DETERMINING THE EGO-MOTION OF AN UNCALIBRATED CAMERA FROM INSTANTANEOUS OPTICAL FLOW

MICHAEL J. BROOKS, WOJCIECH CHOJNACKI, AND LUIS BAUMELA

ABSTRACT. The main result of this paper is a procedure for self-calibration of a moving camera from instantaneous optical flow. Under certain assumptions, this procedure allows the ego-motion 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 upon the use of a differential epipolar equation that relates optical flow to the ego-motion and internal geometry of the camera. The paper presents a detailed derivation of this equation. This aspect of the work may be seen as a recasting into an analytical framework of the pivotal research of Viéville and Faugeras.[1] The information about the camera's ego-motion 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 via closed-form expressions in the composite ratio. The corresponding formulae specify five ego-motion parameters, as well as the focal length and its derivative. An accompanying procedure is presented for reconstructing the viewed scene, up to scale, from the derived self-calibration data and the optical flow data. Experimental results are given to demonstrate the correctness of the approach.

1. Introduction

Of considerable interest in recent years has been to generate 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.[2,3]

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 ego-motion and intrinsic parameters of the camera, and is analogous to self-calibration of a stereo vision set-up using corresponding points.

The proposed method of self-calibration rests on a *differential epipolar equation* that relates optical flow to the ego-motion 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*,[4,5] the differential epipolar equation

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

incorporates two matrices. These matrices encode the information about the ego-motion and 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 via 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, only compute relative motion, our technique applies most generally to a moving camera viewing a moving rigid body.) Suppose the interior characteristics of the camera are known, except for the focal length, and that the focal length is free in that it may vary continuously. We show in this work that, from instantaneous optical flow, we may compute via 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, that describe the camera's ego-motion, accounting for five degrees of freedom. Note that a full description of the ego-motion 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 has as its analogue in stereo vision the indeterminacy of baseline length from corresponding points.)

Our work is inspired by, and closely related to, that of Viéville and Faugeras.[1] These authors were the first to introduce an equation akin to what we term here the differential epipolar equation. However, unlike the latter, the equation from Viéville and Faugeras' work takes the form of an approximation and not 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 ego-motion of a calibrated camera, see for example Refs. 6–9.

2. Scene motion in the camera frame

In order to extract 3D information from an image, a camera model must be adopted. In this paper the camera is modeled as a pinhole (see Figure 1). A detailed exposition of the pinhole model including the relevant terminology can be found in Ref. 10, Section 3. 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 Γ_{w} whose scene configuration will be fixed throughout. Associate with the camera an independent Cartesian coordinate frame Γ_{c}, with origin C and basis $\{\boldsymbol{e}_i\}_{1\leq i\leq 3}$ of unit orthogonal vectors, so that C coincides with the optical centre, $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ span the focal plane, and $\boldsymbol{e}_3$ determines the optical axis (see Figure 1 for a display of the coordinate frames). Ensure that Γ_{c} and Γ_{w} are equi-oriented by swapping two arbitrarily chosen basis vectors of Γ_{w} if initially the frames are counter-oriented. In so doing, it will be guaranteed that the value of the cross product of two vectors is independent of whether the basis of unit orthogonal vectors associated with Γ_{w} or that associated with Γ_{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 Γ_{w}.

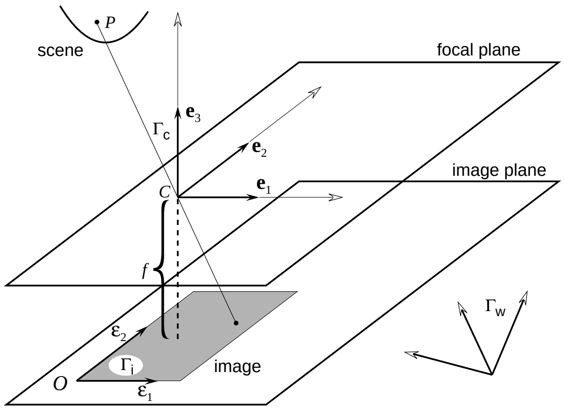

FIGURE 1. Image formation and coordinate frames.

Similarly, for each $i \in \{1,2,3\}$, $\boldsymbol{e}_i$ will be identified with the point in $\mathbb{R}^3$ formed by the components of $\boldsymbol{e}_i$ relative to the vector basis of Γ_{w}.

Suppose that the camera undergoes smooth motion with respect to Γ_{w}. At each time instant t, the location of the camera relative to Γ_{w} is given by $(C(t), \boldsymbol{e}_1(t), \boldsymbol{e}_2(t), \boldsymbol{e}_3(t)) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$. The motion of the camera is then described by the differentiable function $t \mapsto (C(t), \boldsymbol{e}_1(t), \boldsymbol{e}_2(t), \boldsymbol{e}_3(t))$. The derivative $\dot{C}(t)$ captures the instantaneous *translational velocity* of the camera relative to Γ_{w} at t. Expanding this derivative with respect to the basis $\{\boldsymbol{e}_i(t)\}_{1 \le i \le 3}$

$$\dot{C}(t) = \sum_i v_i(t)\boldsymbol{e}_i(t) \tag{1}$$

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

$$\dot{\boldsymbol{e}}_i(t) = \sum_j \omega_{ji}(t)\boldsymbol{e}_j(t). \tag{2}$$

