right)\mathbf{m}'\left(I'(t_2)\right) = 0. \tag{3}$$

This we call *the general time-dependent epipolar equation*. Here, I and I' are image streams obtained from left and right cameras, respectively, and t_1 and t_2 are specific times. The points $\mathbf{m}$ and $\mathbf{m}'$ are images of a fixed 3D point in space. This equation makes explicit the dependencies of the fundamental matrix $\mathbf{F}$. Of critical importance here is to note that the fundamental matrix associated with images obtained from a single camera (in contrast with that associated with a pair of cameras) is dependent upon *two* times. It is this realisation that will shortly enable a precise time-derivative of $\mathbf{F}$ to be defined, and a novel analytical derivation of a differential epipolar equation.

Observe that when $t_1 = t_2 = t$, (3) reduces to (2). Suppose now that we have a single mobile camera, so that $I = I'$, and that successive images are captured at times t_1 and t_2. Equation (3) then becomes

$$\mathbf{m}^T\left(I(t_1)\right)\mathbf{F}\left(I(t_1), I(t_2)\right)\mathbf{m}\left(I(t_2)\right) = 0.$$

Dropping the now superfluous image notation I, we obtain

$$\boxed{\mathbf{m}^T(t_1)\mathbf{F}(t_1, t_2)\mathbf{m}(t_2) = 0.} \tag{4}$$

This we may term *the time-dependent epipolar equation for a single camera*, and it forms the basis for our subsequent considerations.

2 Differential forms of the time-dependent epipolar equation

We now confine our attention to (4), seeking differential forms that enable instantaneous changes in the key parameters to be related to instantaneous changes in the positions of corresponding points.

Assume that a camera undergoes a smooth motion over a period of time, thereby generating an image stream. At times t_1 and t_2, with $t_1 \neq t_2$, (4) will constrain the relationship between the uncalibrated coordinates of the corresponding points and the image formation parameters bound up in $\mathbf{F}$. Clearly, as t_1 and t_2 vary, $\mathbf{F}(t_1, t_2)$ will also vary, with $\mathbf{F}(t_1, t_2)$ tending to $\mathbf{0}$ as $t_2 \to t_1$. Moreover, the derivative of $\mathbf{F}(t_1, t_2)$ will at all times be defined, including at time $t_1 = t_2$. The time-derivatives of $\mathbf{F}$ at $t_1 = t_2 = t$ will be of particular interest, for these will be central to the consideration of ego-motion of a single, moving camera.

We adopt the following convention: Given a function that maps (t_1, t_2) into $\mathbf{X}(t_1, t_2)$, we write for each t

$$\mathbf{X}(t) = \mathbf{X}(t,t), \quad \mathring{\mathbf{X}}(t) = \frac{\partial \mathbf{X}}{\partial t_2}(t_1, t_2)\bigg|_{t_1=t_2=t}, \quad \mathring{\mathring{\mathbf{X}}}(t) = \frac{\partial^2 \mathbf{X}}{\partial t_2^2}(t_1, t_2)\bigg|_{t_1=t_2=t}. \quad (5)$$

With this notation, we clearly have (e.g. see (8) and (12) from the next section) that

$$\mathbf{F}(t) = \mathbf{0}. \quad (6)$$

Differentiating (4) with respect to t_2, we obtain

$$\mathbf{m}^T(t_1)\frac{\partial \mathbf{F}}{\partial t_2}(t_1, t_2)\mathbf{m}(t_2) + \mathbf{m}^T(t_1)\mathbf{F}(t_1, t_2)\dot{\mathbf{m}}(t_2) = 0,$$

whence, on letting $t_1 = t_2 = t$,

$$\mathbf{m}^T(t)\mathring{\mathbf{F}}(t)\mathbf{m}(t) + \mathbf{m}^T(t)\mathbf{F}(t)\dot{\mathbf{m}}(t) = 0.$$

Omitting the notational dependency on time, and using (6), we may rewrite the latter equation as

$$\boxed{\mathbf{m}^T\mathring{\mathbf{F}}\mathbf{m} = 0.}$$

This we term *the first differential form of the epipolar equation*, as it has arisen by once differentiating (4).

We may now follow a similar path to obtain the second form. Differentiating (4) twice with respect to t_2, we obtain

$$\mathbf{m}^T(t_1)\frac{\partial^2 \mathbf{F}}{\partial t_2^2}(t_1, t_2)\mathbf{m}(t_2) + 2\mathbf{m}^T(t_1)\frac{\partial \mathbf{F}}{\partial t_2}(t_1, t_2)\dot{\mathbf{m}}(t_2) + \mathbf{m}^T(t_1)\mathbf{F}(t_1, t_2)\ddot{\mathbf{m}}(t_2) = 0,$$

whence, on letting $t_1 = t_2 = t$,

$$\mathbf{m}^T(t)\mathring{\mathring{\mathbf{F}}}(t)\mathbf{m}(t) + 2\mathbf{m}^T(t)\mathring{\mathbf{F}}(t)\dot{\mathbf{m}}(t) + \mathbf{m}^T(t)\mathbf{F}(t)\ddot{\mathbf{m}}(t) = 0,$$

and further, in view of (6),

$$\boxed{\mathbf{m}^T \overset{\circ\circ}{\mathbf{F}} \mathbf{m} + 2\mathbf{m}^T \overset{\circ}{\mathbf{F}} \dot{\mathbf{m}} = 0.} \qquad (7)$$

This we term *the second differential form of the epipolar equation*. Note that this equation contains both location and velocity of an image point, but not its acceleration, $\ddot{\mathbf{m}}$ having fallen away in the derivation.

