3D RECONTRUCTION FROM OPTICAL FLOW GENERATED BY AN UNCALIBRATED CAMERA UNDERGOING UNKNOWN MOTION

Michael J. Brooks Wojciech Chojnacki
Anton van den Hengel
Department of Computer Science
University of Adelaide
Adelaide, SA 5005, Australia
{mjb,wojtek,hengel}@cs.adelaide.edu.au

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

KEY WORDS: Self-calibration, ego-motion, intrinsic parameters, epipolar equation, fundamental matrix, least-squares minimisation, outlier rejection, robust estimation.

ABSTRACT
A procedure is described 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 on the use of a differential epipolar equation that relates optical flow to the ego-motion and internal geometry of the camera. 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 a scale factor, from the derived self-calibration parameters and the optical flow data. Various least-squares techniques and an outlier rejection scheme are presented to facilitate robust estimation of the critical composite ratio. Experimental results are given that suggest the approach holds promise.

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 [5, 13].

In this paper we present 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 from corresponding points.

The proposed method of self-calibration rests on a constraint that we term a *differential epipolar equation*. It relates optical flow to the ego-motion and intrinsic parameters of the camera. 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* [11, 12], the differential epipolar equation incorporates two matrices. These matrices encode 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 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, 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 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, 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 phenomenon 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 [15]. 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.

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. Fur-

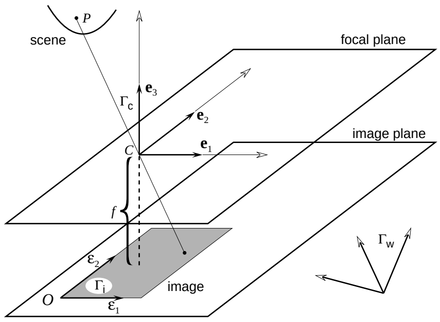

Figure 1: Image formation and coordinate frames.

thermore various least-squares techniques and an outlier rejection scheme are presented to facilitate robust estimation of the critical composite ratio. The self-calibration and reconstruction methods are tested on an optical flow field derived from a real-world image sequence of a calibration grid.

Some calculations will be omitted in this work. A more detailed exposition of part of the development that follows can be found in [3]. For related work dealing with the ego-motion of a calibrated camera, the reader is referred to [6, 7, 9, 10].

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). 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 $\Gamma_{\rm w}$ whose configuration in the scene will be fixed throughout. Associate with the camera an independent Cartesian coordinate frame $\Gamma_{\rm c}$, 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 $\Gamma_{\rm c}$ and $\Gamma_{\rm w}$ are equi-oriented by swapping two arbitrarily chosen basis vectors of $\Gamma_{\rm w}$ if initially the frames are counter-oriented. 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 $\Gamma_{\rm w}$ or that associated with $\Gamma_{\rm c}$ 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 $\Gamma_{\rm w}$. 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 $\Gamma_{\rm w}$.

Suppose that the camera undergoes smooth motion with respect to $\Gamma_{\rm w}$. At each time instant t, the location of the camera relative to $\Gamma_{\rm w}$ 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 $\Gamma_{\rm w}$ at t. Expanding this derivative with respect to the basis $\{\boldsymbol{e}_i(t)\}_{1\leq i\leq 3}$

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

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 $\Gamma_{\rm c}$ 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

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

Leaving the dependency of the $\boldsymbol{e}_i$ on 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

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

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 $\Gamma_{\rm w}$. The direction of $\boldsymbol{\eta}$ determines the axis of the instantaneous rotation of the camera, passing through C, relative to $\Gamma_{\rm w}$. Correspondingly, $\boldsymbol{\omega}$ represents instantaneous angular velocity of the camera relative to $\Gamma_{\rm c}$, with the direction of $\boldsymbol{\omega}$ determining the axis of the instantaneous rotation of the camera relative to $\Gamma_{\rm c}$.

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

$$P = \sum_i x_i \boldsymbol{e}_i + C. \tag{5}$$

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 $t \mapsto \boldsymbol{x}(t)$. Differentiating (5), taking into account that $\dot{P} = \mathbf{0}$, and using (1), (2) and (4), we find that

$$\dot{\boldsymbol{x}} + \widehat{\boldsymbol{\omega}}\boldsymbol{x} + \boldsymbol{v} = \mathbf{0}. \tag{6}$$

This equation shows how the camera motion induces motion of points, static in the scene, relative to the camera frame.

