Determining the egomotion of an uncalibrated camera from instantaneous optical flow

Michael J. Brooks and Wojciech Chojnacki
Department of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia

Luis Baumela
Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte Madrid, Spain

Received August 26, 1996; accepted February 20, 1997

A procedure is described for self-calibration of a moving camera from instantaneous optical flow. Under certain assumptions this procedure allows the egomotion and some intrinsic parameters of the camera to be determined solely from the instantaneous positions and velocities of a set of image features. The proposed method relies on the use of a differential epipolar equation that relates optical flow to the egomotion and the internal geometry of the camera. A detailed derivation of this equation is presented. This aspect of the work may be seen as a recasting into an analytical framework of the pivotal research of Vieville and Faugeras [Proceedings of the Fifth International Conference on Computer Vision (IEEE Computer Society Press, Los Alamitos, Calif., 1995), pp. 750–756]. The information about the camera's egomotion and internal geometry enters the differential epipolar equation via two matrices. It emerges that the optical flow determines the composite ratio of some of the entries of the two matrices. It is shown that a camera with unknown focal length undergoing arbitrary motion can be self-calibrated by means of closed-form expressions in the composite ratio. The corresponding formulas specify five egomotion parameters as well as the focal length and its derivative. A procedure is presented for reconstructing the viewed scene, up to a scale factor, from the derived self-calibration data and the optical flow data. Experimental results are given to demonstrate the correctness of the approach. © 1997 Optical Society of America [S0740-3232(97)02110-8]

1. INTRODUCTION
There has been considerable interest in recent years in the generation of computer vision algorithms able to operate with uncalibrated cameras. One challenge has been to reconstruct a scene, up to scale, from a stereo pair of images obtained by cameras whose internal geometry is not fully known and whose relative orientation is unknown. Remarkably, such a reconstruction is sometimes attainable solely by consideration of corresponding points (that depict a common scene point) identified within the two images. A key process involved here is that of self-calibration, whereby the unknown relative orientation and intrinsic parameters of the cameras are automatically determined.1,2

In this paper we develop a method for self-calibration of a single moving camera from instantaneous optical flow. Here self-calibration amounts to automatically determining the unknown instantaneous egomotion and intrinsic parameters of the camera and is analogous to self-calibration of a stereo vision setup from corresponding points.

The proposed method of self-calibration rests on a constraint that we term a differential epipolar equation, which relates optical flow to the egomotion and intrinsic parameters of the camera. A substantial portion of the paper is devoted to a detailed derivation of this equation. The differential epipolar equation has as its counterpart in stereo vision the familiar (algebraic) epipolar equation. Whereas the standard epipolar equation incorporates a single fundamental matrix,3,4 the differential epipolar equation incorporates two matrices. These matrices encode information about the egomotion and the internal geometry of the camera. Any sufficiently large subset of an optical flow field determines the composite ratio of some of the entries of these matrices. It emerges that, under certain assumptions, the moving camera can be self-calibrated by means of closed-form expressions evolved from this ratio.

Elaborating on the nature of the self-calibration procedure, assume that a camera moves freely through space and views a static world. (Since we can, of course, compute only relative motion, our technique applies most generally to a moving camera viewing a moving rigid body.) Suppose that the interior characteristics of the camera, except for the focal length, are known and that the focal length is free in that it may vary continuously. We show here that, from instantaneous optical flow, we can compute with closed-form expressions the camera's angular velocity, direction of translation, focal length, and rate of change of focal length. These entities embody seven degrees of freedom, with the angular velocity and the direction of translation, which describe the camera's egomotion, accounting for five degrees of freedom. Note that a full description of the egomotion requires six degrees of freedom. However, as is well known, the speed of translation is not computable without the provision of metric information from the scene. (For example, we are unable to discern solely from a radiating optical flow field...
whether we are rushing toward a planet or moving slowly toward a football. This phenomenon has as its analog in stereo vision the indeterminacy of baseline length from corresponding points.)

Our work is inspired by, and is closely related to, that of Viéville and Faugeras. These authors were the first to introduce an equation akin to what we term here the differential epipolar equation. However, unlike the latter equation, that of Viéville and Faugeras takes the form of an approximation and not of a strict equality. One of our aims here has been to clarify this matter and to place the derivation of the differential epipolar equation and ramifications for self-calibration on a firm analytical footing.

In addition to a self-calibration technique, the paper gives a procedure for carrying out scene reconstruction based on the results of self-calibration and the optical flow. Both methods are tested on an optical flow field derived from a real-world image sequence of a calibration grid. For related work dealing with the egomotion of a calibrated camera see, for example, Refs. 6–9.