3 Elaborating the second differential form

In this section, we describe how $\overset{\circ}{\mathbf{F}}(t)$ and $\overset{\circ\circ}{\mathbf{F}}(t)$ may be represented in terms of the component matrices of $\mathbf{F}$. This will enable the differential epipolar equation to be transformed into a form that facilitates self-calibration.

The fundamental matrix $\mathbf{F}(t_1, t_2)$ for a single camera may be expressed as

$$\mathbf{F}(t_1, t_2) = \mathbf{A}^T(t_1)\mathbf{E}(t_1, t_2)\mathbf{A}(t_2), \qquad (8)$$

where the matrix $\mathbf{A}(t)$ describes the intrinsic parameters of the camera at instant t, and $\mathbf{E}(t_1, t_2)$ is the *essential matrix* defined as

$$\mathbf{E}(t_1, t_2) = \mathbf{T}(t_1, t_2)\mathbf{R}(t_1, t_2), \qquad (9)$$

where the matrices $\mathbf{T}(t_1, t_2)$ and $\mathbf{R}(t_1, t_2)$ describe, respectively, the translational and rotational components of the camera's movement from the position at time t_1 to the position at time t_2 (e.g. see [8]). The intrinsic parameters within $\mathbf{A}(t)$ may vary continuously with time. An explicit form of the matrix function $\mathbf{A}$ will be given in a subsequent section. The translation matrix $\mathbf{T}(t_1, t_2)$ takes the form

$$\mathbf{T}(t_1, t_2) = \begin{pmatrix} 0 & -z(t_1, t_2) & y(t_1, t_2) \\ z(t_1, t_2) & 0 & -x(t_1, t_2) \\ -y(t_1, t_2) & x(t_1, t_2) & 0 \end{pmatrix},$$

where $(x(t_1, t_2), y(t_1, t_2), z(t_1, t_2))^T$ is the baseline vector that connects the optical centres of images captured at times t_1 and t_2. The rotation matrix $\mathbf{R}(t_1, t_2)$ is given by

$$\mathbf{R}(t_1, t_2) = \mathbf{R}_1(\alpha)\mathbf{R}_2(\beta)\mathbf{R}_3(\gamma),$$

where the component matrices

$$\mathbf{R}_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix},$$

$$\mathbf{R}_2(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix},$$

$$\mathbf{R}_3(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

correspond to counter clockwise rotations about the camera-centered coordinate axes x, y, and z by the angles α, β and γ, respectively. For convenience, the dependency of α, β and γ upon (t_1, t_2) is left implicit.

In notation from the previous section, we have $x(t) = y(t) = z(t) = \alpha(t) = \beta(t) = \gamma(t) = 0$, so that

$$\mathbf{T}(t) = \mathbf{0}, \tag{10a}$$

$$\mathbf{R}(t) = \mathbf{I}. \tag{10b}$$

A straightforward computation then reveals that

$$\overset{\circ}{\mathbf{T}}(t) = \begin{pmatrix} 0 & -\overset{\circ}{z}(t) & \overset{\circ}{y}(t) \\ \overset{\circ}{z}(t) & 0 & -\overset{\circ}{x}(t) \\ -\overset{\circ}{y}(t) & \overset{\circ}{x}(t) & 0 \end{pmatrix}, \tag{11a}$$

$$\overset{\circ}{\mathbf{R}}(t) = \begin{pmatrix} 0 & \overset{\circ}{\gamma}(t) & -\overset{\circ}{\beta}(t) \\ -\overset{\circ}{\gamma}(t) & 0 & \overset{\circ}{\alpha}(t) \\ \overset{\circ}{\beta}(t) & -\overset{\circ}{\alpha}(t) & 0 \end{pmatrix}. \tag{11b}$$

The vectors $(\overset{\circ}{x}(t), \overset{\circ}{y}(t), \overset{\circ}{z}(t))^T$ and $(\overset{\circ}{\alpha}(t), \overset{\circ}{\beta}(t), \overset{\circ}{\gamma}(t))^T$ associated with $\overset{\circ}{\mathbf{T}}(t)$ and $\overset{\circ}{\mathbf{R}}(t)$ capture the instantaneous *translational* and *angular velocities* of camera ego-motion, respectively. Observe that both $\overset{\circ}{\mathbf{T}}$ and $\overset{\circ}{\mathbf{R}}$ are antisymmetric. Additionally, matrix $\overset{\circ\circ}{\mathbf{T}}$ is readily shown to be antisymmetric.