3 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 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

$$\boldsymbol{p} = -f\frac{\boldsymbol{x}}{x_3}. \tag{7}$$

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 $t \mapsto \boldsymbol{p}(t)$. Exploiting (6) and (7), one can directly verify that

$$\boldsymbol{p}^T\widehat{\boldsymbol{v}}\dot{\boldsymbol{p}} + \boldsymbol{p}^T\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}}\boldsymbol{p} = 0. \tag{8}$$

We call the above relation the *differential epipolar equation*. This term reflects the fact that equation (8) 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 [1] and its short version [2], where an analogue of (8), namely equation (12) presented below, is derived from the standard epipolar equation by applying a special differentiation operator. We also refer the reader to [9], where a similar derivation (though not involving any special differentiation procedure) is presented in the context of images formed on a sphere.

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 $\{\boldsymbol{\epsilon}_i\}_{1 \le i \le 2}$, 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 we append to $[m_1, m_2]^T$ 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

$$\boldsymbol{p} = \boldsymbol{A}\boldsymbol{m}, \tag{9}$$

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 [4, Section 3]).

The differential epipolar equation (8) can be restated so as to use 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 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 [8, Chapter 12]).

Letting $\boldsymbol{B} = \dot{\boldsymbol{A}}\boldsymbol{A}^{-1}$, set

$$\boldsymbol{C} = \frac{1}{2}\boldsymbol{A}^T(\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{v}} + \widehat{\boldsymbol{v}}\boldsymbol{B} - \boldsymbol{B}^T\widehat{\boldsymbol{v}})\boldsymbol{A}, \tag{10}$$

$$\boldsymbol{W} = \boldsymbol{A}^T\widehat{\boldsymbol{v}}\boldsymbol{A}. \tag{11}$$

Using (9), (10) and (11), we can write (8) in the form

$$\boldsymbol{m}^T\boldsymbol{W}\dot{\boldsymbol{m}} + \boldsymbol{m}^T\boldsymbol{C}\boldsymbol{m} = 0. \tag{12}$$

This is the differential epipolar equation for optical flow. A similar constraint, termed the *first-order expansion of the fundamental motion equation*, is derived using quite different means by Viéville and Faugeras [15]. In contrast with the above, however, it takes the form of an approximation rather than a strict equality.

In view of (11) 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 (12) 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*.

In view of (10) and (11),

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

Since $\boldsymbol{w}^T\boldsymbol{W} = \mathbf{0}$ and $\boldsymbol{W}\boldsymbol{w} = \mathbf{0}$, we conclude that

$$\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} = 0. \tag{13}$$

The left-hand side of this equation 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 seven-dimensional submanifold of $\mathbb{P}^8$, a fact already noted in [15].

5 SELF-CALIBRATION WITH FREE FOCAL LENGTH

Of the key parameters, six describe the ego-motion of the camera, and the rest describe the internal geometry of the camera. Only five ego-motion parameters can, however, be determined from image data, as one parameter is lost because of scale indeterminacy. Given that $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is a member of a seven-dimensional hypersurface in $\mathbb{P}^8$, the total number of key parameters that can be recovered by exploiting $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ does not exceed seven. If we want to recover all five 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

$$\boldsymbol{A}(t) = \begin{bmatrix} 1 & 0 & -i_1 \\ 0 & 1 & -i_2 \\ 0 & 0 & -f(t) \end{bmatrix}, \tag{14}$$

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 $\boldsymbol{\pi}(\boldsymbol{v})$ 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 five ego-motion parameters ($\boldsymbol{\omega}$ accounting for three parameters and $\boldsymbol{\pi}(\boldsymbol{v})$ accounting for two parameters), and f and $\dot{f}$ account for two intrinsic parameters. Note that $\boldsymbol{v}$ is not wholly recoverable, as the length of $\boldsymbol{v}$ is indeterminate.

We now outline the self-calibration procedure. We first make a reduction to the case $i_1 = i_2 = 0$. To this end, we represent $\boldsymbol{A}$ as

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

where

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

and let

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

It emerges that by 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 $i_1 = i_2 = 0$.