2. SCENE MOTION IN THE CAMERA FRAME

In order to extract three-dimensional information from an image, a camera model must be adopted. In this paper the camera is modeled as a pinhole (see Fig. 1). A detailed exposition of the pinhole model including the relevant terminology can be found in Sec. 3 of Ref. 10. To describe the position, orientation, and internal geometry of the camera as well as the image formation process, it is convenient to introduce two coordinate frames. Select a Cartesian (world) coordinate frame \( \Gamma_w \) whose scene configuration will be fixed throughout. Associate with the camera an independent Cartesian coordinate frame \( \Gamma_c \), with origin \( C \) and basis \( \{ e_i \}_{1 \leq i \leq 3} \) of unit orthogonal vectors, so that \( C \) coincides with the optical center, \( e_1 \) and \( e_2 \) span the focal plane, and \( e_3 \) determines the optical axis (see Fig. 1 for a display of the coordinate frames). Ensure that \( \Gamma_c \) and \( \Gamma_w \) are equioriented by swapping two arbitrarily chosen basis vectors of \( \Gamma_w \) if initially the frames are counteroriented. In so doing, we guarantee that the value of the cross product of two vectors is independent of whether the basis of unit orthogonal vectors associated with \( \Gamma_w \) or that associated with \( \Gamma_c \) is used for calculation. For reasons of tractability, \( C \) will be identified with the point in \( \mathbb{R}^3 \) formed by the coordinates of \( C \) relative to \( \Gamma_w \). Similarly, for each \( i \in \{ 1, 2, 3 \} \), \( e_i \) will be identified with the point in \( \mathbb{R}^3 \) formed by the components of \( e_i \) relative to the vector basis of \( \Gamma_w \).

Suppose that the camera undergoes smooth motion with respect to \( \Gamma_w \). At each time instant \( t \), the location of the camera relative to \( \Gamma_w \) is given by \([ C(t), e_1(t), e_2(t), e_3(t) ] \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \). The motion of the camera is then described by the differentiable function \( t \rightarrow [ C(t), e_1(t), e_2(t), e_3(t) ] \). The derivative \( \dot{C}(t) \) captures the instantaneous translational velocity of the camera relative to \( \Gamma_w \) at \( t \). Expanding this derivative with respect to the basis \( \{ e_i(t) \}_{1 \leq i \leq 3} \),

\[
\dot{C}(t) = \sum_{i} v_i(t) e_i(t),
\]

defines \( \mathbf{v}(t) = [v_1(t), v_2(t), v_3(t)]^T \). This vector represents the instantaneous translational velocity of the camera relative to \( \Gamma_c \) at \( t \). Each of the derivatives \( e_i(t) \) can be expanded in a similar fashion, yielding

\[
\dot{e}_i(t) = \sum_{j} \omega_{ji}(t) e_j(t).
\]

The coefficients thus arising can be arranged in the matrix

\[
\Omega(t) = [ \omega_{ji}(t) ]_{1 \leq i,j \leq 3}.
\]

Leaving the dependency of the \( e_i \) on \( t \) implicit, we can express the orthogonality and normalization conditions satisfied by the \( e_i \) as

\[
e_i^T e_j = \delta_{ij},
\]

where

\[
\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.
\]

Differentiating both sides of Eq. (3) with respect to \( t \), we obtain

\[
\dot{e}_i^T e_j + e_i^T \dot{e}_j = 0.
\]

In view of Eq. (2),

\[
\omega_{ji} = e_j^T \dot{e}_i,
\]

which together with the previous equation yields

\[
\omega_{ji} = -\omega_{ij}.
\]

We see then that \( \Omega \) is antisymmetric and as such can be represented as

\[
\Omega = \dot{\omega}
\]

for some vector \( \omega = [ \omega_1, \omega_2, \omega_3 ]^T \), where \( \dot{\omega} \) is defined as

\[
\dot{\omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}.
\]

Writing Eq. (2) as
\[
\begin{align*}
\dot{e}_1 &= \omega_2 e_2 - \omega_3 e_3, \\
\dot{e}_2 &= \omega_1 e_3 - \omega_3 e_1, \\
\dot{e}_3 &= \omega_2 e_1 - \omega_1 e_2,
\end{align*}
\]

Introducing
\[
\eta = \omega_1 e_1 + \omega_2 e_2 + \omega_3 e_3,
\]
and noting that, for any \(z = z_1 e_1 + z_2 e_2 + z_3 e_3\),
\[
\eta \times z = (\omega_2 z_3 - \omega_3 z_2) e_1 + (\omega_3 z_1 - \omega_1 z_3) e_2 + (\omega_1 z_2 - \omega_2 z_1) e_3,
\]
we have, for each \(i \in \{1,2,3\}\),
\[
\dot{e}_i = \eta \times e_i.
\]

It is clear from this system of equations that \(\eta\) represents the instantaneous angular velocity of the camera relative to \(\Gamma_w\). The direction of \(\eta\) determines the axis of the instantaneous rotation of the camera, passing through \(C\), relative to \(\Gamma_w\). Correspondingly, \(\omega\) represents the instantaneous angular velocity of the camera relative to \(\Gamma_e\), with the direction of \(\omega\) determining the axis of the instantaneous rotation of the camera relative to \(\Gamma_e\).