The coefficients thus arising can be arranged in the matrix

$$\boldsymbol{\Omega}(t) = [\omega_{ij}(t)]_{1 \le i,j \le 3}.$$

Leaving the dependency of the $\boldsymbol{e}_i$ upon t implicit, we can express the orthogonality and normalisation conditions satisfied by the $\boldsymbol{e}_i$ as

$$\boldsymbol{e}_i{}^T \boldsymbol{e}_j = \delta_{ij}, \tag{3}$$

where

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

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

$$\dot{\boldsymbol{e}}_i{}^T \boldsymbol{e}_j + \boldsymbol{e}_i{}^T \dot{\boldsymbol{e}}_j = 0.$$

In view of (2),

$$\omega_{ji} = \boldsymbol{e}_j{}^T \dot{\boldsymbol{e}}_i,$$

which together with the previous equation yields

$$\omega_{ij} = -\omega_{ji}.$$

We see then that $\boldsymbol{\Omega}$ is antisymmetric and as such can be represented as

$$\boldsymbol{\Omega} = \widehat{\boldsymbol{\omega}} \tag{4}$$

for some vector $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$, where $\widehat{\boldsymbol{\omega}}$ is defined as

$$\widehat{\boldsymbol{\omega}} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \tag{5}$$

Writing (2) as

$$\dot{\boldsymbol{e}}_1 = \omega_3 \boldsymbol{e}_2 - \omega_2 \boldsymbol{e}_3,$$

$$\dot{\boldsymbol{e}}_2 = \omega_1 \boldsymbol{e}_3 - \omega_3 \boldsymbol{e}_1,$$

$$\dot{\boldsymbol{e}}_3 = \omega_2 \boldsymbol{e}_1 - \omega_1 \boldsymbol{e}_2,$$

introducing

$$\boldsymbol{\eta} = \omega_1 \boldsymbol{e}_1 + \omega_2 \boldsymbol{e}_2 + \omega_3 \boldsymbol{e}_3,$$

and noting that, for any $\boldsymbol{z} = z_1 \boldsymbol{e}_1 + z_2 \boldsymbol{e}_2 + z_3 \boldsymbol{e}_3$,

$$\boldsymbol{\eta} \times \boldsymbol{z} = (\omega_2 z_3 - \omega_3 z_2)\boldsymbol{e}_1 + (\omega_3 z_1 - \omega_1 z_3)\boldsymbol{e}_2 + (\omega_1 z_2 - \omega_2 z_1)\boldsymbol{e}_3,$$

we have, for each $i \in \{1, 2, 3\}$,

$$\dot{\boldsymbol{e}}_i = \boldsymbol{\eta} \times \boldsymbol{e}_i.$$

It is clear from this system of equations that $\boldsymbol{\eta}$ represents the instantaneous *angular velocity* of the camera relative to Γ_{w}. The direction of $\boldsymbol{\eta}$ determines the axis of the instantaneous rotation of the camera, passing through C, relative to Γ_{w}. Correspondingly, $\boldsymbol{\omega}$ represents instantaneous angular velocity of the camera relative to Γ_{c}, and relative to Γ_{c} the axis of the instantaneous rotation of the camera is aligned along $\boldsymbol{\omega}$.

Let P be a point in space. Identify P with the point in $\mathbb{R}^3$ formed by the coordinates of P relative to Γ_{w}. With the earlier identification of C and the $\boldsymbol{e}_i$ with respective points of $\mathbb{R}^3$ still in force, the location of P relative to Γ_{c} can be expressed in terms of a coordinate vector $\boldsymbol{x} = [x_1, x_2, x_3]^T$ determined from the equation

$$P = \sum_i x_i \boldsymbol{e}_i + C. \tag{6}$$

This equation 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 Γ_{c}. Suppose that P is static with respect to Γ_{w}. As the camera moves, the position of P relative to Γ_{c} will change accordingly and will be recorded in the function $t \mapsto \boldsymbol{x}(t)$. This function satisfies an equation reflecting the kinematics of the moving camera. We derive this equation next.

Differentiating (6) and taking into account that $\dot{P} = \mathbf{0}$, we obtain

$$\sum_i (\dot{x}_i \boldsymbol{e}_i + x_i \dot{\boldsymbol{e}}_i) + \dot{C} = \mathbf{0}.$$

In view of (1) and (2),

$$\begin{aligned} \sum_i (\dot{x}_i \boldsymbol{e}_i + x_i \dot{\boldsymbol{e}}_i) + \dot{C} &= \sum_i \big(\dot{x}_i \boldsymbol{e}_i + x_i \sum_j \omega_{ji} \boldsymbol{e}_j + v_i \boldsymbol{e}_i\big) \\ &= \sum_i \big(\dot{x}_i + \sum_j \omega_{ij} x_j + v_i\big)\boldsymbol{e}_i. \end{aligned}$$

Therefore, for each $i \in \{1, 2, 3\}$,

$$\dot{x}_i + \sum_j \omega_{ij} x_j + v_i = 0,$$

or in matrix notation

$$\dot{\boldsymbol{x}} + \boldsymbol{\Omega} \boldsymbol{x} + \boldsymbol{v} = \boldsymbol{0}.$$