Let

$$\delta_1 = -\frac{\omega_1}{f}, \qquad \delta_2 = -\frac{\omega_2}{f}, \qquad \delta_3 = -\omega_3, \qquad \delta_4 = f^2, \qquad \delta_5 = \frac{\dot{f}}{f}. \tag{15}$$

Detailed calculation shows that δ_1, δ_2, and δ_3 satisfy

$$\begin{aligned} \delta_1 &= \frac{2c_{12}w_2 - (c_{22} - c_{11})w_1}{w_1^2 + w_2^2}, \\ \delta_2 &= \frac{2c_{12}w_1 + (c_{22} - c_{11})w_2}{w_1^2 + w_2^2}, \\ \delta_3 &= \frac{c_{11}w_1^2 + 2c_{12}w_1w_2 + c_{22}w_2^2}{w_3(w_1^2 + w_2^2)}. \end{aligned} \tag{16}$$

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 $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

Let

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

Further calculation shows that

$$\begin{aligned} \delta_4 &= \frac{1}{\Gamma}\left(w_1w_3d_1 + w_2w_3d_2 - (w_1^2 + w_2^2)d_3\right), \\ \delta_5 &= \frac{1}{\Gamma}\big(((w_1w_2\delta_1 + (w_2^2 + w_3^2)\delta_2)d_1 \\ &\quad - ((w_1^2 + w_3^2)\delta_1 + w_1w_2\delta_2)d_2 \\ &\quad + (w_2w_3\delta_1 - w_1w_3\delta_2)d_3\big), \end{aligned} \tag{17}$$

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 (15), (16) and (17), we obtain

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

By (11),

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

and, since f has already been specified, we have

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

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

It should be borne in mind that for the above self-calibration procedure to work a number of conditions must be met. Closer examination reveals that it is necessary 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

We now tackle the problem of scene reconstruction. We show 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 a scale factor, from instantaneous optical flow.

We adopt the form of $\boldsymbol{A}$ given in (14). Assuming that $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ are known, we solve for $\boldsymbol{x}$ given $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$.

First, using (9) and the equation obtained by differentiating both sides of (9), we determine the values of $\boldsymbol{p}$ and $\dot{\boldsymbol{p}}$. Combining (6), (7), and the equation obtained by differentiating both sides of (7), we find that

$$x_3\big(\dot{f}\boldsymbol{p} - f(\dot{\boldsymbol{p}} + \widehat{\boldsymbol{\omega}}\boldsymbol{p})\big) - \dot{x}_3 f\boldsymbol{p} + f^2\boldsymbol{v} = \boldsymbol{0}. \tag{18}$$

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 (18) 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 three entries, one can regard (18) 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 (7) to determine $\boldsymbol{x}$. With $\boldsymbol{x}$ thus specified, scene reconstruction is complete.

A moment's reflection reveals that in order for this method to work we need to assume that $\widehat{\boldsymbol{x}}\boldsymbol{v} \neq \boldsymbol{0}$ whenever $x_3 \neq 0$. In particular, this means that $\boldsymbol{v}$ has to be non-zero.

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

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

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 (18) be non-negative. This requirement simply reflects the fact that the scene is in front of the camera.

7 ESTIMATING $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$

Let $\mathcal{S} = \{[\boldsymbol{m}_i^T, \dot{\boldsymbol{m}}_i^T]^T \mid i = 1, \ldots, n\}$ with $n \geq 7$ be a data set representing measurements of a portion of instantaneous optical flow. It is of great practical importance to consider ways in which *estimates* of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ can be derived from $\mathcal{S}$. Note that the ratio $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ can be identified with the pair $(\boldsymbol{C}, \boldsymbol{W})$ satisfying the normalisation condition $\|\boldsymbol{C}\|^2 + \|\boldsymbol{W}\|^2 = 1$ (here, for a given matrix $\boldsymbol{S} = [s_{ij}]_{1\leq i,j\leq k}$, we let $\|\boldsymbol{S}\| = (\sum_{i,j=1}^k s_{ij}^2)^{1/2}$). Therefore estimates of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ can always be expressed in terms of normalised pairs $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$.