Let \(P\) be a point in space. Identify \(P\) with the point in \(\mathbb{R}^3\) formed by the coordinates of \(P\) relative to \(\Gamma_w\). With the earlier identification of \(C\) and \(e_i\) with respective points of \(\mathbb{R}^3\) still in force, the location of \(P\) relative to \(\Gamma_e\) can be expressed in terms of a coordinate vector \(x = [x_1, x_2, x_3]^T\) determined from the equation
\[
P = \sum_i x_i e_i + C.
\]

Equation (6) can be viewed as the expansion of the vector connecting \(C\) with \(P\), identifiable with the point \(P-C\), relative to the vector basis of \(\Gamma_e\). Suppose that \(P\) is static with respect to \(\Gamma_w\). As the camera moves, the position of \(P\) relative to \(\Gamma_e\) will change accordingly and will be recorded in the function \(t \rightarrow x(t)\). This function satisfies an equation that reflects the kinematics of the moving camera. We derive this equation next.

Differentiating Eq. (6) and taking into account that \(\dot{P} = 0\), we obtain
\[
\sum_i (\dot{x}_i e_i + x_i \dot{e}_i) + \dot{C} = 0.
\]

In view of Eqs. (1) and (2),
\[
\sum_i (\dot{x}_i e_i + x_i \dot{e}_i) + \dot{C} = \sum_i \left( \dot{x}_i e_i + x_i \sum_j \omega_{ij} e_j + v_i e_i \right) = \sum_i \left( \dot{x}_i + \sum_j \omega_{ij} x_j + v_i \right) e_i.
\]

Therefore, for each \(i \in \{1,2,3\}\),
\[
\dot{x}_i + \sum_j \omega_{ij} x_j + v_i = 0,
\]
or in matrix notation
\[
\dot{x} + \Omega x + v = 0.
\]

Coupling this equation with Eq. (4), we obtain
\[
\dot{x} + \omega \times x + v = 0.
\]

This is the equation that governs the evolution of \(x\). Taking into account that \(\dot{\omega} x = \omega \times \dot{x}\), we can also state it in a more traditional form as
\[
\dot{x} + \omega \times x + v = 0.
\]

\[\text{(7)}\]

\section{DIFFERENTIAL EPIPOLAR EQUATION}

The camera image is formed by perspective projection of the viewed scene, through \(C\), onto the plane parallel to the focal plane (again, see Fig. 1). In coordinates relative to \(\Gamma_e\), the image plane is described by \(x \in \mathbb{R}^2: x_3 = -f\), where \(f\) is the focal length. If \(P\) is a point in space, and if \(x\) and \(p\) are the coordinates relative to \(\Gamma_e\) of \(P\) and its image, then
\[
p = -\frac{f}{x_3} x.
\]

Suppose again that \(P\) is static and that the camera moves with respect to \(\Gamma_w\). The evolution of the image of \(P\) will then be described by the function \(t \rightarrow p(t)\). This function is subject to a constraint deriving from Eq. (7). We proceed to determine this constraint.

First, note that Eq. (8) can be equivalently rewritten as
\[
x = \frac{x_3 p}{f},
\]

which immediately leads to
\[
\dot{x} = \frac{x_3 \dot{p} - x_3 \dot{x}_3}{f^2} p - \frac{x_3}{f} \dot{p}.
\]

Next, applying the matrix \(\tilde{v}\) (formed according to definition (5)) to both sides of Eq. (7) and noting that \(\tilde{v} v = 0\), we get
\[
\tilde{v} \dot{x} + \tilde{v} \dot{\omega} x = 0.
\]

Now, in view of Eqs. (9) and (10),
\[
\frac{x_3 \dot{p} - x_3 \dot{x}_3}{f^2} v p - \frac{x_3}{f} \dot{v} p - \frac{x_3}{f} \tilde{v} \dot{\omega} p = 0.
\]

Applying \(\tilde{v}^T\) to both sides of this equation, dropping the summand with \(\tilde{v}^T v p\) on the left-hand side (in view of the antisymmetry of \(\tilde{v}\), we have \(\tilde{v}^T v p = 0\)), and canceling out the common factor \(-x_3/f\) in the remaining summands, we obtain
\[
\tilde{v}^T \dot{v} p + \tilde{v}^T \dot{\omega} p = 0.
\]

This is the sought-after constraint. We call it the differential epipolar equation. This term reflects the fact that Eq. (11) is a limiting case of the familiar epipolar equation in stereo vision. We shall not discuss here the relationship between the two types of epipolar equation, referring the reader to Ref. 11 and its short version, Ref. 12, where an analog of Eq. (11), namely, Eq. (20) below, is derived from the standard epipolar equation by applying a special differentiation operator. We also refer the reader to Ref. 8, where a similar derivation (though it does not involve any special differentiation procedure) is presented in the context of images formed on a sphere. It is due to
a suggestion of Torr\textsuperscript{13} that here we derive the differential epipolar equation from first principles rather than from the standard epipolar equation.

