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 single, mobile camera. Differential forms of the epipolar equation are thereby obtained. This may be seen as a recasting of the recent work of Viéville and Faugeras [11] 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

\[ m^T F m' = 0, \]  

(1)

where \( m \) and \( m' \) represent corresponding points in the images obtained by left and right cameras, respectively, expressed in terms of homogeneous coordinates, and \( F \) is the fundamental matrix influenced by both extrinsic and intrinsic imaging factors, henceforth termed the key parameters \[7\]. (Note that a slightly non-standard notation is used here, as described in Appendix A.) Given sufficiently many non-degenerate corresponding points, it is sometimes possible, via a process of self-calibration, to determine various of the key parameters \[2, 8\]. It is well known that using corresponding points extracted from a single image pair, at most 7 key parameters may be determined. These might, for example, comprise 5 relative orientation parameters and 2 focal lengths (e.g. see \[3, 9\]).

Our aim in this work is to introduce into (1) a dependency on time and to establish relationships taking form of 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 \[11\] into an analytical framework. For related work dealing with ego-motion of a calibrated camera, see for example \[4–6\]. (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

\[ m^T(t) F(t) m'(t) = 0, \]  

(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 in a stereo setup. At a later time, this information may again be recomputed. If the key parameters are unchanged, then so too will be the fundamental matrix. 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 (for 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) enables then recovery of some of the key parameters at various times. Any variation in these key parameters conveys information about the motion of one camera relative to another. Again, however, no information may be inferred about the motion of either camera relative to a fixed frame of reference, or for that matter the ego-motion of either camera.

Respite is to be found by considering the epipolar equation associated with two snapshots in time taken by a single moving camera. In this way, a form
of discrete motion analysis of the stereo pair may be undertaken, and this has
been pursued by Zhang et al. [12] and Brooks et al. [1]. A limiting case of the
analysis of discrete motion leads to consideration of a differential form of the
epipolar equation. This in turn opens up the possibility of determining the way
in which the key parameters change over time. That is, a differential form of (2)
offers the possibility of computing the instantaneous changes in both relative
orientation and intrinsic parameters, expressed perhaps in terms of the location
and movement of various image points.

Differentiating (2) with respect to time, we have

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

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

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

There is now the possibility of computing over time the relative orientation of
the right camera relative 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 camera at successive time intervals.
The limiting case, where the time interval tends to zero, might then permit
computation of both the ego-motion and the intrinsic parameters of the camera.
(Note that, as is well known, results pertaining to camera ego-motion and a
stationary scene are equally applicable to a stationary camera and a moving,
rigid scene.)

In order to use the time-dependent epipolar equation (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 [12]
and [1] involved stereo pairs generated by a pair of cameras at successive, dis-
crete times. Of interest here is to contemplate the limiting case of the time
difference between images tending to zero (as in [11]). Note, immediately, that
the following equation holds little value:

$$m^T(t)F(t)m(t) = 0.$$  

This deals merely with identical left and right images and points. In this sit-
uation, we clearly have $F(t) = 0$. 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:

$$m^T(I(t_1))F(I(t_1), I'(t_2))m'(I'(t_2)) = 0.$$  

(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 $m$ and $m'$ are images of a fixed 3D point in space. This equation makes explicit the dependencies of the fundamental matrix $F$.

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 $F$ 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

$$m^T(I(t_1))F(I(t_1), I(t_2))m(I(t_2)) = 0.$$  

Dropping the now superfluous image notation $I$, we obtain

$$m^T(t_1)F(t_1, t_2)m(t_2) = 0. \quad (4)$$

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 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 $F$. Clearly, as $t_1$ and $t_2$ vary, $F(t_1, t_2)$ will also vary, with $F(t_1, t_2)$ tending to 0 as $t_2 \to t_1$. Moreover, the derivative of $F(t_1, t_2)$ will at all times be defined, including at time $t_1 = t_2$. The time-derivatives of $F$ 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 $X(t_1, t_2)$, we write for each $t$