7.1 Seven-point estimator

If $n = 7$, then an estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ can be obtained by solving a system of seven linear equations

$$\boldsymbol{m}_i^T \boldsymbol{W} \dot{\boldsymbol{m}}_i + \boldsymbol{m}_i^T \boldsymbol{C} \boldsymbol{m}_i = 0 \tag{19}$$

and the non-linear equation (13). These equations are homogeneous in the entries of $\boldsymbol{C}$ and $\boldsymbol{W}$, and effectively provide seven constraints for the ratio $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$. Since (13) is cubic, one or three estimates $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ can be obtained by exploiting these constraints (the complex solutions are discarded).

7.2 Least squares estimator based on algebraic distances

If $n \geq 8$, then the linear homogeneous equations forming system (19) provide $n - 1 \geq 7$ constraints for $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$. They can serve as a basis for estimation of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$. The redundancy in system (19) suggests a least squares solution. In order to develop such a solution, a cost function has to be specified. The simplest choice is the function J_1 given by

$$J_1(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S}) = \sum_{i=1}^n |\boldsymbol{m}_i^T \boldsymbol{W} \dot{\boldsymbol{m}}_i + \boldsymbol{m}_i^T \boldsymbol{C} \boldsymbol{m}_i|^2.$$

Here, a residual $|\boldsymbol{m}^T \boldsymbol{W} \dot{\boldsymbol{m}} + \boldsymbol{m}^T \boldsymbol{C} \boldsymbol{m}|$ measures the *algebraic distance* between the vector $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$ and the manifold

$$\mathcal{M}_{\boldsymbol{C},\boldsymbol{W}} = \{[\boldsymbol{n}^T, \dot{\boldsymbol{n}}^T]^T \mid \boldsymbol{n}^T \boldsymbol{W} \dot{\boldsymbol{n}} + \boldsymbol{n}^T \boldsymbol{C} \boldsymbol{n} = 0\}.$$

The estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ based on J_1 is a unique pair $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ that satisfies

$$J_1(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}}; \mathcal{S}) = \min\{J_1(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S}) \mid \|\boldsymbol{C}\|^2 + \|\boldsymbol{W}\|^2 = 1\}.$$

Because J_1 is quadratic in the entries of $\boldsymbol{C}$ and $\boldsymbol{W}$, this estimate can be computed explicitly with the use of Lagrange multipliers. If we identify $(\boldsymbol{C}, \boldsymbol{W})$ with the vector

$$\boldsymbol{d} = [c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{33}, w_{12}, w_{13}, w_{23}]^T,$$

let

$$\boldsymbol{u}_i = \begin{bmatrix} m_{i,1}^2 \\ 2m_{i,1}m_{i,2} \\ 2m_{i,1}m_{i,3} \\ m_{i,2}^2 \\ 2m_{i,2}m_{i,3} \\ m_{i,3}^3 \\ m_{i,1}\dot{m}_{i,2} - m_{i,2}\dot{m}_{i,1} \\ m_{i,1}\dot{m}_{i,3} - m_{i,3}\dot{m}_{i,1} \\ m_{i,2}\dot{m}_{i,3} - m_{i,3}\dot{m}_{i,2} \end{bmatrix},$$

$$\boldsymbol{U}_n = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]^T,$$

and identify the estimate $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ with a corresponding vector $\widehat{\boldsymbol{d}}$, then $\widehat{\boldsymbol{d}}$ turns out to coincide with the eigenvector of $\boldsymbol{U}_n^T \boldsymbol{U}_n$ associated with the smallest eigenvalue. This eigenvector can be efficiently calculated by employing the method of singular value decomposition.

7.3 Enforcing the cubic constraint

The estimate $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ developed above may fail to satisfy equation (13). A procedure for modifying estimates to accommodate this constraint is therefore needed.

Given a (normalised) estimate $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$, let

$$\widehat{\boldsymbol{C}}_\rho = \frac{\widehat{\boldsymbol{C}} - \boldsymbol{P}\widehat{\boldsymbol{C}}\boldsymbol{P}}{\|\widehat{\boldsymbol{C}} - \boldsymbol{P}\widehat{\boldsymbol{C}}\boldsymbol{P}\|^2 + \|\widehat{\boldsymbol{W}}\|^2},$$
$$\widehat{\boldsymbol{W}}_\rho = \frac{\widehat{\boldsymbol{W}}}{\|\widehat{\boldsymbol{C}} - \boldsymbol{P}\widehat{\boldsymbol{C}}\boldsymbol{P}\|^2 + \|\widehat{\boldsymbol{W}}\|^2},$$

where

$$\boldsymbol{P} = \boldsymbol{I} + \|\widehat{\boldsymbol{w}}\|^{-2}\widehat{\boldsymbol{W}}^2.$$