Coupling this with (4), we obtain

$$\dot{\boldsymbol{x}} + \widehat{\boldsymbol{\omega}} \boldsymbol{x} + \boldsymbol{v} = \boldsymbol{0}. \tag{7}$$

This is the equation governing the evolution of $\boldsymbol{x}$. Taking into account that $\widehat{\boldsymbol{\omega}} \boldsymbol{x} = \boldsymbol{\omega} \times \boldsymbol{x}$, it can also be stated in a more traditional form as

$$\dot{\boldsymbol{x}} + \boldsymbol{\omega} \times \boldsymbol{x} + \boldsymbol{v} = \boldsymbol{0}.$$

3. Differential epipolar equation

The camera image is formed via perspective projection of the viewed scene, through C, onto the plane parallel to the focal plane (again, see Figure 1). In coordinates relative to Γ_{c} the image plane is described by $\{\boldsymbol{x} \in \mathbb{R}^3 \colon x_3 = -f\}$, where f is the focal length. If P is a point in space, and if $\boldsymbol{x}$ and $\boldsymbol{p}$ are the coordinates relative to Γ_{c} of P and its image, then

$$\boldsymbol{p} = -f \frac{\boldsymbol{x}}{x_3}. \tag{8}$$

Suppose again that P is static and the camera moves with respect to Γ_{w}. The evolution of the image of P will then be described by the function $t \mapsto \boldsymbol{p}(t)$. This function is subject to a constraint deriving from equation (7). We proceed to determine this constraint.

First, note that (8) can be equivalently rewritten as

$$\boldsymbol{x} = -\frac{x_3 \boldsymbol{p}}{f}, \tag{9}$$

which immediately leads to

$$\dot{\boldsymbol{x}} = \frac{x_3 \dot{f} - \dot{x}_3 f}{f^2} \boldsymbol{p} - \frac{x_3}{f} \dot{\boldsymbol{p}}. \tag{10}$$

Next, applying the matrix $\widehat{\boldsymbol{v}}$ (formed according to the definition (5)) to both sides of (7) and noting that $\widehat{\boldsymbol{v}} \boldsymbol{v} = \boldsymbol{0}$, we get

$$\widehat{\boldsymbol{v}} \dot{\boldsymbol{x}} + \widehat{\boldsymbol{v}} \widehat{\boldsymbol{\omega}} \boldsymbol{x} = \boldsymbol{0}.$$

Now, in view of (9) and (10),

$$\frac{x_3 \dot{f} - \dot{x}_3 f}{f^2} \widehat{\boldsymbol{v}} \boldsymbol{p} - \frac{x_3}{f} \widehat{\boldsymbol{v}} \dot{\boldsymbol{p}} - \frac{x_3}{f} \widehat{\boldsymbol{v}} \widehat{\boldsymbol{\omega}} \boldsymbol{p} = \boldsymbol{0}.$$

Applying $\boldsymbol{p}^T$ to both sides of this equation, dropping the summand with $\boldsymbol{p}^T \widehat{\boldsymbol{v}} \boldsymbol{p}$ in the left-hand side (in view of the antisymmetry of $\widehat{\boldsymbol{v}}$, we have $\boldsymbol{p}^T \widehat{\boldsymbol{v}} \boldsymbol{p} = 0$), and cancelling out the common factor $-x_3/f$ in the remaining summands, we obtain

$$\boldsymbol{p}^T \widehat{\boldsymbol{v}} \dot{\boldsymbol{p}} + \boldsymbol{p}^T \widehat{\boldsymbol{v}} \widehat{\boldsymbol{\omega}} \boldsymbol{p} = 0. \tag{11}$$

This is the sought-after constraint. We call it *the differential epipolar equation.* This term reflects the fact that equation (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 equations, referring the reader to Ref. 11 and its short version Ref. 12, where an analogue of (11), namely equation (20) presented 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 not involving any special differentiation procedure) is presented in the context of images formed on a sphere. It is due to a suggestion of Torr[13] that in this work 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 $t \mapsto \boldsymbol{p}(t)$. As shown by Åström and Heyden,[14] for every $n \geq 2$, such functions satisfy an nth order differential equation that reduces to the differential epipolar equation when $n = 2$. The nth equation in the series is the infinitesimal version of the analogue 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 Γ_{i}, with origin O and basis of vectors $\{\boldsymbol{\epsilon}_i\}_{1\leq i\leq 2}$, in the image plane. It is natural to align the $\boldsymbol{\epsilon}_i$ 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, Γ_{i} and Γ_{c} are customarily adjusted so that $\boldsymbol{\epsilon}_i = s_i\boldsymbol{e}_i$, where s_i characterises the pixel size in the direction of $\boldsymbol{\epsilon}_i$ in length units of Γ_{c}. Suppose that a point in the image plane has coordinates $\boldsymbol{p} = [p_1, p_2, -f]^T$ and $[m_1, m_2]^T$ relative to Γ_{c} and Γ_{i}, respectively. If $[m_1, m_2]^T$ is appended by an extra entry equal to 1 to yield the vector $\boldsymbol{m} = [m_1, m_2, 1]^T$, then the relation between $\boldsymbol{p}$ and $\boldsymbol{m}$ can be conveniently written as