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

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

$$F(t) = 0. \quad (6)$$
Differentiating (4) with respect to $t_2$, we obtain
\[
m^T(t_1) \frac{\partial F}{\partial t_2}(t_1, t_2)m(t_2) + m^T(t_1)F(t_1, t_2)\dot{m}(t_2) = 0,
\]
whence, on letting $t_1 = t_2 = t$, we have
\[
m^T(t)\ddot{F}(t)m(t) + m^T(t)F(t)\dot{m}(t) = 0.
\]
Omitting the notational dependency on time, and using (6), we may rewrite the latter equation as
\[
m^T \dddot{F} m = 0.
\]
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
\[
m^T(t_1) \frac{\partial^2 F}{\partial t_2^2}(t_1, t_2)m(t_2) + 2m^T(t_1) \frac{\partial F}{\partial t_2}(t_1, t_2)\dot{m}(t_2)
+ m^T(t_1)F(t_1, t_2)\ddot{m}(t_2) = 0,
\]
whence, on letting $t_1 = t_2 = t$, we have
\[
m^T(t)\dddot{F}(t)m(t) + 2m^T(t)\ddot{F}(t)\dot{m}(t) + m^T(t)F(t)\dddot{m}(t) = 0,
\]
and hence, in view of (6),
\[
m^T \dddddot{F} m + 2m^T \dddot{F} \dot{m} = 0. \tag{7}
\]
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, $\ddot{m}$ having fallen away in the derivation.

3 Elaborating the second differential form

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

The fundamental matrix, $F(t_1, t_2)$, for a single camera may be expressed in terms of intrinsic parameter matrices $A(t_1)$ and $A(t_2)$, a rotation matrix $R(t_1, t_2)$ and a translation matrix $T(t_1, t_2)$ (e.g. see [7]). Specifically, we have that
\[
F(t_1, t_2) = A^T(t_1)E(t_1, t_2)A(t_2), \tag{8}
\]
where $E(t_1, t_2)$ is the essential matrix defined as
\[
E(t_1, t_2) = T(t_1, t_2)R(t_1, t_2). \tag{9}
\]
Note that the intrinsic parameters within $A(t)$ may vary continuously with time. Henceforth any parameter that is unknown and may vary continuously with time will be termed free. $T(t_1, t_2)$ is an antisymmetric matrix associated with the baseline vector $(x(t_1, t_2), y(t_1, t_2), z(t_1, t_2))^T$ that connects the optical centres of images captured at times $t_1$ and $t_2$; it takes the form

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

$R(t_1, t_2)$ describes the relative rotation of the camera and is given by

$$
R(t_1, t_2) = R_1(\alpha)R_2(\beta)R_3(\gamma),
$$

where the component matrices

$$
R_1(\alpha) = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos \alpha & \sin \alpha \\
0 & -\sin \alpha & \cos \alpha
\end{pmatrix},
$$

$$
R_2(\beta) = \begin{pmatrix}
\cos \beta & 0 & -\sin \beta \\
0 & 1 & 0 \\
\sin \beta & 0 & \cos \beta
\end{pmatrix},
$$

$$
R_3(\gamma) = \begin{pmatrix}
\cos \gamma & \sin \gamma & 0 \\
-\sin \gamma & \cos \gamma & 0 \\
0 & 0 & 1
\end{pmatrix}
$$

correspond to counter clockwise rotations about the camera-centered coordinate axes $x$, $y$, and $z$ by the angles $\alpha$, $\beta$ and $\gamma$, respectively. Here, for convenience, the dependency of $\alpha$, $\beta$ and $\gamma$ upon $(t_1, t_2)$ is left implicit. An explicit form of the matrices $A(t)$ will be given in a subsequent section.