Note that the pair $(\widehat{\boldsymbol{C}}_\rho, \widehat{\boldsymbol{W}}_\rho)$ comes automatically normalised. It is easily verified that if $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ satisfies (13), then $\boldsymbol{P}\widehat{\boldsymbol{C}}\boldsymbol{P} = \boldsymbol{0}$. Hence $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}_\rho$ whenever (13) holds for $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$. Since $\boldsymbol{P}\widehat{\boldsymbol{w}} = \widehat{\boldsymbol{w}}$ and $\widehat{\boldsymbol{w}}^T\boldsymbol{P} = \widehat{\boldsymbol{w}}^T$, it follows that $\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{C}}_\rho\widehat{\boldsymbol{w}} = 0$, which in turn immediately implies that $\widehat{\boldsymbol{w}}_\rho{}^T\widehat{\boldsymbol{C}}_\rho\widehat{\boldsymbol{w}}_\rho = 0$. Thus passing from $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ to $(\widehat{\boldsymbol{C}}_\rho, \widehat{\boldsymbol{W}}_\rho)$ gives the required modification procedure.

7.4 Least squares estimator based on Euclidean distances

The algebraic distance mentioned above has no geometric significance. In contrast, the expression

$$\delta(\boldsymbol{C}, \boldsymbol{W}, \boldsymbol{m}, \dot{\boldsymbol{m}}) = \frac{|\boldsymbol{m}^T\boldsymbol{W}\dot{\boldsymbol{m}} + \boldsymbol{m}^T\boldsymbol{C}\boldsymbol{m}|}{\sqrt{\|2\boldsymbol{C}\boldsymbol{m} + \boldsymbol{W}\dot{\boldsymbol{m}}\|^2 + \|\boldsymbol{W}\boldsymbol{m}\|^2}}$$

is geometrically meaningful being an approximation of the *Euclidean distance* between $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$ and $\mathcal{M}_{\boldsymbol{C},\boldsymbol{W}}$. This observation suggests using the cost function

$$J_2(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S}) = \sum_{i=1}^{n} |\delta(\boldsymbol{C}, \boldsymbol{W}, \boldsymbol{m}, \dot{\boldsymbol{m}})|^2$$

instead of $J_1(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S})$. The least-square estimate $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ based on J_2 is characterised by the requirement

$$J_2(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}}; \mathcal{S}) = \min\{J_2(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S}) \mid \|\boldsymbol{C}\|^2 + \|\boldsymbol{W}\|^2 = 1\}.$$

Because of the complicated way in which $\boldsymbol{C}$ and $\boldsymbol{W}$ enter J_2, it is not clear whether $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ can be given an explicit form. Various techniques might be employed to evolve an approximation of $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$. One such technique is proposed next.

7.5 Iteratively reweighted least squares estimator

Let $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ be the least-square estimate based on J_2, and let

$$J_3(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S}) = \sum_{i=1}^{n} |\widetilde{\delta}(\boldsymbol{C}, \boldsymbol{W}, \boldsymbol{m}_i, \dot{\boldsymbol{m}}_i)|^2,$$

where

$$\widetilde{\delta}(\boldsymbol{C}, \boldsymbol{W}, \boldsymbol{m}, \dot{\boldsymbol{m}}) = \frac{|\boldsymbol{m}^T\boldsymbol{W}\dot{\boldsymbol{m}} + \boldsymbol{m}^T\boldsymbol{C}\boldsymbol{m}|}{\sqrt{\|2\widehat{\boldsymbol{C}}\boldsymbol{m} + \widehat{\boldsymbol{W}}\dot{\boldsymbol{m}}\|^2 + \|\widehat{\boldsymbol{W}}\boldsymbol{m}\|^2}}.$$