The differential epipolar equation is not the only constraint that can be imposed on functions of the form \( t \rightarrow \mathbf{p}(t) \). As was shown by Aström and Heyden,\textsuperscript{14} for every \( n > 2 \) such functions satisfy an \( n \)-th order differential equation that reduces to the differential epipolar equation when \( n = 2 \). The \( n \)-th equation in the series is the infinitesimal version of the analog of the standard epipolar equation satisfied by a set of corresponding points identified within a sequence of \( n \) images depicting a common scene point. This paper rests solely on the differential epipolar equation, which is the simplest of these equations.

4. ALTERNATIVE FORM OF THE DIFFERENTIAL EPIPOLAR EQUATION

To account for the geometry of the image, it is useful to adopt an image-related coordinate frame \( \Gamma_i \), with origin \( O \) and basis of vectors \( \{\mathbf{e}_i\}_{1 \leq i \leq 2} \), in the image plane. It is natural to align the \( \mathbf{e}_i \) along the sides of pixels and take one of the four corners of the rectangular image boundary for \( O \). In a typical situation when image pixels are rectangular, \( \Gamma_i \) and \( \Gamma_j \) are customarily adjusted so that \( \mathbf{e}_i = s_i\mathbf{e}_i \), where \( s_i \) characterizes the pixel size in the direction of \( \mathbf{e}_i \) in length units of \( \Gamma_i \). Suppose that a point in the image plane has coordinates \( \mathbf{p} = [p_1, p_2, -f]^T \) and \( [m_1, m_2]^T \) relative to \( \Gamma_i \) and \( \Gamma_j \), respectively. If we append to \( [m_1, m_2]^T \) an extra entry equal to 1 to yield the vector \( \mathbf{m} = [m_1, m_2, 1]^T \), then the relation between \( \mathbf{p} \) and \( \mathbf{m} \) can be conveniently written as

\[
\mathbf{p} = \mathbf{A}\mathbf{m},
\]

(12)

where \( \mathbf{A} \) is a \( 3 \times 3 \) invertible matrix called the intrinsic-parameter matrix. With the assumption that \( \mathbf{e}_i = s_i\mathbf{e}_i \) in force, if \( [\mathbf{f}_1, \mathbf{f}_2]^T \) is the \( \Gamma_i \)-based coordinate representation of the principal point (that is, the point at which the optical axis intersects the image plane), then \( \mathbf{A} \) takes the form

\[
\mathbf{A} = \begin{bmatrix}
s_1 & 0 & -s_1 f_1 \\
0 & s_2 & -s_2 f_2 \\
0 & 0 & -f
\end{bmatrix}.
\]

(13)

When pixels are nonrectangular, \( \mathbf{A} \) takes a more complicated form that accounts for one more parameter that encodes shear in the camera axes (see Sec. 3 of Ref. 10).

Differential epipolar equation (11) can be restated for, say, for any given instant, the \( \Gamma_i \)-based vector \([\mathbf{m}^T, \mathbf{m}^T]^T \) in place of the \( \Gamma_i \)-based vector \([\mathbf{p}^T, \mathbf{p}^T]^T \).

This equation, in conjunction with Eq. (12), implies that

\[
\ddot{\mathbf{p}} + \dot{\mathbf{A}}\dot{\mathbf{m}} + \dot{\mathbf{A}}\mathbf{m} = 0.
\]

(14)

This is the differential epipolar equation for optical flow. A similar constraint, termed the first-order expansion of the fundamental motion equation, is derived by quite different means by Viéeville and Faugeras.\textsuperscript{5} In contrast with Eq. (20) above, however, it takes the form of an approximation rather than a strict equality.

In view of Eq. (19) and the antisymmetry of \( \mathbf{\tilde{v}} \), \( \mathbf{W} \) is antisymmetric, so \( \mathbf{W} = \mathbf{W}^T \) for some vector \( \mathbf{w} = [w_1, w_2, w_3]^T \). \( \mathbf{C} \) is symmetric, and hence it is uniquely determined by the entries \( c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{33} \), and \( c_{33} \). Let \( \pi(\mathbf{C}, \mathbf{W}) \) be the joint projective form of \( \mathbf{C} \) and \( \mathbf{W} \), which is, the point in the eight-dimensional real projective space \( \mathbb{P}^7 \) with homogeneous coordinates given by the composite ratio

\[
\pi(\mathbf{C}, \mathbf{W}) = \langle c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{33}, w_1, w_2, w_3 \rangle.
\]

Clearly, \( \pi(\lambda\mathbf{C}, \lambda\mathbf{W}) = \pi(\mathbf{C}, \mathbf{W}) \) for any nonzero scalar \( \lambda \). Thus knowing \( \pi(\mathbf{C}, \mathbf{W}) \) amounts to knowing \( \mathbf{C} \) and \( \mathbf{W} \) within a common scalar factor.