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

$$
T(t) = 0, \quad (10a)
$$

$$
R(t) = I. \quad (10b)
$$

A straightforward computation then shows that

$$
\mathring{T}(t) = \begin{pmatrix}
\mathring{0} & -\mathring{z}(t) & \mathring{y}(t) \\
\mathring{z}(t) & 0 & -\mathring{\xi}(t) \\
-\mathring{y}(t) & \mathring{\xi}(t) & 0
\end{pmatrix},
$$

$$
\mathring{R}(t) = \begin{pmatrix}
0 & \mathring{\gamma}(t) & -\mathring{\beta}(t) \\
-\mathring{\gamma}(t) & 0 & \mathring{\alpha}(t) \\
\mathring{\beta}(t) & -\mathring{\alpha}(t) & 0
\end{pmatrix}. \quad (11b)
$$
The vectors $(\mathbf{\hat{x}}(t), \mathbf{\hat{y}}(t), \mathbf{\hat{z}}(t))^T$ and $(\mathbf{\ddot{\alpha}}(t), \mathbf{\ddot{\beta}}(t), \mathbf{\ddot{\gamma}}(t))^T$ associated with $\mathbf{T}(t)$ and $\mathbf{R}(t)$ capture the instantaneous translational and angular velocities of camera ego-motion, respectively. Observe that both $\mathbf{T}$ and $\mathbf{R}$ are antisymmetric.

Additionally, matrix $\mathbf{\ddot{\mathbf{E}}}$ is readily shown to be antisymmetric.

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

$$\mathbf{E}(t) = 0.$$  \hspace{1cm} (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),

$$\mathbf{\dot{E}}(t) = \mathbf{\dot{T}}(t) \mathbf{R}(t) + \mathbf{T}(t) \mathbf{\dot{R}}(t) = \mathbf{\ddot{T}}(t).$$  \hspace{1cm} (13)

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

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

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

$$\mathbf{\ddot{F}}(t) = \mathbf{A}^T(t) \mathbf{\dot{E}}(t) \mathbf{A}(t) = \mathbf{A}^T(t) \mathbf{\ddot{T}}(t) \mathbf{A}(t).$$  \hspace{1cm} (14)

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

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

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

$$\mathbf{\dddot{F}}(t) = \mathbf{A}^T(t) \mathbf{\dddot{E}}(t) \mathbf{A}(t) + 2 \mathbf{A}^T(t) \mathbf{\dddot{E}}(t) \mathbf{\ddot{A}}(t).$$  \hspace{1cm} (15)

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

$$\mathbf{E} = \mathbf{T} \mathbf{R} + \mathbf{\ddot{\mathbf{R}}} + \mathbf{\dddot{\mathbf{R}}} = \mathbf{\ddot{T}} + 2 \mathbf{\dddot{T}}.$$  \hspace{1cm} (16)

In view of (14),

$$\mathbf{m}^T \mathbf{\dddot{F}} \mathbf{\ddot{m}} = \mathbf{m}^T \mathbf{A}^T \mathbf{T} \mathbf{\ddot{A}} \mathbf{\ddot{m}}.$$  \hspace{1cm} (17)

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

$$\mathbf{m}^T \mathbf{\dddot{F}} \mathbf{\ddot{m}} = \mathbf{m}^T \mathbf{A}^T \mathbf{T} \mathbf{\ddot{A}} \mathbf{\ddot{m}} + 2 \mathbf{m}^T \mathbf{A}^T \mathbf{\dddot{R}} \mathbf{\ddot{A}} \mathbf{\ddot{m}} + 2 \mathbf{m}^T \mathbf{A}^T \mathbf{\dddot{T}} \mathbf{\ddot{A}} \mathbf{\ddot{m}}.$$  \hspace{1cm} (18)

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

$$\mathbf{m}^T \mathbf{A}^T \mathbf{\dddot{F}} \mathbf{\ddot{m}} = 0.$$
Therefore (18) can be rewritten as

\[ m^T \tilde{F} m = 2m^T A^T \tilde{R} A m + 2m^T A^T \tilde{\dot{R}} A m. \]