The denominator in the expression for $\widetilde{\delta}$ does not depend on $(\boldsymbol{C}, \boldsymbol{W})$, and so minimisation of $J_3(\boldsymbol{C}, \boldsymbol{W}; \mathcal{S})$ subject to the constraint $\|\boldsymbol{C}\|^2 + \|\boldsymbol{W}\|^2 = 1$ leads to an estimator falling into the category of weighted least squares techniques. Employing Lagrange multipliers, we verify at once that the least-square estimate based on J_3 can be identified with the eigenvector of $\boldsymbol{U}_n^T\boldsymbol{R}_{\widehat{\boldsymbol{C}},\widehat{\boldsymbol{W}}}{}^2\boldsymbol{U}_n$ corresponding to the smallest eigenvalue, where $\boldsymbol{R}_{\widehat{\boldsymbol{C}},\widehat{\boldsymbol{W}}}$ is a weight matrix given by

$$\boldsymbol{R}_{\widehat{\boldsymbol{C}},\widehat{\boldsymbol{W}}} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_n \end{bmatrix},$$
$$\lambda_i = \left(\|2\widehat{\boldsymbol{C}}\boldsymbol{m}_i + \widehat{\boldsymbol{W}}\dot{\boldsymbol{m}}_i\|^2 + \|\widehat{\boldsymbol{W}}\boldsymbol{m}_i\|^2\right)^{-1}.$$

Using this observation and bearing in mind that J_3 is an approximation of J_2, we can now propose the following iteratively reweighted least squares estimator that simultaneously seeks to minimise J_2 and to accommodate the cubic constraint (13):

1. (i) Compute $(\widehat{\boldsymbol{C}}_0, \widehat{\boldsymbol{W}}_0)$ using least-square fitting based on J_1.
   (ii) Generate $(\widehat{\boldsymbol{C}}_{0,\rho}, \widehat{\boldsymbol{W}}_{0,\rho})$ from $(\widehat{\boldsymbol{C}}_0, \widehat{\boldsymbol{W}}_0)$ using the procedure described in Subsection 7.3.
2. Compute the weight matrix $\boldsymbol{R}_{\widehat{\boldsymbol{C}}_{k-1,\rho},\widehat{\boldsymbol{W}}_{k-1,\rho}}$.
3. (i) Compute the eigenvector of $\boldsymbol{U}_n^T(\boldsymbol{R}_{\widehat{\boldsymbol{C}}_{k-1,\rho},\widehat{\boldsymbol{W}}_{k-1,\rho}})^2\boldsymbol{U}_n$ corresponding to the smallest eigenvalue and represent this eigenvector as $(\widehat{\boldsymbol{C}}_k, \widehat{\boldsymbol{W}}_k)$.
   (ii) Generate $(\widehat{\boldsymbol{C}}_{k,\rho}, \widehat{\boldsymbol{W}}_{k,\rho})$ from $(\widehat{\boldsymbol{C}}_k, \widehat{\boldsymbol{W}}_k)$ using the procedure described in Subsection 7.3.
4. If $(\widehat{\boldsymbol{C}}_{k,\rho}, \widehat{\boldsymbol{W}}_{k,\rho})$ is sufficiently close to $(\widehat{\boldsymbol{C}}_{k-1,\rho}, \widehat{\boldsymbol{W}}_{k-1,\rho})$, then terminate the procedure; otherwise return to Step 2.

7.6 Robust estimation

Typically, a data set comprises two subsets: a large, dominant subset of valid data or *inliers*, and a relatively small subset of *outliers* or *contaminants*. Least squares minimisation is global in nature and hence vulnerable to distortion by outliers. To obtain robust estimates, outliers have to be detected and rejected. To identify the outliers, we use the method of least median of squares (LMedS) as developed in [14]. This technique is representative of *robust statistics* methods that find a fit without removing the outliers. Once an LMedS fit is generated, the outliers can then be identified (if necessary) as those data which are inconsistent with the fit. The remaining inliers can next be processed with the use of a least squares technique, which results in a final, relatively robust estimate.

The LMedS estimator repeatedly samples the data set, computing a statistic from each sample and combining the results of many such computations to evolve a dominant fit. The robustness of this method stems from the fact that at any one time it considers only a subset of the data set.