In view of (9) and (10a), we have

$$\mathbf{E}(t) = \mathbf{0}. \tag{12}$$

Differentiating (9) with respect to t_2, we obtain

$$\frac{\partial \mathbf{E}}{\partial t_2}(t_1, t_2) = \frac{\partial \mathbf{T}}{\partial t_2}(t_1, t_2)\mathbf{R}(t_1, t_2) + \mathbf{T}(t_1, t_2)\frac{\partial \mathbf{R}}{\partial t_2}(t_1, t_2),$$

whence, on letting $t_1 = t_2 = t$ and using (10a),

$$\overset{\circ}{\mathbf{E}}(t) = \overset{\circ}{\mathbf{T}}(t)\mathbf{R}(t) + \mathbf{T}(t)\overset{\circ}{\mathbf{R}}(t) = \overset{\circ}{\mathbf{T}}(t). \tag{13}$$

Differentiating (8) with respect to t_2, we find that

$$\frac{\partial \mathbf{F}}{\partial t_2}(t_1, t_2) = \mathbf{A}^T(t_1)\frac{\partial \mathbf{E}}{\partial t_2}(t_1, t_2)\mathbf{A}(t_2) + \mathbf{A}^T(t_1)\mathbf{E}(t_1, t_2)\dot{\mathbf{A}}(t_2),$$

whence, on letting $t_1 = t_2 = t$ and taking into account (12) and (13),

$$\overset{\circ}{\mathbf{F}}(t) = \mathbf{A}^T(t)\overset{\circ}{\mathbf{E}}(t)\mathbf{A}(t) = \mathbf{A}^T(t)\overset{\circ}{\mathbf{T}}(t)\mathbf{A}(t). \tag{14}$$

Differentiating (8) twice with respect to t_2, we conclude that

$$\frac{\partial^2 \mathbf{F}}{\partial t_2^2}(t_1, t_2) = \mathbf{A}^T(t_1)\left[\frac{\partial^2 \mathbf{E}}{\partial t_2^2}(t_1, t_2)\mathbf{A}(t_2) + 2\frac{\partial \mathbf{E}}{\partial t_2}(t_1, t_2)\dot{\mathbf{A}}(t_2) + \mathbf{E}(t_1, t_2)\ddot{\mathbf{A}}(t_2)\right],$$

whence, on letting $t_1 = t_2 = t$ and taking into account (12),

$$\overset{\circ\circ}{\mathbf{F}}(t) = \mathbf{A}^T(t)\overset{\circ\circ}{\mathbf{E}}(t)\mathbf{A}(t) + 2\mathbf{A}^T(t)\overset{\circ}{\mathbf{E}}(t)\dot{\mathbf{A}}(t). \tag{15}$$

Dropping henceforth the dependency on t, we may now apply a similar procedure to (9), obtaining

$$\overset{\circ\circ}{\mathbf{E}} = \overset{\circ\circ}{\mathbf{T}}\mathbf{R} + 2\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}} + \mathbf{T}\overset{\circ\circ}{\mathbf{R}} = \overset{\circ\circ}{\mathbf{T}} + 2\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}}. \tag{16}$$

In view of (14),

$$\mathbf{m}^T\overset{\circ}{\mathbf{F}}\dot{\mathbf{m}} = \mathbf{m}\mathbf{A}^T\overset{\circ}{\mathbf{T}}\mathbf{A}\dot{\mathbf{m}}. \tag{17}$$

By (13), (15), and (16),

$$\mathbf{m}^T\overset{\circ\circ}{\mathbf{F}}\mathbf{m} = \mathbf{m}^T\mathbf{A}^T\overset{\circ\circ}{\mathbf{T}}\mathbf{A}\mathbf{m} + 2\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}}\mathbf{A}\mathbf{m} + 2\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\dot{\mathbf{A}}\mathbf{m}. \tag{18}$$

Since $\overset{\circ\circ}{\mathbf{T}}$ is antisymmetric, it follows that

$$\mathbf{m}^T\mathbf{A}^T\overset{\circ\circ}{\mathbf{T}}\mathbf{A}\mathbf{m} = 0.$$

Therefore (18) can be rewritten as

$$\mathbf{m}^T\overset{\circ\circ}{\mathbf{F}}\mathbf{m} = 2\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}}\mathbf{A}\mathbf{m} + 2\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\dot{\mathbf{A}}\mathbf{m}.$$

This equation along with (7) and (17) leads to the second differential epipolar equation in the following form:

$$\boxed{\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}}\mathbf{A}\mathbf{m} + \mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\dot{\mathbf{A}}\mathbf{m} + \mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\mathbf{A}\dot{\mathbf{m}} = 0.} \tag{19}$$

Observe that even though this equation incorporates the first and second derivatives of the fundamental matrix, no second derivatives of its component matrices survive the elaboration.