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

\[ m^T A^T \tilde{R} A m + m^T A^T \tilde{\dot{R}} A m + m^T A^T \tilde{\dot{T}} A m = 0. \]  

(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 \( \dot{B} = \dot{A} A^{-1} \), we first rewrite (19) as

\[ m^T A^T \tilde{\dot{R}} A m + m^T A^T \tilde{\dot{T}} A m = 0. \]  

(20)

Given a matrix \( X \), denote by \( X_{\text{sym}} \) and \( X_{\text{asym}} \) the symmetric and antisymmetric parts of \( X \) defined, respectively, by

\[ X_{\text{sym}} = \frac{1}{2} (X + X^T), \quad X_{\text{asym}} = \frac{1}{2} (X - X^T). \]

Evidently

\[ m^T X_{\text{sym}} m = m^T X m, \quad m^T X_{\text{asym}} m = 0. \]

(21a)

(21b)

Since \( \tilde{R} \) and \( \tilde{T} \) are antisymmetric, we have

\[ (\tilde{R} \tilde{T})_{\text{sym}} = \frac{1}{2} (\tilde{R} \tilde{T} + \tilde{T} \tilde{R}), \quad (\tilde{T} \tilde{B})_{\text{sym}} = \frac{1}{2} (\tilde{T} \tilde{B} - \tilde{B}^T \tilde{T}). \]  

(22)

Denote by \( C \) the symmetric part of \( A^T \tilde{\dot{R}} (\tilde{R} + \tilde{B}) A \). In view of (22), we have

\[ C = \frac{1}{2} A^T (\tilde{R} \tilde{T} + \tilde{T} \tilde{R} + \tilde{T} \tilde{B} - \tilde{B}^T \tilde{T}) A. \]  

(23)

Let

\[ V = A^T \tilde{\dot{T}} A. \]  

(24)

On account of (20), (21a) and (23), we can write

\[ m^T C m + m^T V \dot{m} = 0. \]  

(25)
A constraint similar to that of (25), termed the first-order expansion of the fundamental motion equation, is derived by Viéville and Faugeras [11]. In contrast with the above, however, it takes the form of an approximate rather than strict equality.

In view of (24) and the antisymmetry of $T$, $V$ is antisymmetric. Hence, for some vector $v = (v_1, v_2, v_3)^T$, $V$ can be written as

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

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

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

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

It is important to realise that, by applying (25), $\pi(C, V)$ can be determined directly from image data. Namely, if, at any given instant $t$, we supply sufficiently many independent $m_i(t)$ and $\dot{m}_i(t)$, then $C(t)$ and $V(t)$ can be determined, up to a common scalar factor, from the following system of equations:

$$m_i(t)^T C(t) m_i(t) + m_i(t)^T V(t) \dot{m}_i(t) = 0.$$  \hfill (26)

These equations are linear in the entries of $C(t)$ and $V(t)$. Note that, in view of (21b), the antisymmetric part of $A(t)^T T(t)(\dot{R}(t) + B(t)) A(t)$ cannot be found along similar lines.

Since $\pi(C, V)$ is a member of the 8-dimensional projective space, we see that at most 8 key parameters may be determined from $\pi(C, V)$. In fact, only 7 key parameters can be inferred on the basis of $\pi(C, V)$. This is a consequence of the fact that the matrices $C$ and $V$ are not independent. To see this, note first that, by (23) and (24), we have

$$C = \frac{1}{2} [VA^{-1}(\dot{R} + B)A + A^T(\dot{R} - B^T)(A^T)^{-1}V].$$  \hfill (27)

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

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

It is readily verified that

$$||v||^2 P = ||v||^2 I + V^2 = \begin{bmatrix} v_1 v \mid v_2 v \mid v_3 v \end{bmatrix}.$$