Differential epipolar equation (20) forms the basis for our method of self-calibration. We use this equation to determine \( \pi(\mathbf{C}, \mathbf{W}) \) from the optical flow. Knowing \( \pi(\mathbf{C}, \mathbf{W}) \) will in turn allow recovery of some of the parameters that describe the egomotion and internal geometry of the camera, henceforth termed the key parameters.

Finding \( \pi(\mathbf{C}, \mathbf{W}) \) from the optical flow is in theory straightforward. If, at any given instant \( t \), we supply
sufficiently many independent vectors \([\mathbf{m}_i(t)^T, \mathbf{m}_i(t)^T]^T\), then \(C(t)\) and \(W(t)\) can be determined, up to a common scalar factor, from the following system of equations:

\[
\mathbf{m}_i(t)^T W(t) \mathbf{m}_i(t) + \mathbf{m}_i(t)^T C(t) \mathbf{m}_i(t) = 0.
\]  

(21)

Note that each of these equations is linear in the entries of \(C(t)\) and \(W(t)\). Therefore solving Eqs. (21) reduces to finding the null space of a matrix, and this problem can be tackled, for example, by employing the method of singular-value decomposition.

The extraction of key parameters from \(\pi(C, W)\) is discussed in Section 5. We close the present section by showing that \(\pi(C, W)\) lies upon a hypersurface of \(\mathbb{P}^8\), a seven-dimensional manifold. Indeed, from Eqs. (18) and (19) we have

\[
C = \frac{1}{2} (WA^{-1} (\hat{\mathbf{c}} + \mathbf{B}) A + A^T (\hat{\mathbf{c}} - \mathbf{B}^T) (A^T)^{-1} W).
\]

(22)

Since \(w^T W = 0\) and \(W w = 0\), it follows from Eq. (22) that

\[
w^T C w = 0.
\]

The left-hand side of this equation is a homogeneous polynomial of degree 3 in the entries of \(C\) and \(W\), so the equation defines a hypersurface in \(\mathbb{P}^8\). Clearly, \(\pi(C, W)\) is a member of this hypersurface. Thus \(\pi(C, W)\) is not an arbitrary point in \(\mathbb{P}^8\) but is constrained to a seven-dimensional submanifold of \(\mathbb{P}^8\), a fact already noted in Ref. 5.

5. SELF-CALIBRATION WITH FREE FOCAL LENGTH

Of the five parameters, six describe the egomotion of the camera and the rest describe the internal geometry of the camera. Only five egomotion parameters can, however, be determined from image data, as one parameter is lost because of scale indeterminacy. Given that \(\pi(C, W)\) is a member of a seven-dimensional hypersurface in \(\mathbb{P}^8\), the total number of key parameters that can be recovered by exploiting \(\pi(C, W)\) does not exceed seven. If we want to recover all five computable egomotion parameters, we have to accept that not all intrinsic parameters can be retrieved. Accordingly, we have to adopt a particular form of \(A\), deciding which intrinsic parameters will be known and which will be unknown, as well as which will be fixed and which will be free. We define a free parameter to be one that can vary continuously with time.

Assume that the focal length is unknown and free, that pixels are square with unit length (in length units of \(\Gamma_c\)), and that the principal point is fixed and known. In this situation, for each time instant \(t\), \(A(t)\) is given by

\[
A(t) = \begin{bmatrix}
1 & 0 & -i_1 \\
0 & 1 & -i_2 \\
0 & 0 & -f(t)
\end{bmatrix},
\]

(23)

where \(i_1\) and \(i_2\) are the coordinates of the known principal point and \(f(t)\) is the unknown focal length at time \(t\). From now on we shall omit in notation the dependence on time. Let \(\pi(v)\) be the projective form of \(v\), that is, the point in the two-dimensional real projective space \(\mathbb{P}^2\) with homogeneous coordinates given by the composite ratio

\[
\pi(v) = (v_1 : v_2 : v_3).
\]

As is clear, \(\pi(v)\) captures the direction of \(v\). It emerges that, with the adoption of the above form of \(A\), one can conduct self-calibration by explicitly expressing the entities \(\omega, \pi(v), f\), and \(f\) in terms of \(\pi(C, W)\). Of these entities, \(\omega\) and \(\pi(v)\) account for five egomotion parameters [\(\omega\) accounts for three parameters and \(\pi(v)\) accounts for two parameters] and \(f\) and \(f\) account for two intrinsic parameters. Note that \(v\) is not wholly recoverable, as the length of \(v\) is indeterminate. Retrieving \(\omega, \pi(v), f\), and \(f\) from \(\pi(C, W)\) has as its counterpart in stereo vision Hartley's10 procedure to determine five relative orientation parameters and two focal lengths from a fundamental matrix whose intrinsic-parameter parts have a form analogous to that given in Eq. (25) (with \(i_1\) and \(i_2\) known).