We use the LMedS estimator involving samples formed by subsets of $\mathcal{S}$ containing seven elements. The size of samples is such that it is minimal to allow an estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ to be determined from a single sample. Ideally, the estimator should consider the set of all seven-element samples. In practice, to make the search computationally

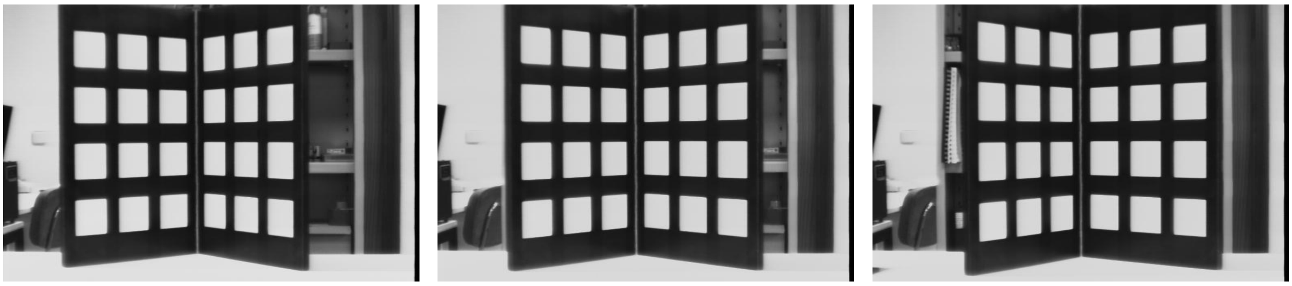

Figure 2: Image sequence of a calibration grid.

feasible, the sample space is reduced to a family of m randomly chosen samples. The number m is determined as follows. Assume that the proportion of outliers in $\mathcal{S}$ does not exceed ϵ, where $0 \leq \epsilon \leq 1$. Then the probability P that a family of m samples contains at least one element that is outlier-free is approximatively given by

$$P = 1 - (1 - (1 - \epsilon)^7)^m.$$

Consequently,

$$m = \left[\frac{\log(1 - P)}{\log(1 - (1 - \epsilon)^7)} \right],$$

where $[x]$ denotes the integral part of x. We exploit this formula by assuming that $\epsilon = 0.2$ and $P = 0.95$.

Once m is fixed, the LMedS estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is obtained in the following steps:

1. Using a Monte Carlo type technique, select a family $\mathcal{S}_0$ consisting of m subsets of $\mathcal{S}$, each subset containing seven elements.
2. For each $s \in \mathcal{S}_0$, compute three estimates $(\widehat{\boldsymbol{C}}_{s,k}, \widehat{\boldsymbol{W}}_{s,k})$ ($k \in \{1, 2, 3\}$) by using the seven-point algorithm (all three estimates may coincide).
3. For each $(s, k) \in \mathcal{S}_0 \times \{1, 2, 3\}$, determine the median

$$M_{s,k} = \operatorname{med}\{\delta(\boldsymbol{m}_i, \dot{\boldsymbol{m}}_i, \widehat{\boldsymbol{C}}_{s,k}, \widehat{\boldsymbol{W}}_{s,k})^2 \mid i = 1, \ldots, n\}.$$

4. Letting $(s_m, k_m) \in \mathcal{S}_0 \times \{1, 2, 3\}$ be such that

$$M_{s_m,k_m} = \min\{M_{s,k} \mid (s, k) \in \mathcal{S}_0 \times \{1, 2, 3\}\},$$

take $(\widehat{\boldsymbol{C}}_{s_m,k_m}, \widehat{\boldsymbol{W}}_{s_m,k_m})$ for the LMedS estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

With the LMedS estimate at hand, we proceed to identify outliers by applying the following procedure:

1. Take

$$\hat{\sigma} = 1.4826 \left(1 + \frac{5}{n - 7}\right) \sqrt{M_{s_m,k_m}}$$

for the *robust standard deviation.*

2. Declare $[\boldsymbol{m}_i^T, \dot{\boldsymbol{m}}_i^T]^T$ to be an outlier if and only if

$$\delta(\boldsymbol{m}, \dot{\boldsymbol{m}}, \widehat{\boldsymbol{C}}_{s_m,k_m}, \widehat{\boldsymbol{W}}_{s_m,k_m}) > 2.5\hat{\sigma}.$$

Once the outliers have been detected and removed, we can apply one of least-squares techniques proposed earlier to the remaining elements of $\mathcal{S}$ and thereby obtain a robust estimate of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

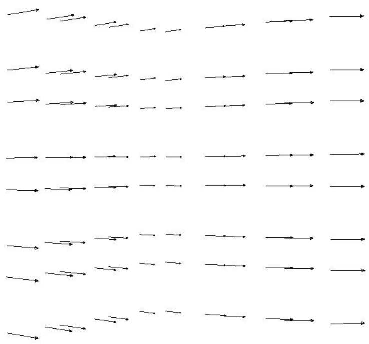

Figure 3: Optical flow.