A straightforward computation employing this identity shows that

$$PV = VP = 0.$$

Now, using the latter equation and (27), we immediately find that $C$ and $V$ are interrelated by means of the identity

$$PCP = 0.$$
5 Special case: free focal length, known principal point

We now introduce some intrinsic parameters into our analysis. This will amount to deciding which camera parameters will be known or will be free, or equivalently to adopting, for each time instant $t$, a particular form of the intrinsic parameter matrix $A(t)$. For reasons of tractability, we shall assume that the focal length is free and the principal point is fixed and known. In this situation, $A(t)$ is given by

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

where $u_0$ and $v_0$ are the coordinates of the principal point, and $f(t)$ is the focal length at time $t$. Omitting in notation the dependence on time, observe first that

$$
A = A_1 A_2,
$$

where

$$
A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -f \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

Noting that

$$
A_2^{-1} = \begin{pmatrix} 1 & 0 & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{pmatrix},
$$

let

$$
C_1 = (A_2^{-1})^T CA_2^{-1}, \quad V_1 = (A_2^{-1})^T VA_2^{-1}.
$$

It is readily verified that $C_1$ and $V_1$ satisfy (23) and (24), respectively, in which $A$ is replaced by $A_1$. Therefore, passing to $A_1$, $C_1$ and $V_1$ in lieu of $A$, $C$ and $V$, 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 $A_1$, $C_1$ and $V_1$ be equal to $A$, $C$ and $V$, respectively. Then

$$
A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} \end{pmatrix}, \quad \dot{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\dot{f} \end{pmatrix},
$$

and further

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

(28)

In view of (11b),

$$
A^{-1}\dot{R} = \begin{pmatrix} 0 & \hat{\gamma} & -\hat{\beta} \\ -\hat{\gamma} & 0 & \hat{\alpha} \\ -\hat{\beta} & \hat{\alpha} & 0 \end{pmatrix}
$$

(29)
Let 

\[ S = A^{-1}(R + B)A. \]

Then

\[
A^{-1}(R + B) = \begin{pmatrix}
0 & \mathring{o} & -\mathring{\beta} \\
-\mathring{\gamma} & 0 & \mathring{\alpha} \\
-\mathring{d} & \mathring{f} & -\mathring{f}^2
\end{pmatrix}
\]

and further

\[
S = \begin{pmatrix}
0 & \mathring{o} & +f\mathring{\beta} \\
-\mathring{\gamma} & 0 & -f\mathring{\alpha} \\
-\mathring{d} & \mathring{f} & \mathring{f}^2
\end{pmatrix}
\]  

(30)

Now (27) can be rewritten as

\[
C = \frac{1}{2}(VS - S^T V).
\]  

(31)

Since \( C \) 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 \( S \) treated as unknowns. Of these only five equations are independent, as \( C \) and \( V \) are interrelated. As we shall see shortly, we can use these five equations to express \( \mathring{\alpha}, \mathring{\beta}, \mathring{\gamma}, f \) and \( \mathring{f} \) in terms of \( \pi(C, V) \). Once \( f \) is determined, \( A \) becomes known, and next \( T \) can be found by using the equality

\[
\mathring{T} = (A^T)^{-1}VA^{-1}
\]

(32)

which immediately follows from (24). Obviously, the resulting formulae for the entries of \( T \), or equivalently, in view of (11a), for the translational velocity \( (\mathring{x}, \mathring{y}, \mathring{z})^T \) will be linear in the entries of \( V \). Therefore the projectivised translational velocity \( (\mathring{x} : \mathring{y} : \mathring{z}) \), or equivalently the direction of the translational part of the camera’s ego-motion, will be uniquely determined in terms of \( \pi(C, V) \).

In this way, we shall be able to determine 7 key parameters \( \mathring{\alpha}, \mathring{\beta}, \mathring{\gamma}, f, \mathring{f} \) and \( (\mathring{x} : \mathring{y} : \mathring{z}) \) (note that \( (\mathring{x} : \mathring{y} : \mathring{z}) \), being a member of the two-dimensional real projective space, accounts for 2 parameters). Explicit formulae for \( \mathring{\alpha}, \mathring{\beta}, \mathring{\gamma}, f, \mathring{f} \) and \( (\mathring{x} : \mathring{y} : \mathring{z}) \) are deferred to the next section.