$$\boldsymbol{p} = \boldsymbol{A}\boldsymbol{m}, \tag{12}$$

where $\boldsymbol{A}$ is a 3×3 invertible matrix called the *intrinsic-parameter matrix.* With the assumption $\boldsymbol{\epsilon}_i = s_i\boldsymbol{e}_i$ in force, if $[i_1, i_2]^T$ is the Γ_{i}-based coordinate representation of the principal point (that is the point at which the optical axis intersects the image plane), then $\boldsymbol{A}$ takes the form

$$\boldsymbol{A} = \begin{bmatrix} s_1 & 0 & -s_1 i_1 \\ 0 & s_2 & -s_1 i_2 \\ 0 & 0 & -f \end{bmatrix}.$$

When pixels are non-rectangular, $\boldsymbol{A}$ takes a more complicated form accounting for one more parameter that encodes shear in the camera axes (see Ref. 10, Section 3).

The differential epipolar equation (11) can be restated so as to use, for any given instant, the Γ_{i}-based vector $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$ in place of the Γ_{c}-based vector $[\boldsymbol{p}^T, \dot{\boldsymbol{p}}^T]^T$. The time-labeled set of all vectors of the form $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$, describing the position and velocity of the images of various elements of the scene, constitutes the true image motion field which, as is usual, we assume to correspond to the observed image velocity field or *optical flow* (see Ref. 15, Chapter 12).

It follows from (12) that

$$\dot{\boldsymbol{p}} = \dot{\boldsymbol{A}}\boldsymbol{m} + \boldsymbol{A}\dot{\boldsymbol{m}}. \tag{13}$$

This equation in conjunction with (12) implies that

$$\boldsymbol{p}^T\widehat{\boldsymbol{v}}\dot{\boldsymbol{p}} = \boldsymbol{m}^T\boldsymbol{A}^T\widehat{\boldsymbol{v}}\dot{\boldsymbol{A}}\boldsymbol{m} + \boldsymbol{m}^T\boldsymbol{A}\widehat{\boldsymbol{v}}\boldsymbol{A}\dot{\boldsymbol{m}},$$
$$\boldsymbol{p}^T\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}}\boldsymbol{p} = \boldsymbol{m}^T\boldsymbol{A}^T\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}}\boldsymbol{A}\boldsymbol{m},$$

and so (11) can be rewritten as

$$\boldsymbol{m}^T\boldsymbol{A}\widehat{\boldsymbol{v}}\boldsymbol{A}\dot{\boldsymbol{m}} + \boldsymbol{m}^T\big(\boldsymbol{A}^T\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}}\boldsymbol{A} + \boldsymbol{A}^T\widehat{\boldsymbol{v}}\dot{\boldsymbol{A}}\big)\boldsymbol{m} = 0.$$

Letting

$$\boldsymbol{B} = \dot{\boldsymbol{A}}\boldsymbol{A}^{-1}, \tag{14}$$

we have

$$\boldsymbol{m}^T \boldsymbol{A}^T \widehat{\boldsymbol{v}} \boldsymbol{A} \dot{\boldsymbol{m}} + \boldsymbol{m}^T \boldsymbol{A}^T \widehat{\boldsymbol{v}} (\widehat{\boldsymbol{\omega}} + \boldsymbol{B}) \boldsymbol{A} \boldsymbol{m} = 0. \tag{15}$$

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

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

Evidently

$$\boldsymbol{m}^T \boldsymbol{X}_{\text{sym}} \boldsymbol{m} = \boldsymbol{m}^T \boldsymbol{X} \boldsymbol{m}, \tag{16a}$$

$$\boldsymbol{m}^T \boldsymbol{X}_{\text{asym}} \boldsymbol{m} = 0. \tag{16b}$$

Since $\widehat{\boldsymbol{\omega}}$ and $\widehat{\boldsymbol{v}}$ are antisymmetric, we have

$$(\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}})_{\text{sym}} = \frac{1}{2}(\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{v}}), \qquad (\widehat{\boldsymbol{v}}\boldsymbol{B})_{\text{sym}} = \frac{1}{2}(\widehat{\boldsymbol{v}}\boldsymbol{B} - \boldsymbol{B}^T\widehat{\boldsymbol{v}}). \tag{17}$$

Denote by $\boldsymbol{C}$ the symmetric part of $\boldsymbol{A}^T \widehat{\boldsymbol{v}} (\widehat{\boldsymbol{\omega}} + \boldsymbol{B}) \boldsymbol{A}$. In view of (17), we have

$$\boldsymbol{C} = \frac{1}{2} \boldsymbol{A}^T (\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{v}} + \widehat{\boldsymbol{v}}\boldsymbol{B} - \boldsymbol{B}^T\widehat{\boldsymbol{v}}) \boldsymbol{A}. \tag{18}$$

Let

$$\boldsymbol{W} = \boldsymbol{A}^T \widehat{\boldsymbol{v}} \boldsymbol{A}. \tag{19}$$

On account of (15), (16a) and (18), we can write

$$\boldsymbol{m}^T \boldsymbol{W} \dot{\boldsymbol{m}} + \boldsymbol{m}^T \boldsymbol{C} \boldsymbol{m} = 0. \tag{20}$$

This is the differential epipolar equation for optical flow. A constraint similar, termed *the first-order expansion of the fundamental motion equation*, is derived by quite different means by Viéville and Faugeras.[1] In contrast with the above, however, it takes the form of an approximation rather than a strict equality.