4 An alternative second differential form

We now derive an alternative form of (19) that is more amenable to numerical solution.

Introducing

$$\mathbf{B} = \dot{\mathbf{A}}\mathbf{A}^{-1}, \tag{20}$$

we first rewrite (19) as

$$\mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}(\overset{\circ}{\mathbf{R}} + \mathbf{B})\mathbf{A}\mathbf{m} + \mathbf{m}^T\mathbf{A}^T\overset{\circ}{\mathbf{T}}\mathbf{A}\dot{\mathbf{m}} = 0. \tag{21}$$

Given a matrix $\mathbf{X}$, denote by $\mathbf{X}_{\text{sym}}$ and $\mathbf{X}_{\text{asym}}$ the symmetric and antisymmetric parts of $\mathbf{X}$ defined, respectively, by

$$\mathbf{X}_{\text{sym}} = \frac{1}{2}(\mathbf{X} + \mathbf{X}^T), \qquad \mathbf{X}_{\text{asym}} = \frac{1}{2}(\mathbf{X} - \mathbf{X}^T).$$

Evidently

$$\mathbf{m}^T \mathbf{X}_{\text{sym}} \mathbf{m} = \mathbf{m}^T \mathbf{X} \mathbf{m}, \tag{22a}$$

$$\mathbf{m}^T \mathbf{X}_{\text{asym}} \mathbf{m} = 0. \tag{22b}$$

Since $\overset{\circ}{\mathbf{R}}$ and $\overset{\circ}{\mathbf{T}}$ are antisymmetric, we have

$$(\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}})_{\text{sym}} = \frac{1}{2}(\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}} + \overset{\circ}{\mathbf{R}}\overset{\circ}{\mathbf{T}}), \qquad (\overset{\circ}{\mathbf{T}}\mathbf{B})_{\text{sym}} = \frac{1}{2}(\overset{\circ}{\mathbf{T}}\mathbf{B} - \mathbf{B}^T\overset{\circ}{\mathbf{T}}). \tag{23}$$

Denote by $\mathbf{C}$ the symmetric part of $\mathbf{A}^T \overset{\circ}{\mathbf{T}}(\overset{\circ}{\mathbf{R}} + \mathbf{B})\mathbf{A}$. In view of (23), we have

$$\mathbf{C} = \frac{1}{2}\mathbf{A}^T(\overset{\circ}{\mathbf{T}}\overset{\circ}{\mathbf{R}} + \overset{\circ}{\mathbf{R}}\overset{\circ}{\mathbf{T}} + \overset{\circ}{\mathbf{T}}\mathbf{B} - \mathbf{B}^T\overset{\circ}{\mathbf{T}})\mathbf{A}. \tag{24}$$

Let

$$\mathbf{V} = \mathbf{A}^T \overset{\circ}{\mathbf{T}} \mathbf{A}. \tag{25}$$

On account of (21), (22a) and (24), we can write

$$\boxed{\mathbf{m}^T \mathbf{C} \mathbf{m} + \mathbf{m}^T \mathbf{V} \dot{\mathbf{m}} = 0.} \tag{26}$$

A constraint similar to that of (26), termed *the first-order expansion of the fundamental motion equation*, is derived by Viéville and Faugeras [12]. In contrast with the above, however, it takes the form of an approximative equality rather than a strict equality.

In view of (25) and the antisymmetry of $\overset{\circ}{\mathbf{T}}$, $\mathbf{V}$ is antisymmetric. Hence, for some vector $\mathbf{v} = (v_1, v_2, v_3)^T$, $\mathbf{V}$ can be written as

$$\mathbf{V} = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}.$$

$\mathbf{C}$ is symmetric, and hence it is uniquely determined by the entries $c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{33}$. Let $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$ be the *joint projectivisation* of $\mathbf{C}$ and $\mathbf{V}$, that is, the point in the 8-dimensional real projective space whose homogeneous coordinates are formed by all the independent entries of $\mathbf{C}$ and $\mathbf{V}$. More specifically,

$$\boldsymbol{\pi}(\mathbf{C}, \mathbf{V}) = (c_{11} : c_{12} : c_{13} : c_{22} : c_{23} : c_{33} : v_1 : v_2 : v_3).$$

Clearly, $\boldsymbol{\pi}(\lambda\mathbf{C}, \lambda\mathbf{V}) = \boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$ for any non-zero scalar λ. Thus knowing $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$ amounts to knowing $\mathbf{C}$ and $\mathbf{V}$ to within a common scalar factor.