6 Explicit formulae computation

Write \( S \) as

\[
S = \begin{pmatrix}
s_{11} & s_{12} & s_{13} \\
s_{21} & s_{22} & s_{23} \\
s_{31} & s_{32} & s_{33}
\end{pmatrix}
\]  

(33)
Then
\[
\mathbf{VS} = \begin{pmatrix}
v_2 s_{31} - v_3 s_{21} & v_2 s_{32} - v_3 s_{22} & v_2 s_{33} - v_3 s_{23} \\
v_3 s_{11} - v_1 s_{31} & v_3 s_{12} - v_1 s_{32} & v_3 s_{13} - v_1 s_{33} \\
v_1 s_{21} - v_2 s_{11} & v_1 s_{22} - v_2 s_{12} & v_1 s_{23} - v_2 s_{13}
\end{pmatrix},
\]
and so, by (31),
\[
c_{11} = v_2 s_{31} - v_3 s_{21}, \\
c_{12} = c_{21} = \frac{1}{2} (v_2 s_{32} + v_3 (s_{11} - s_{22}) - v_1 s_{31}), \\
c_{13} = c_{31} = \frac{1}{2} (v_2 (s_{33} - s_{11}) - v_3 s_{23} + v_1 s_{21}), \\
c_{22} = v_3 s_{12} - v_1 s_{32}, \\
c_{23} = \frac{1}{2} (v_3 s_{13} + v_1 (s_{22} - s_{33}) - v_2 s_{12}), \\
c_{33} = v_1 s_{23} - v_2 s_{13}
\]
Set
\[
\delta_1 = \frac{\phi}{f}, \quad \delta_2 = \frac{\theta}{f}, \quad \delta_3 = \frac{\gamma}{f}.
\]
In view of (30) and (33), we have
\[
s_{11} = s_{22} = 0, \quad s_{12} = -s_{21} = \delta_3, \quad s_{31} = -\delta_2 \quad s_{32} = \delta_1.
\]
Using (34), we can now write
\[
c_{11} = -v_2 \delta_2 + v_3 \delta_3, \\
2c_{12} = v_2 \delta_1 + v_1 \delta_2, \\
c_{22} = -v_1 \delta_1 + v_3 \delta_3.
\]
Hence
\[
\delta_1 = \frac{2c_{12} v_2 - (c_{22} - c_{11}) v_1}{v_1^2 + v_2^2}, \\
\delta_2 = \frac{2c_{12} v_1 + (c_{22} - c_{11}) v_2}{v_1^2 + v_2^2}, \\
\delta_3 = \frac{c_{11} v_1^2 + 2c_{12} v_1 v_2 + c_{22} v_2^2}{v_1^2 + v_2^2}.
\]
Set
\[
\delta_4 = f^2, \quad \delta_5 = \frac{f}{f}.
\]
Having established formulae (38) for \(\delta_1, \delta_2, \delta_3\), we now give analogous formulae for \(\delta_4\) and \(\delta_5\). In deriving the latter we shall assume—as we may—that \(\delta_1, \delta_2, \delta_3\) are known. In view of (30) and (33), we have
\[
s_{13} = \delta_2 \delta_4, \quad s_{23} = -\delta_1 \delta_4, \quad s_{33} = \delta_5.
\]
Using (34), we can now write
\[ 2c_{13} = v_3\delta_1\delta_4 + v_2\delta_5 - v_1\delta_3, \]
\[ 2c_{23} = v_3\delta_2\delta_4 - v_1\delta_5 - v_2\delta_3, \]
\[ c_{33} = -(v_1\delta_1 + v_2\delta_2)\delta_4. \] (40)