In view of (19) and the antisymmetry of $\widehat{\boldsymbol{v}}$, $\boldsymbol{W}$ is antisymmetric, and so $\boldsymbol{W} = \widehat{\boldsymbol{w}}$ for some vector $\boldsymbol{w} = [w_1, w_2, w_3]^T$. $\boldsymbol{C}$ is symmetric, and hence it is uniquely determined by the entries $c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{33}$. Let $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ be the *joint projective form* of $\boldsymbol{C}$ and $\boldsymbol{W}$, that is, the point in the 8-dimensional real projective space $\mathbb{P}^8$ with homogeneous coordinates given by the composite ratio

$$\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W}) = (c_{11} : c_{12} : c_{13} : c_{22} : c_{23} : c_{33} : w_1 : w_2 : w_3).$$

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

The differential epipolar equation (20) forms the basis for our method of self-calibration. We use this equation to determine $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ from the optical flow. Knowing $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ will in turn allow recovery of some of the parameters describing the ego-motion and internal geometry of the camera, henceforth termed the *key parameters*.

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

$$\boldsymbol{m}_i(t)^T \boldsymbol{W}(t) \dot{\boldsymbol{m}}_i(t) + \boldsymbol{m}_i(t)^T \boldsymbol{C}(t) \boldsymbol{m}_i(t) = 0. \tag{21}$$

Note that each of these equations is linear in the entries of $\boldsymbol{C}(t)$ and $\boldsymbol{W}(t)$. Therefore solving (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 $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ will be discussed in the next section. We close the present section by showing that $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ lies on a hypersurface of $\mathbb{P}^8$, a 7-dimensional manifold. Indeed, by (18) and (19), we have

$$\boldsymbol{C} = \frac{1}{2}\big[\boldsymbol{W}\boldsymbol{A}^{-1}(\widehat{\boldsymbol{\omega}} + \boldsymbol{B})\boldsymbol{A} + \boldsymbol{A}^T(\widehat{\boldsymbol{\omega}} - \boldsymbol{B}^T)(\boldsymbol{A}^T)^{-1}\boldsymbol{W}\big]. \tag{22}$$

Taking into account that $\boldsymbol{w}^T\boldsymbol{W} = \mathbf{0}$ and $\boldsymbol{W}\boldsymbol{w} = \mathbf{0}$, we see that

$$\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} = 0.$$

The left-hand side is a homogeneous polynomial of degree 3 in the entries of $\boldsymbol{C}$ and $\boldsymbol{W}$, and so the equation defines a hypersurface in $\mathbb{P}^8$. Clearly, $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is a member of this hypersurface. Thus $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is not an arbitrary point in $\mathbb{P}^8$ but is constrained to a 7-dimensional submanifold of $\mathbb{P}^8$, a fact already noted in Ref. 1.

5. Self-calibration with free focal length

Of the key parameters, 6 describe the ego-motion of the camera, and the rest describe the internal geometry of the camera. Only 5 ego-motion parameters can, however, be determined from image data, as one parameter is lost due to scale indeterminacy. Given that $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is a member of a 7-dimensional hypersurface in $\mathbb{P}^8$, the total number of key parameters that can be recovered by exploiting $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ does not exceed 7. If we want to recover all 5 computable ego-motion parameters, we have to accept that not all intrinsic parameters can be retrieved. Accordingly, we have to adopt a particular form of $\boldsymbol{A}$, deciding which intrinsic parameters will be known and which will be unknown, and also which will be fixed and which will be free. We define a *free* parameter to be one that may vary continuously with time.

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

$$\boldsymbol{A}(t) = \begin{bmatrix} 1 & 0 & -i_1 \\ 0 & 1 & -i_2 \\ 0 & 0 & -f(t) \end{bmatrix}, \tag{23}$$

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 upon time. Let $\boldsymbol{\pi}(\boldsymbol{v})$ be the *projective form* of $\boldsymbol{v}$, that is, the point in the 2-dimensional real projective space $\mathbb{P}^2$ with homogeneous coordinates given by the composite ratio

$$\boldsymbol{\pi}(\boldsymbol{v}) = (v_1 : v_2 : v_3).$$

As is clear, $\boldsymbol{\pi}(\boldsymbol{v})$ captures the direction of $\boldsymbol{v}$. It emerges that, with the adoption of the above form of $\boldsymbol{A}$, one can conduct self-calibration by explicitly expressing the entities $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ in terms of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$. Of these entities, $\boldsymbol{\omega}$ and $\boldsymbol{\pi}(\boldsymbol{v})$ account for 5 ego-motion parameters ($\boldsymbol{\omega}$ accounting for 3 parameters and $\boldsymbol{\pi}(\boldsymbol{v})$ accounting for 2 parameters), and f and $\dot{f}$ account for 2 intrinsic parameters. Note that $\boldsymbol{v}$ is not wholly recoverable, the length of $\boldsymbol{v}$ being indeterminate. Retrieving $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ from $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ has as its counterpart in stereo vision Hartley's[16] procedure to determine 5 relative orientation parameters and 2 focal lengths from a fundamental matrix whose intrinsic-parameter parts have a form analogous to that given in (23) (with i_1 and i_2 being known).