It is important to realise that, by applying (26), $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$ can be determined directly from image data. Namely, if, at any given instant t, we supply sufficiently many independent $\mathbf{m}_i(t)$ and $\dot{\mathbf{m}}_i(t)$, then $\mathbf{C}(t)$ and $\mathbf{V}(t)$ can be determined, up to a common scalar factor, from the following system of equations:

$$\mathbf{m}_i(t)^T \mathbf{C}(t) \mathbf{m}_i(t) + \mathbf{m}_i(t)^T \mathbf{V}(t) \dot{\mathbf{m}}_i(t) = 0. \tag{27}$$

These equations are linear in the entries of $\mathbf{C}(t)$ and $\mathbf{V}(t)$. Note that, in view of (22b), the antisymmetric part of $\mathbf{A}(t)^T\overset{\circ}{\mathbf{T}}(t)(\overset{\circ}{\mathbf{R}}(t)+\mathbf{B}(t))\mathbf{A}(t)$ cannot be found along similar lines.

Since $\boldsymbol{\pi}(\mathbf{C},\mathbf{V})$ is a member of the 8-dimensional projective space, we see that at most 8 key parameters may be determined from $\boldsymbol{\pi}(\mathbf{C},\mathbf{V})$. In fact, only 7 key parameters can be inferred on the basis of $\boldsymbol{\pi}(\mathbf{C},\mathbf{V})$. This is a consequence of $\mathbf{C}$ and $\mathbf{V}$ not being independent. To see this, note first that, by (24) and (25), we have

$$\mathbf{C} = \frac{1}{2}\Big[\mathbf{V}\mathbf{A}^{-1}(\overset{\circ}{\mathbf{R}} + \mathbf{B})\mathbf{A} + \mathbf{A}^T(\overset{\circ}{\mathbf{R}} - \mathbf{B}^T)(\mathbf{A}^T)^{-1}\mathbf{V}\Big]. \tag{28}$$

Denote by $\|\mathbf{v}\|$ the length of $\mathbf{v}$, that is $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$, and set

$$\mathbf{P} = \mathbf{I} + \|\mathbf{v}\|^{-2}\mathbf{V}^2.$$

It is readily verified that

$$\|\mathbf{v}\|^2\mathbf{P} = \|\mathbf{v}\|^2\mathbf{I} + \mathbf{V}^2 = (v_1\mathbf{v} \mid v_2\mathbf{v} \mid v_3\mathbf{v}).$$

A straightforward computation employing this identity shows that

$$\mathbf{P}\mathbf{V} = \mathbf{V}\mathbf{P} = \mathbf{0}.$$

Now, using the latter equation and (28), we immediately find that $\mathbf{C}$ and $\mathbf{V}$ are interrelated by means of the identity

$$\mathbf{P}\mathbf{C}\mathbf{P} = \mathbf{0}.$$

5 Special case: free focal length

We now introduce some intrinsic parameters into our analysis. This will amount to deciding which camera parameters will be known or *free*, or equivalently to adopting a particular form of the intrinsic-parameter matrix $\mathbf{A}$. We define a free parameter to be one that is unknown and which may vary continuously with time.

Assume that the focal length is free and the principal point is fixed and known. In this situation, for each time instant t, $\mathbf{A}(t)$ is given by

$$\mathbf{A}(t) = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & -f(t) \end{pmatrix},$$

where u_0 and v_0 are the coordinates of the known principal point, and $f(t)$ is the focal length at time t. Omitting in notation the dependence on time, observe first that $\mathbf{A}$ can be represented as

$$\mathbf{A} = \mathbf{A}_1\mathbf{A}_2,$$

where

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -f \end{pmatrix}, \qquad \mathbf{A}_2 = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let

$$\mathbf{C}_1 = (\mathbf{A}_2^{-1})^T \mathbf{C} \mathbf{A}_2^{-1}, \qquad \mathbf{V}_1 = (\mathbf{A}_2^{-1})^T \mathbf{V} \mathbf{A}_2^{-1}.$$

Letting $\mathbf{B}_1$ be the matrix function obtained from (20) by substituting $\mathbf{A}_1$ for $\mathbf{A}$, and taking into account that $\dot{\mathbf{A}}_2 = 0$, we find that

$$\mathbf{B} = \dot{\mathbf{A}}\mathbf{A}^{-1} = \dot{\mathbf{A}}_1\mathbf{A}_2(\mathbf{A}_1\mathbf{A}_2)^{-1} = \mathbf{B}_1.$$

Using this identity, it is easy to verify that $\mathbf{C}_1$ and $\mathbf{V}_1$ satisfy (24) and (25), respectively, provided $\mathbf{A}$ and $\mathbf{B}$ in these equations are replaced by $\mathbf{A}_1$ and $\mathbf{B}_1$. Therefore, passing to $\mathbf{A}_1$, $\mathbf{C}_1$ and $\mathbf{V}_1$ in lieu of $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{V}$, respectively, we may always assume that $u_0 = v_0 = 0$. Henceforth we shall assume that such an initial reduction has been made, letting effectively $\mathbf{A}_1$, $\mathbf{C}_1$ and $\mathbf{V}_1$ be equal to $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{V}$, respectively. We then have