These three equations for \( \delta_4 \) and \( \delta_5 \) are not linearly independent. To determine from them \( \delta_4 \) and \( \delta_5 \) in an efficient way, we proceed as follows. Let \( \mathbf{\delta} = (\delta_4, \delta_5)^T \), and let \( \mathbf{d} = (d_1, d_2, d_3)^T \) be such that
\[ d_1 = 2c_{13} + v_1\delta_3, \quad d_2 = 2c_{23} + v_2\delta_3, \quad d_3 = c_{33}, \]
and let
\[
D = \begin{pmatrix}
v_3\delta_1 & v_2 \\
v_3\delta_2 & -v_1 \\
-v_1\delta_1 - v_2\delta_2 & 0
\end{pmatrix}.
\]
With this notation, (40) can be rewritten in the form
\[ D\mathbf{\delta} = \mathbf{d}. \] (41)

Now \( \mathbf{\delta} \) is given by
\[ \mathbf{\delta} = (D^T D)^{-1} D^T \mathbf{d}. \] (42)

More explicitly, we have the following formulae:
\[ \delta_4 = \frac{1}{\Gamma} \left( v_1 v_3 d_1 + v_2 v_3 d_2 - (v_1^2 + v_2^2) d_3 \right), \]
\[ \delta_5 = \frac{1}{\Gamma} \left( (v_1 v_2 \delta_1 + (v_2^2 + v_3^2) \delta_2) d_1 - ((v_1^2 + v_3^2) \delta_1 + v_1 v_2 \delta_2) d_2 + (v_2 v_3 \delta_1 - v_1 v_3 \delta_2) d_3 \right), \] (43)
where
\[ \Gamma = (v_1^2 + v_2^2 + v_3^2)(v_1\delta_1 + v_2\delta_2). \]

With formulae (38) and (43) at hand, the parameters \( \alpha, \beta, \gamma, f \) and \( \dot{f} \) can now be determined by employing the following equalities resulting from (35) and (39):
\[ \alpha = \delta_1 \sqrt{\delta_4}, \quad \beta = \delta_2 \sqrt{\delta_4}, \quad \gamma = \delta_3, \quad f = \sqrt{\delta_4}, \quad \dot{f} = \delta_5 \sqrt{\delta_4}. \]

Having found \( f \), we can use (11a) and (32) to express the translational velocities as
\[ \dot{x} = -\frac{v_1}{f}, \quad \dot{y} = -\frac{v_2}{f}, \quad \dot{z} = v_3. \]

Hence, we finally obtain
\[ (\dot{x} : \dot{y} : \dot{z}) = (-v_1 : -v_2 : f v_3). \]
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 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 $F$, $E$, $T$, $R$ and $A$ 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

$$m^T F m' = 0,$$

where

$$F = A^T T R A'.$$

This contrasts with Faugeras, where

$$m'^T F m = 0,$$

and

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

We note that the epipolar relationship in (44) and (45) is expressed with respect to the left and right camera coordinate systems, respectively. In this sense, $TR$ and $TR$ act in opposite directions from different coordinate systems. We also observe that the intrinsic parameter matrix $A$ 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 3-space maps to an image point $(u, v)$ according to the equation

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

where $m = (u, v, 1)^T$ and $M = (X, Y, Z)^T$, this contrasting with the more naturally expressed

$$Z m = AM.$$

The full list of notational relationships is now given:

$$F = \sqrt{\det(A) \det(A')} F^T$$
\[ E = E^T \]
\[ R = R^T \]
\[ T = -R^T TR \]
\[ A = \pm \sqrt{\det(A)A^{-1}} \]
\[ A' = \pm \sqrt{\det(A')A'^{-1}}, \]

or, alternatively,

\[ F = \frac{1}{\det(A)\det(A')} F^T \]
\[ E = E^T \]
\[ R = R^T \]
\[ T = -R^T TR \]
\[ A = \det(A)A^{-1} \]
\[ A' = \det(A')A'^{-1}. \]

References