We now describe the self-calibration procedure in detail. First we make a reduction to the case \(i_1 = i_2 = 0\). Represent \(A\) as

\[
A = A_1 A_2,
\]

where

\[
A_1 = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & f
\end{bmatrix}, \quad A_2 = \begin{bmatrix}
1 & 0 & -i_1 \\
0 & 1 & -i_2 \\
0 & 0 & 1
\end{bmatrix}.
\]

Let

\[
C_1 = (A_2^{-1})^T C A_2^{-1}, \quad W_1 = (A_2^{-1})^T W A_2^{-1}.
\]

Letting \(B_1\) be the matrix function obtained from Eq. (14) by substituting \(A_1\) for \(A\), and taking into account that \(\dot{A}_2 = 0\), we find that

\[
B = \dot{A} A^{-1} = \dot{A}_1 A_2 (A_1 A_2)^{-1} = B_1.
\]

With this identity, it is easy to verify that \(C_1\) and \(W_1\) satisfy Eqs. (18) and (19), respectively, provided that \(A\) and \(B\) in these equations are replaced by \(A_1\) and \(B_1\). Therefore, passing to \(A_1, C_1,\) and \(W_1\) in lieu of \(A, C,\) and \(W\), respectively, we may assume that \(i_1 = i_2 = 0\).

Henceforth we shall assume that such an initial reduction has been made, letting \(A, C,\) and \(W\) be equal to \(A_1, C_1,\) and \(W_1\), respectively. Let \(S\) be the matrix defined as

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

A straightforward calculation shows that

\[
S = \begin{bmatrix}
0 & \omega_3 & f \omega_2 \\
-\omega_3 & 0 & -f \omega_1 \\
-\omega_2 / f & \omega_1 / f & \dot{f} / f
\end{bmatrix}.
\]

(24)

With the use of \(S\), Eq. (22) can be rewritten as

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

(25)

Regarding \(C\) and \(W\) as being known and \(S\) as being unknown, and taking into account that \(C\) a \(3 \times 3\) symmetric matrix—has only six independent entries, we can view the above matrix equation as a system of six inhomogeneous linear equations in the entries of \(S\). Of these only five equations are independent, as \(C\) and \(W\) are interre-
lated. Solving for the entries of $S$ and using on the way the explicit form of $S$ given by Eq. (24), we can express—as we shall see shortly—$\omega$, $f$, and $\tilde{f}$ in terms of $\pi(C, W)$. Once $f$ and hence $A$ is represented as a function of $\pi(C, W)$, $\tilde{v}$ can be found from

$$\tilde{v} = (A^T)^{-1}WA^{-1}, \quad (26)$$

which immediately follows from Eq. (19). Note that $W$ is known only up to a scalar factor, so $\tilde{v}$ (and hence $v$) cannot be fully determined. However, as $W$ depends linearly on $\pi(C, W)$, it is clear that $\pi(v)$ can be regarded as being a function of $\pi(C, W)$. In this way, all the parameters $\omega, \pi(v), f$, and $\tilde{f}$ are determined from $\pi(C, W)$.

We now give explicit formulas for $\omega, \pi(v), f$, and $\tilde{f}$. Set

$$\delta_1 = \frac{\omega_1}{f}, \quad \delta_2 = \frac{\omega_2}{f}, \quad \delta_3 = \omega_3,$$

$$\delta_4 = f^2, \quad \delta_5 = \frac{\tilde{f}}{f}. \quad (27)$$

In view of Eqs. (24) and (25), we have

$$c_{11} = -w_2 \delta_2 + w_3 \delta_3,$$
$$2c_{12} = w_2 \delta_1 + w_1 \delta_2,$$
$$c_{22} = w_1 \delta_1 + w_3 \delta_3.$$

Hence

$$\delta_1 = \frac{2c_{12}w_2 - (c_{22} - c_{11})w_1}{w_1^2 + w_2^2},$$
$$\delta_2 = \frac{2c_{12}w_1 + (c_{22} - c_{11})w_2}{w_1^2 + w_2^2},$$
$$\delta_3 = \frac{c_{11}w_1^2 + 2c_{12}w_1w_2 + c_{22}w_2^2}{w_3(w_1^2 + w_2^2)}.$$ 

(28)

The expressions on the right-hand sides of Eqs. (28) are homogeneous of degree 0 in the entries of $C$ and $W$, that is, they do not change if $C$ and $W$ are multiplied by a common scalar factor. Therefore Eqs. (28) can be regarded as formulas for $\delta_1, \delta_2,$ and $\delta_3$ in terms of $\pi(C, W)$. Assuming—as we now may—that $\delta_1, \delta_2, \delta_3$ are known, we again use Eqs. (24) and (25) to derive the following formulas for $\delta_4$ and $\delta_5$:

$$2c_{13} = w_3 \delta_1 \delta_4 + w_2 \delta_3 - w_1 \delta_5,$$
$$2c_{23} = w_3 \delta_2 \delta_4 - w_1 \delta_5 - w_2 \delta_3,$$
$$c_{33} = -(w_1 \delta_1 + w_2 \delta_2) \delta_4.$$ 