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/f \end{pmatrix}, \qquad \dot{\mathbf{A}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\dot{f} \end{pmatrix},$$

and further

$$\mathbf{A}^{-1}\mathbf{B} = \mathbf{A}^{-1}\dot{\mathbf{A}}\mathbf{A}^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\dot{f}/f^2 \end{pmatrix}.$$

Moreover, in view of (11b),

$$\mathbf{A}^{-1}\mathring{\mathbf{R}} = \begin{pmatrix} 0 & \mathring{\gamma} & -\mathring{\beta} \\ -\mathring{\gamma} & 0 & \mathring{\alpha} \\ -\mathring{\beta}/f & \mathring{\alpha}/f & 0 \end{pmatrix},$$

and so

$$\mathbf{A}^{-1}(\mathring{\mathbf{R}} + \mathbf{B}) = \begin{pmatrix} 0 & \mathring{\gamma} & -\mathring{\beta} \\ -\mathring{\gamma} & 0 & \mathring{\alpha} \\ -\mathring{\beta}/f & \mathring{\alpha}/f & -\dot{f}/f^2 \end{pmatrix}.$$

Letting

$$\mathbf{S} = \mathbf{A}^{-1}(\mathring{\mathbf{R}} + \mathbf{B})\mathbf{A},$$

we clearly have

$$\mathbf{S} = \begin{pmatrix} 0 & \mathring{\gamma} & +f\mathring{\beta} \\ -\mathring{\gamma} & 0 & -f\mathring{\alpha} \\ -\mathring{\beta}/f & \mathring{\alpha}/f & \dot{f}/f \end{pmatrix}. \tag{29}$$

Now (28) can be rewritten as

$$\mathbf{C} = \frac{1}{2}(\mathbf{V}\mathbf{S} - \mathbf{S}^T\mathbf{V}). \tag{30}$$

Since $\mathbf{C}$ is symmetric, with six independent entries, the above matrix equation can be seen as a system of six inhomogeneous linear equations for the entries of $\mathbf{S}$ treated as unknowns. Of these only five equations are independent, as $\mathbf{C}$ and $\mathbf{V}$ are interrelated. As we shall see shortly, we can use (30) to express $\overset{\circ}{\alpha}$, $\overset{\circ}{\beta}$, $\overset{\circ}{\gamma}$, f and $\dot{f}$ in terms of $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$. Once f is determined, $\mathbf{A}$ becomes known, and $\overset{\circ}{\mathbf{T}}$ can next be found by using the equality

$$\overset{\circ}{\mathbf{T}} = (\mathbf{A}^T)^{-1}\mathbf{V}\mathbf{A}^{-1}, \tag{31}$$

which immediately follows from (25). Obviously, the resulting formulae for the entries of $\overset{\circ}{\mathbf{T}}$, or equivalently, in view of (11a), for the translational velocity $(\overset{\circ}{x}, \overset{\circ}{y}, \overset{\circ}{z})^T$ will be linear in the entries of $\mathbf{V}$. Therefore the projectivised translational velocity $(\overset{\circ}{x} : \overset{\circ}{y} : \overset{\circ}{z})$, or equivalently the direction of the translational part of the camera's ego-motion, will be uniquely expressed in terms of $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$. In this way, we shall be able to determine 7 key parameters $\overset{\circ}{\alpha}$, $\overset{\circ}{\beta}$, $\overset{\circ}{\gamma}$, f, $\dot{f}$ and $(\overset{\circ}{x} : \overset{\circ}{y} : \overset{\circ}{z})$. (Note that $(\overset{\circ}{x} : \overset{\circ}{y} : \overset{\circ}{z})$ accounts for 2 parameters, being a member of the 2-dimensional real projective space.) Explicit formulae for $\overset{\circ}{\alpha}$, $\overset{\circ}{\beta}$, $\overset{\circ}{\gamma}$, f, $\dot{f}$ and $(\overset{\circ}{x} : \overset{\circ}{y} : \overset{\circ}{z})$ are deferred to the next section.

6 Explicit formulae computation

Set

$$\delta_1 = \frac{\overset{\circ}{\alpha}}{f}, \qquad \delta_2 = \frac{\overset{\circ}{\beta}}{f}, \qquad \delta_3 = \overset{\circ}{\gamma}, \qquad \delta_4 = f^2, \qquad \delta_5 = \frac{\dot{f}}{f}. \tag{32}$$

In view of (29) and (30), we have

$$\begin{aligned} c_{11} &= -v_2\delta_2 + v_3\delta_3, \\ 2c_{12} &= v_2\delta_1 + v_1\delta_2, \\ c_{22} &= -v_1\delta_1 + v_3\delta_3. \end{aligned}$$