We now describe the self-calibration procedure in detail. We first make a reduction to the case $i_1 = i_2 = 0$. Represent $\boldsymbol{A}$ as

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

where

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -f \end{bmatrix}, \qquad \boldsymbol{A}_2 = \begin{bmatrix} 1 & 0 & -i_1 \\ 0 & 1 & -i_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let

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

Letting $\boldsymbol{B}_1$ be the matrix function obtained from (14) by substituting $\boldsymbol{A}_1$ for $\boldsymbol{A}$, and taking into account that $\dot{\boldsymbol{A}}_2 = \boldsymbol{0}$, we find that

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

Using this identity, it is easy to verify that $\boldsymbol{C}_1$ and $\boldsymbol{W}_1$ satisfy (18) and (19), respectively, provided $\boldsymbol{A}$ and $\boldsymbol{B}$ in these equations are replaced by $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$. Therefore, passing to $\boldsymbol{A}_1$, $\boldsymbol{C}_1$ and $\boldsymbol{W}_1$ in lieu of $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{W}$, respectively, we may assume that $i_1 = i_2 = 0$.

Henceforth we shall assume that such an initial reduction has been made, letting $\boldsymbol{A}$, $\boldsymbol{C}$ and $\boldsymbol{W}$ be equal to $\boldsymbol{A}_1$, $\boldsymbol{C}_1$ and $\boldsymbol{W}_1$, respectively. Let $\boldsymbol{S}$ be the matrix defined as

$$\boldsymbol{S} = \boldsymbol{A}^{-1}(\widehat{\boldsymbol{\omega}} + \boldsymbol{B})\boldsymbol{A}.$$

A straightforward calculation shows that

$$\boldsymbol{S} = \begin{bmatrix} 0 & -\omega_3 & -f\omega_2 \\ \omega_3 & 0 & f\omega_1 \\ \omega_2/f & -\omega_1/f & \dot{f}/f \end{bmatrix}. \tag{24}$$

With the use of $\boldsymbol{S}$, (22) can be rewritten as

$$\boldsymbol{C} = \frac{1}{2}(\boldsymbol{W}\boldsymbol{S} - \boldsymbol{S}^T\boldsymbol{W}). \tag{25}$$

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

$$\widehat{\boldsymbol{v}} = (\boldsymbol{A}^T)^{-1}\boldsymbol{W}\boldsymbol{A}^{-1}, \tag{26}$$

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

We now give explicit formulae for $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f, and $\dot{f}$. Set

$$\delta_1 = -\frac{\omega_1}{f}, \qquad \delta_2 = -\frac{\omega_2}{f}, \qquad \delta_3 = -\omega_3, \qquad \delta_4 = f^2, \qquad \delta_5 = \frac{\dot{f}}{f}. \tag{27}$$

In view of (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

$$\delta_1 = \frac{2c_{12}w_2 - (c_{22} - c_{11})w_1}{w_1^2 + w_2^2},$$
$$\delta_2 = \frac{2c_{12}w_1 + (c_{22} - c_{11})w_2}{w_1^2 + w_2^2}, \tag{28}$$
$$\delta_3 = \frac{c_{11}w_1^2 + 2c_{12}w_1w_2 + c_{22}w_2^2}{w_3(w_1^2 + w_2^2)}.$$

The expressions on the right-hand side are homogeneous of degree 0 in the entries of $\boldsymbol{C}$ and $\boldsymbol{W}$; that is, they do not change if $\boldsymbol{C}$ and $\boldsymbol{W}$ are multiplied by a common scalar factor. Therefore the above equations can be regarded as formulae for δ_1, δ_2, and δ_3 in terms of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$. Assuming—as we now may—that δ_1, δ_2, δ_3 are known, we again use (24) and (25) to derive the following formulae for δ_4 and δ_5:

$$2c_{13} = w_3\delta_1\delta_4 + w_2\delta_5 - w_1\delta_3,$$
$$2c_{23} = w_3\delta_2\delta_4 - w_1\delta_5 - w_2\delta_3, \tag{29}$$
$$c_{33} = -(w_1\delta_1 + w_2\delta_2)\delta_4.$$

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 $\boldsymbol{\delta} = [\delta_4, \delta_5]^T$, let $\boldsymbol{d} = [d_1, d_2, d_3]^T$ be such that

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

and let

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

With this notation, (29) can be rewritten as

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

whence

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

More explicitly, we have the following formulae:

$$\delta_4 = \frac{1}{\Gamma}\left(w_1w_3d_1 + w_2w_3d_2 - (w_1^2 + w_2^2)d_3\right),$$
$$\delta_5 = \frac{1}{\Gamma}\big((w_1w_2\delta_1 + (w_2^2 + w_3^2)\delta_2)d_1 - ((w_1^2 + w_3^2)\delta_1 + w_1w_2\delta_2)d_2 + (w_2w_3\delta_1 - w_1w_3\delta_2)d_3\big), \tag{30}$$