(29)

These three equations in $\delta_4$ and $\delta_5$ are not linearly independent. To determine $\delta_4$ and $\delta_5$ in an efficient way, we proceed as follows. Let $\delta = [\delta_1, \delta_2, \delta_3]^T$ be such that

$$d_1 = 2c_{13} + w_1 \delta_3, \quad d_2 = 2c_{23} + w_2 \delta_3, \quad d_3 = c_{33},$$

and let

$$D = \begin{bmatrix} w_3 \delta_1 & w_2 \\ w_1 \delta_2 & -w_1 \\ -w_1 \delta_1 - w_2 \delta_2 & 0 \end{bmatrix}. $$

With this notation, Eqs. (29) can be rewritten as

$$D \delta = d,$$

whence

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

More explicitly, we have the following formulas:

$$\delta_4 = \frac{1}{\Gamma} [(w_1 w_2 \delta_1 + (w_2^2 + w_3^2) \delta_2) d_1 + [(w_1^2 + w_3^2) \delta_1 + w_1 w_2 \delta_3] d_2 + (w_2 w_3 \delta_1 - w_1 w_2 \delta_3) d_3],$$

$$\delta_5 = \frac{1}{\Gamma} [(w_1^2 + w_3^2) \delta_1 + w_1 w_2 \delta_3] d_2 + (w_2 w_3 \delta_1 - w_1 w_2 \delta_3) d_3, \quad (30)$$

where $\Gamma = (w_1^2 + w_2^2 + w_3^2)(w_1 \delta_1 + w_2 \delta_2).$ Again the expressions on the right-hand sides are homogeneous of degree 0 in the entries of $C$ and $W$, so Eq. (30) can be regarded as formulas for $\delta_4$ and $\delta_5$ in terms of $\pi(C, W)$. Combining Eqs. (27), (28), and (30), we obtain

$$\omega_1 = \delta_1 \sqrt{\delta_4}, \quad \omega_2 = \delta_2 \sqrt{\delta_4}, \quad \omega_3 = \delta_3,$$
$$f = \sqrt{\delta_4}, \quad \tilde{f} = \delta_5 \sqrt{\delta_4}. $$

Rewriting Eq. (26) as

$$v_1 = \frac{w_1}{f}, \quad v_2 = \frac{w_2}{f}, \quad v_3 = w_3, \quad (31)$$

and taking into account that $\tilde{f}$ has already been specified, we find that

$$\pi(v) = (-w_1: -w_2: w_3). $$

In this way, all the parameters $\omega, \pi(v), f$, and $\tilde{f}$ are determined from $\pi(C, W)$.

Note that, for the above self-calibration procedure to work, a number of conditions must be met. Inspecting Eqs. (28), we see the need to assume that $v_3 \neq 0$ and also that either $v_1 \neq 0$ or $v_2 \neq 0$. In particular, $v$ has to be nonzero. Furthermore, $\Gamma$ in Eqs. (30) also has to be nonzero. With the assumption that $\pi \neq 0$ in place, we have that $\Gamma \neq 0$ if and only if $w_1 \delta_1 + w_2 \delta_2 \neq 0$. Taking into account the first two equations of Eqs. (27) and the first two equations of Eqs. (31), we see that the latter condition is equivalent to $v_1 \omega_1 + v_2 \omega_2 \neq 0$. Altogether we have then to assume that $v_3 \neq 0$, that either $v_1 \neq 0$ or $v_2 \neq 0$, and, furthermore, that $v_1 \omega_1 + v_2 \omega_2 \neq 0$.

6. SCENE RECONSTRUCTION

Here we tackle the problem of scene reconstruction. It is shown that if the camera's intrinsic-parameter matrix matrix assumes the form given in Section 5, knowledge of the entities $\omega, \pi(v), f$, and $\tilde{f}$ allows scene structure to be computed, up to scale, from instantaneous optical flow.

We adopt the form of $A$ given in Eq. (23). Assuming that $\omega, \pi(v), f$, and $\tilde{f}$ are known, we solve for $[\mathbf{x}^T, \mathbf{x}^T]^T$ given $[\mathbf{m}^T, \mathbf{m}^T]^T$. Note that, of the entities $\mathbf{x}$ and $\tilde{\mathbf{x}}$, only $\mathbf{x}$ is needed for scene reconstruction.
First, using Eqs. (12) and (13), we determine the values of $p$ and $\dot{p}$. Next, substituting Eqs. (9) and (10) into Eq. (7), we find that

$$x_3 [\dot{p} - f' (\dot{p} + \omega \dot{p})] - x_3 \dot{p} + f^2 \nu = 0. \quad (32)$$

Clearly, $\dot{p} - f' (\dot{p} + \omega \dot{p})$ and $f'/p$ are known, $\nu$ is partially known [namely, $\pi(v)$ is known], and $x_3$ and $\dot{x}_3$ are unknown. Assume temporarily that $\nu$ is known. Then Eq. (32) can immediately be employed to find $x_3$ and $\dot{x}_3$. Indeed, bearing in mind that $\dot{p} - f' (\dot{p} + \omega \dot{p})$, $f'/p$ and $f^2 \nu$ are column vectors with three entries, one can regard Eq. (32) as being a system of three linear equations (algebraic, not differential) in $x_3$ and $\dot{x}_3$, and this system can easily be solved for the two unknowns. On finding $x_3$ and $\dot{x}_3$, we use Eqs. (9) and (10) to determine $x$ and $\dot{x}$. With $x$ thus specified, scene reconstruction is complete.