Hence

$$\begin{aligned} \delta_1 &= \frac{2c_{12}v_2 - (c_{22} - c_{11})v_1}{v_1^2 + v_2^2}, \\ \delta_2 &= \frac{2c_{12}v_1 + (c_{22} - c_{11})v_2}{v_1^2 + v_2^2}, \\ \delta_3 &= \frac{c_{11}v_1^2 + 2c_{12}v_1v_2 + c_{22}v_2^2}{v_3(v_1^2 + v_2^2)}. \end{aligned} \tag{33}$$

Note that the expressions on the right-hand side of the above equalities are homogeneous of degree 0 in the (essential) entries of $\mathbf{C}$ and $\mathbf{V}$, meaning that these expressions are uniquely representable in terms of $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$. Assuming—as we now may—that δ_1, δ_2, δ_3 are known, we again use (29) and (30) to derive the following formulae for δ_4 and δ_5:

$$\begin{aligned} 2c_{13} &= v_3\delta_1\delta_4 + v_2\delta_5 - v_1\delta_3, \\ 2c_{23} &= v_3\delta_2\delta_4 - v_1\delta_5 - v_2\delta_3, \\ c_{33} &= -(v_1\delta_1 + v_2\delta_2)\delta_4. \end{aligned} \tag{34}$$

These three equations in δ_4 and δ_5 are not linearly independent. To determine δ_4 and δ_5 in an efficient way, we proceed as follows. Let $\boldsymbol{\delta} = (\delta_4, \delta_5)^T$, and let $\mathbf{d} = (d_1, d_2, d_3)^T$ be such that

$$d_1 = 2c_{13} + v_1\delta_3, \qquad d_2 = 2c_{23} + v_2\delta_3, \qquad d_3 = c_{33},$$

and let

$$\mathbf{D} = \begin{pmatrix} v_3\delta_1 & v_2 \\ v_3\delta_2 & -v_1 \\ -v_1\delta_1 - v_2\delta_2 & 0 \end{pmatrix}.$$

With this notation, (34) can be rewritten as

$$\mathbf{D}\boldsymbol{\delta} = \mathbf{d},$$

whence

$$\boldsymbol{\delta} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{d}.$$

More explicitly, we have the following formulae:

$$\begin{aligned} \delta_4 &= \frac{1}{\Gamma}\Big(v_1v_3d_1 + v_2v_3d_2 - (v_1^2 + v_2^2)d_3\Big), \\ \delta_5 &= \frac{1}{\Gamma}\Big((v_1v_2\delta_1 + (v_2^2 + v_3^2)\delta_2)d_1 - ((v_1^2 + v_3^2)\delta_1 + v_1v_2\delta_2)d_2 \\ &\qquad + (v_2v_3\delta_1 - v_1v_3\delta_2)d_3\Big), \end{aligned} \tag{35}$$

where $\Gamma = (v_1^2 + v_2^2 + v_3^2)(v_1\delta_1 + v_2\delta_2)$. Again the expressions on the right-hand side of the above formulae are homogeneous of degree 0 in the entries of $\mathbf{C}$ and $\mathbf{V}$, and so these expressions are univalent functions of $\boldsymbol{\pi}(\mathbf{C}, \mathbf{V})$. With formulae (33) and (35) at hand, the parameters $\mathring{\alpha}$, $\mathring{\beta}$, $\mathring{\gamma}$, f and $\dot{f}$ can now be determined by employing the following equalities resulting from (32):

$$\mathring{\alpha} = \delta_1\sqrt{\delta_4}, \qquad \mathring{\beta} = \delta_2\sqrt{\delta_4}, \qquad \mathring{\gamma} = \delta_3, \qquad f = \sqrt{\delta_4}, \qquad \dot{f} = \delta_5\sqrt{\delta_4}.$$

Having found f, we now use (11a) and (31) to express the translational velocities as

$$\mathring{x} = -\frac{v_1}{f}, \qquad \mathring{y} = -\frac{v_2}{f}, \qquad \mathring{z} = v_3.$$

Hence, we finally obtain

$$(\mathring{x} : \mathring{y} : \mathring{z}) = (-v_1 : -v_2 : fv_3).$$

7 Conclusion

In this paper, we have considered, from an analytical perspective, the problem of determining the ego-motion of a camera having free intrinsic parameters. We have presented a solution based on a differential form of the time-dependent epipolar equation. This latter form was derived by using a variant of the time-derivative of the fundamental matrix for a single mobile camera. Starting from the differential time-dependent epipolar equation, closed-form expressions for the ego-motion parameters, as well as the focal length and its derivative, were obtained.

A Notation semantics

The notation employed in this work differs from the standard notation of Faugeras et al. [2] (henceforth termed the Faugeras notation). Recall that $\mathbf{F}$, $\mathbf{E}$, $\mathbf{T}$, $\mathbf{R}$ and $\mathbf{A}$ denote in this work the fundamental, essential, translation, rotation and intrinsic-parameter matrices, respectively. Let the corresponding matrices of Faugeras be denoted F, E, T, R and A.