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 side are homogeneous of degree 0 in the entries of $\boldsymbol{C}$ and $\boldsymbol{W}$, and so the above equations can be regarded as formulae for δ_4 and δ_5 in terms of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

Combining (27), (28) and (30), we obtain

$$\omega_1 = -\delta_1\sqrt{\delta_4}, \qquad \omega_2 = -\delta_2\sqrt{\delta_4}, \qquad \omega_3 = -\delta_3, \qquad f = \sqrt{\delta_4}, \qquad \dot{f} = \delta_5\sqrt{\delta_4}.$$

Rewriting (26) as

$$v_1 = -\frac{w_1}{f}, \qquad v_2 = -\frac{w_2}{f}, \qquad v_3 = w_3, \tag{31}$$

and taking into account that f has already been specified, we find that

$$\boldsymbol{\pi}(\boldsymbol{v}) = (-w_1 : -w_2 : fw_3).$$

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

Note that, for the above self-calibration procedure to work, a number of conditions must be met. Inspecting (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, $\boldsymbol{v}$ has to be non-zero. Furthermore, Γ appearing in (30) also has to be non-zero. With the assumption $\boldsymbol{v} \neq \mathbf{0}$ 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 (27) and the first two equations of (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

The present section tackles the problem of scene reconstruction. It is shown that if the camera's intrinsic-parameter matrix assumes the form given in the previous section, then knowledge of the entities $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ allows scene structure to be computed, up to scale, from instantaneous optical flow.

We adopt the form of $\boldsymbol{A}$ given in (23). Assuming that $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ are known, we solve for $[\boldsymbol{x}^T, \dot{\boldsymbol{x}}^T]^T$ given $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$. Note that, of the entities $\boldsymbol{x}$ and $\dot{\boldsymbol{x}}$, solely $\boldsymbol{x}$ is needed for scene reconstruction.

First, using (12) and (13), we determine the values of $\boldsymbol{p}$ and $\dot{\boldsymbol{p}}$. Next, substituting (9) and (10) into (7), we find that

$$x_3\big(\dot{f}\boldsymbol{p} - f(\dot{\boldsymbol{p}} + \widehat{\boldsymbol{\omega}}\boldsymbol{p})\big) - \dot{x}_3 f\boldsymbol{p} + f^2\boldsymbol{v} = \mathbf{0}. \tag{32}$$

Clearly, $\dot{f}\boldsymbol{p} - f(\dot{\boldsymbol{p}} + \widehat{\boldsymbol{\omega}}\boldsymbol{p})$ and $f\boldsymbol{p}$ are known, $\boldsymbol{v}$ is partially known (namely $\boldsymbol{\pi}(\boldsymbol{v})$ is known), and x_3 and $\dot{x}_3$ are unknown. Assume temporarily that $\boldsymbol{v}$ is known. Then (32) can immediately be employed to find x_3 and $\dot{x}_3$. Indeed, bearing in mind that $\dot{f}\boldsymbol{p} - f(\dot{\boldsymbol{p}} + \widehat{\boldsymbol{\omega}}\boldsymbol{p})$, $f\boldsymbol{p}$ and $f^2\boldsymbol{v}$ are column vectors with 3 entries, one can regard (32) as being a system of 3 linear equations (algebraic not differential!) in x_3 and $\dot{x}_3$, and this system can easily be solved for the two unknowns. Upon finding x_3 and $\dot{x}_3$, we use (9) and (10) to determine $\boldsymbol{x}$ and $\dot{\boldsymbol{x}}$. With $\boldsymbol{x}$ thus specified, scene reconstruction is complete.

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

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

In view of (9) and (10), if $x_3 \neq 0$, then the last equation is equivalent to

$$\widehat{\boldsymbol{x}}(\dot{\boldsymbol{x}} + \widehat{\boldsymbol{\omega}}\boldsymbol{x}) = \mathbf{0}$$

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

We are left with the task of determining $\boldsymbol{v}$. Fix $\|\boldsymbol{v}\|$ arbitrarily as a positive value. In view of $\boldsymbol{v} \neq \mathbf{0}$, one of the components of $\boldsymbol{v}$, say v_3, is non-zero. Since

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

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

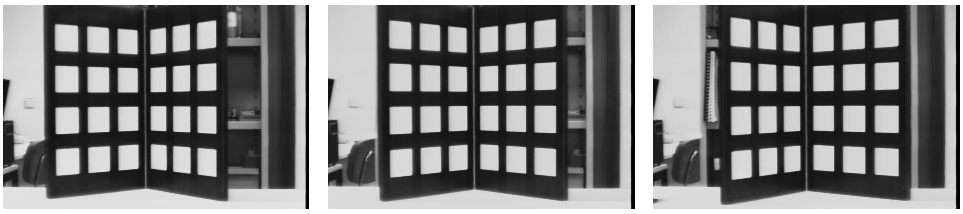

Figure 2. Image sequence of a calibration grid.

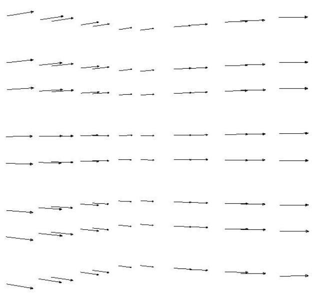

Figure 3. Optical flow.