8 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 by 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 least-squares estimation based on algebraic distances 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.

ACKNOWLEDGEMENTS

This research was in part supported by a grant of the Australian Research Council. The authors are grateful for the comments of Phil Torr, Andrew Zisserman, Lourdes de Agapito, Zhengyou Zhang, and Kenichi Kanatani.

REFERENCES

[1] M. J. Brooks, L. Baumela, and W. Chojnacki, *An analytical approach to determining the egomotion of a camera having free intrinsic parameters*, Tech. Report

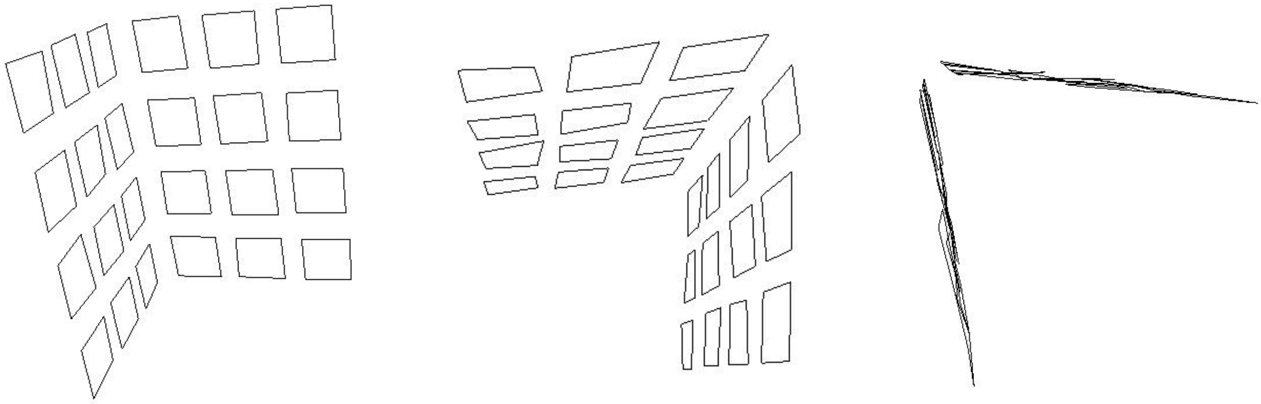

Figure 4: Reconstruction from various views.

96-04, Department of Computer Science, University of Adelaide, South Australia, June 1996.

[2] ———, *Egomotion from optical flow with an uncalibrated camera*, Visual Communications and Image Processing '97 (J. Biemond and E. J. Delp, eds.), Proceedings of SPIE, vol. 3024, 1997, pp. 220–228.

[3] M. J. Brooks, W. Chojnacki, and L. Baumela, *Determining the egomotion of an uncalibrated camera from instantaneous optical flow*, Journal of the Optical Society of America A **14** (1997), no. 10, 2670–2677.

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

[5] O. D. Faugeras, Q. T. Luong, and S. J. Maybank, *Camera self-calibration: Theory and experiments*, Computer Vision—ECCV '92, Second European Conference on Computer Vision (Santa Margherita Ligure, Italy, May 19–22, 1992) (G. Sandini, ed.), Lecture Notes in Computer Science, vol. 588, Springer-Verlag, Berlin, 1992, pp. 321–334.

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

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

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

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

[10] ———, *Geometric Computation for Machine Vision*, Clarendon Press, Oxford, 1993.

[11] Q.-T. Luong, R. Deriche, O. D. Faugeras, and T. Papadopoulo, *On determining the fundamental matrix: analysis of different methods and experimental results*, Tech. Report 1894, Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France, April 1993.

[12] Q.-T. Luong and O. D. Faugeras, *The fundamental matrix: theory, algorithms, and stability analysis*, International Journal of Computer Vision **17** (1996), no. 1, 43–75.

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

[14] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Regression*, John Wiley & Sons, New York, 1987.

[15] T. Viéville and O. D. Faugeras, *Motion analysis with a camera with unknown, and possibly varying intrinsic parameters*, Proceedings of the Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995), IEEE Computer Society Press, Los Alamitos, CA, 1995, pp. 750–756.