Herein, the epipolar equation (already given in (1)) has the form

$$\mathbf{m}^T\mathbf{F}\mathbf{m}' = 0, \qquad (36)$$

where

$$\mathbf{F} = \mathbf{A}^T\mathbf{T}\mathbf{R}\mathbf{A}'.$$

This contrasts with Faugeras, where

$$\mathbf{m}'^T F\mathbf{m} = 0, \qquad (37)$$

and

$$F = A'^{-T}TRA^{-1}.$$

We note that the epipolar relationship in (36) and (37) is expressed with respect to the left and right camera coordinate systems, respectively. In this sense, $\mathbf{TR}$ and TR act in opposite directions from different coordinate systems. We also observe that the intrinsic parameter matrix $\mathbf{A}$ takes a more convenient form than A. However, this is at the cost of its less convenient role in the process of image projection. A point $(X, Y, Z)^T$ in 3D-space maps to an image point (u, v) according to the equation

$$Z\mathbf{m} = \det(\mathbf{A})\mathbf{A}^{-1}\mathbf{M},$$

where $\mathbf{m} = (u, v, 1)^T$ and $\mathbf{M} = (X, Y, Z)^T$, this contrasting with the more natural relation

$$Z\mathbf{m} = A\mathbf{M}.$$

The full list of notational relationships is as follows:

$$\begin{aligned}
\mathbf{F} &= \sqrt{\det(A)\det(A')}F^T \\
\mathbf{E} &= E^T \\
\mathbf{R} &= R^T \\
\mathbf{T} &= -R^T T R \\
\mathbf{A} &= \pm\sqrt{\det(A)}A^{-1} \\
\mathbf{A}' &= \pm\sqrt{\det(A')}A'^{-1},
\end{aligned}$$

or, alternatively,

$$F = \frac{1}{\det(\mathbf{A})\det(\mathbf{A}')}\mathbf{F}^T$$

$$
\begin{aligned}
E &= \mathbf{E}^T \\
R &= \mathbf{R}^T \\
T &= -\mathbf{R}^T \mathbf{T} \mathbf{R} \\
A &= \det(\mathbf{A}) \mathbf{A}^{-1} \\
A' &= \det(\mathbf{A}') \mathbf{A}'^{-1}.
\end{aligned}
$$

References

[1] M. J. Brooks, L. de Agapito, D. Q. Huynh, and L. Baumela, *Direct methods for self-calibration of a moving stereo head*, Computer Vision—ECCV '96 (Fourth European Conference on Computer Vision, Cambridge, UK, April 14–18, 1996) (B. Buxton and R. Cipolla, eds.), Lecture Notes in Computer Science, vol. 1065, Springer, Berlin, 1996, pp. 415–426.

[2] O. D. Faugeras, Q. T. Luong, and S. J. Maybank, *Camera self-calibration: Theory and experiments*, In Sandini [11], pp. 321–334.

[3] N. C. Gupta and L. N. Kanal, *3-D motion estimation from motion field*, Artificial Intelligence **78** (1995), no. 1-2, 45–86.

[4] R. I. Hartley, *Estimation of relative camera positions for uncalibrated cameras*, In Sandini [11], pp. 579–587.

[5] D. J. Heeger and A. D. Jepson, *Subspace methods for recovering rigid motion I: algorithm and implementation*, International Journal of Computer Vision **7** (1992), no. 2, 95–117.

[6] K. Kanatani, *3-D interpretation of optical flow by renormalization*, International Journal of Computer Vision **11** (1993), no. 3, 267–282.

[7] ______, *Geometric Computation for Machine Vision*, Clarendon Press, Oxford, 1993.

[8] Q.-T. Luong, R. Deriche, O. D. Faugeras, and T. Papadopoulo, *On determining the fundamental matrix: analysis of different methods and experimental results*, Tech. Report 1894, INRIA, April 1993.

[9] S. J. Maybank and O. D. Faugeras, *A theory of self-calibration of a moving camera*, International Journal of Computer Vision **8** (1992), no. 2, 123–151.

[10] H.-P. Pan, M. J. Brooks, and G. N. Newsam, *Image resituation: initial theory*, Videometrics IV (Philadelphia, Pennsylvania, USA, Oct. 25–26, 1995) (S. F. El-Hakim, ed.), Proceedings SPIE – The International Society for Optical Engineering, vol. 2598, 1995, pp. 162–173.

[11] G. Sandini (ed.), *Computer Vision—ECCV '92*, Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19–22, 1992, Springer, Berlin, 1992.

[12] T. Viéville and O. D. Faugeras, *Motion analysis with a camera with unknown, and possibly varying intrinsic parameters*, Proceedings of the Fifth International Conference on Computer Vision (Cambridge, MA, June 1995), IEEE Computer Society Press, Los Alamitos, CA, 1995, pp. 750–756.

[13] Z. Zhang, Q. T. Luong, and O. D. Faugeras, *Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction.*, Tech. Report 2079, INRIA, June 1994.