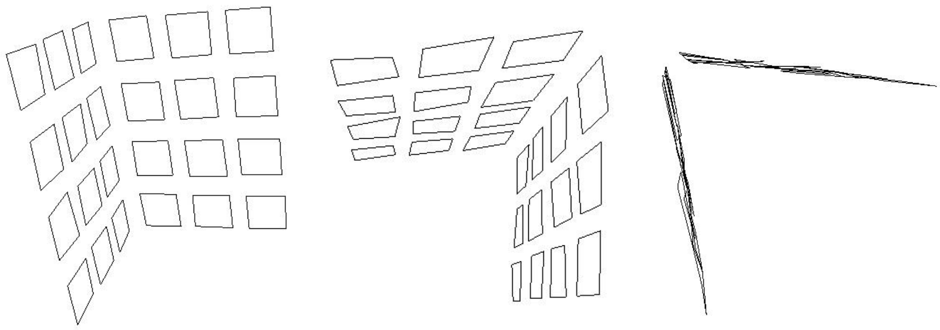

Figure 4. Reconstruction from various views.

7. Experimental results

In order to assess the applicability and correctness of the approach, a simple test with real-world imagery was performed. The three images shown in Figure 2 were captured via a Phillips CCD camera with a 12.5 mm lens. Corners were localised to sub-pixel accuracy with the use of a corner detector, correspondences between the images were obtained, and the optical flow depicted in Figure 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 $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ from the optical flow. Closed-form expressions described earlier were employed to self-calibrate the system. With the seven key parameters recovered, the reconstruction displayed in Figure 4 was finally obtained. Note that reconstructed points in 3-space have been connected by line segments so as to convey clearly the patterns of the calibration 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 may 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 intrinsics, permitted the specification of closed form expressions (in terms of the composite ratio of some entries of the two matrices) for the five computable ego-motion 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 3D reconstruction of the grid squares was visually pleasing, confirming the validity of the theory, and suggesting that the approach holds promise.

Acknowledgements

This research was in part supported by a grant of the Australian Research Council, and 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.

References

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

[2] O. D. Faugeras, Q. T. Luong, and S. J. Maybank, "Camera self-calibration: theory and experiments," in *Computer Vision—ECCV '92*, G. Sandini, ed., Vol. 588 of Lecture Notes in Computer Science, (Springer, Berlin, 1992), pp. 321–334.

[3] S. J. Maybank and O. D. Faugeras, "A theory of self-calibration of a moving camera," Int. J. Computer Vision **8**, 123–151 (1992).

[4] Q.-T. Luong, R. Deriche, O. D. Faugeras, and T. Papadopoulo, "On determining the fundamental matrix: analysis of different methods and experimental results," Technical Report No. 1894 (INRIA, Sophia Antipolis, France, 1993).

[5] Q.-T. Luong and O. D. Faugeras, "The fundamental matrix: theory, algorithms, and stability analysis," Int. J. Computer Vision **17**, 43–75 (1996).

[6] N. C. Gupta and L. N. Kanal, "3-D motion estimation from motion field," Artificial Intelligence **78**, 45–86 (1995).

[7] D. J. Heeger and A. D. Jepson, "Subspace methods for recovering rigid motion I: algorithm and implementation," Int. J. Computer Vision **7**, 95–117 (1992).

[8] K. Kanatani, "3-D interpretation of optical flow by renormalization," Int. J. Computer Vision **11**, 267–282 (1993).

[9] K. Kanatani, *Geometric Computation for Machine Vision* (Clarendon Press, Oxford, 1993).

[10] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint* (The MIT Press, Cambridge, Mass., 1993).

[11] M. J. Brooks, L. Baumela, and W. Chojnacki, "An analytical approach to determining the egomotion of a camera having free intrinsic parameters," Technical Report No. 96-04 (Department of Computer Science, University of Adelaide, South Australia, 1996).

[12] M. J. Brooks, L. Baumela, and W. Chojnacki, "Egomotion from optical flow with an uncalibrated camera," in *Visual Communications and Image Processing '97*, J. Biemond and E. J. Delp, eds., Vol. 3024 of Proc. SPIE (1997), pp. 220–228.

[13] P. H. S. Torr, personal communication.

[14] K. Åström and A. Heyden, "Multilinear forms in the infinitesimal-time case," in *Proceedings, CVPR '96, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (IEEE Computer Society Press, Los Alamitos, CA, 1996), pp. 833–838.

[15] B. K. P. Horn, *Robot Vision* (The MIT Press, Cambridge, Mass., 1986).

[16] R. I. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Computer Vision—ECCV '92*, G. Sandini, ed., Vol. 588 of Lecture Notes in Computer Science, (Springer, Berlin, 1992), pp. 579–587.

(M. J. Brooks and W. Chojnacki) Department of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia

E-mail address, M. J. Brooks: mjb@cs.adelaide.edu.au

E-mail address, W. Chojnacki: wojtek@cs.adelaide.edu.au

(L. Baumela) Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus de Monteganedo s/n, 28660 Boadilla del Monte Madrid, Spain

E-mail address, L. Baumela: lbaumela@mayor.dia.fi.upm.es