Note that this method breaks down when $\dot{p} - f' (\dot{p} + \omega \dot{p})$ and $f'/p$ are linearly dependent, or equivalently if

$$\dot{p} (\dot{p} + \omega \dot{p}) = 0.$$

In view of Eqs. (9) and (10), if $x_3 = 0$ then this equation is equivalent to

$$\dot{x} (\dot{x} + \omega \dot{x}) = 0,$$

which, by Eq. (7), is equivalent to $\dot{x} \nu = 0$. We need therefore to assume that $\dot{x} \nu \neq 0$, or equivalently that $x$ and $\nu$ are linearly independent, whenever $x_3 \neq 0$. In particular, this means that $\nu \neq 0$.

We are left with the task of determining $\nu$. Fix $|\nu|$ arbitrarily as a positive value. In view of $\nu \neq 0$, one of the components of $\nu$, say $v_3$, is nonzero. Since

$$\frac{(\text{sgn} v_3)}{|\nu|} \nu = \left[ \frac{v_1^2}{v_3} + \frac{v_1^2}{v_3} + 1 \right]^{-1/2} \left[ \frac{v_1}{v_3}, \frac{v_1}{v_3}, 1 \right]^T,$$

where $\text{sgn} v_3$ denotes the sign of $v_3$ and $|\nu| = (w_1^2 + w_2^2 + w_3^2)^{1/2}$ and since the right-hand side is expressible in terms of $\pi(v)$, one can regard $(\text{sgn} v_3)|\nu|/|\nu|$ as being known. With the assumed value of $|\nu|$, we see that $\nu$ is determined up to a sign. The sign is $a \text{ priori}$ unknown because $v_3$ is unknown. However, we can determine it uniquely by requiring that all the $x_3$ calculated by solving Eq. (32) be nonnegative. This requirement simply reflects the fact that the scene is in front of the camera.

7. EXPERIMENTAL RESULTS

To assess the applicability and correctness of the approach, we performed a simple test with real-world imagery. The three images shown in Fig. 2 were captured by a Phillips CCD camera with a 12.5-mm lens. Corners were localized to subpixel accuracy with the use of a corner detector, correspondences between the images were obtained, and the optical flow depicted in Fig. 3 was computed by exploiting these correspondences (no intensity-based method was used in the process). A straightforward singular-value-decomposition method was used to determine the corresponding ratio $\pi(C, W)$ from the optical flow. Closed-form expressions described above were employed to self-calibrate the system. With the seven key parameters recovered, the reconstruction displayed in Fig. 4 was finally obtained. Note that reconstructed
points in 3-space have been connected by line segments to convey clearly the patterns of the calibrated grid. This simple reconstruction is visually pleasing and suggests that the approach holds promise.

8. CONCLUSION

The primary aim in this work has been to elucidate the means by which a moving camera can be self-calibrated from instantaneous optical flow. Our approach was to model the way in which optical flow is induced when a freely moving camera views a static scene and then to derive a differential epipolar equation incorporating two critical matrices. We noted that these matrices are retrievable, up to a scalar factor, directly from the optical flow. Adoption of a specific camera model, in which the focal length and its derivative are the sole unknown intrinsic, permitted the specification of closed-form expressions (in terms of the composite ratio of some entries of the two matrices) for the five computable egomotion parameters and the two unknown intrinsic parameters. A procedure was also given for reconstructing a scene from the optical flow and the results of self-calibration. The self-calibration and reconstruction procedures were implemented and tested on an optical flow field derived from a real-image sequence of a calibration grid. The ensuing three-dimensional reconstruction of the grid squares was visually pleasing, confirming the validity of the theory and suggesting that the approach holds promise.

ACKNOWLEDGMENTS

This research was in part supported by a grant from the Australian Research Council and by a grant issued by the Dirección General de Investigación Científica y Técnica of Spain. The authors are grateful for the comments of Phil Torr, Andrew Zisserman, Lourdes de Agapito, Zhengyou Zhang, Kenichi Kanatani, and Dario Maravall. Anton van den Hengel provided valuable advice and some implementation support.

The author's e-mail addresses are mjb@cs.adelaide.edu.au, wojtek@cs.adelaide.edu.au, and lbaumela@fl.upm.es.

REFERENCES