Estimation of structure from motion in the uncalibrated case

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`

Abstract

A procedure is described for determining structure from motion in the uncalibrated case. 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 of elements of two special matrices. Techniques are given for robust determination of this ratio via least-squares methods, and for the elimination of outlying flow vectors. Experimental results confirm that the approach holds promise.

Keywords: Self-calibration, ego-motion, intrinsic parameters, epipolar equation, fundamental matrix, least squares estimation, outlier rejection.

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 [3, 6].

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* [4, 5], 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.

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. Furthermore 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 crucial if tedious calculations will be omitted in this work. For a more detailed exposition of part of the development that follows, the interested reader is referred to [2].

2 Differential epipolar equation

Consider a camera with an associated coordinate frame such that the origin of the frame coincides with the camera's optical centre, two basis vectors span the focal plane, and the other basis vector passes through the optical axis. Suppose that the camera undergoes smooth motion. Denote by $\boldsymbol{v} = [v_1, v_2, v_3]^T$ and $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$ the camera's instantaneous *translational velocity* and instantaneous *angular velocity* with respect to the camera frame. Let P be a static point in space and let $\boldsymbol{x} = [x_1, x_2, x_3]^T$ be the coordinates of the vector connecting C with P in the vector basis of the camera frame. As the camera moves, the position of P relative to the camera frame will change accordingly and will be recorded in the function $t \mapsto \boldsymbol{x}(t)$. Straightforward calculation shows that this function satisfies

$$\dot{\boldsymbol{x}} + \widehat{\boldsymbol{\omega}}\boldsymbol{x} + \boldsymbol{v} = \boldsymbol{0}, \tag{1}$$

where

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

Let $\boldsymbol{p} = [p_1, p_2, p_3]^T$ be the coordinates of the perspective projection of P, through C, onto the image plane $\{\boldsymbol{x} \in \mathbb{R}^3 \colon x_3 = -f\}$, relative to the camera frame; here f is the focal length. It is easily seen that

$$\boldsymbol{p} = -f\frac{\boldsymbol{x}}{x_3}. \tag{2}$$

Exploiting (1) and (2), one can directly verify that the function $t \mapsto \boldsymbol{p}(t)$ satisfies

$$\boldsymbol{p}^T\widehat{\boldsymbol{v}}\dot{\boldsymbol{p}} + \boldsymbol{p}^T\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}}\boldsymbol{p} = 0. \tag{3}$$

We call this relation the *differential epipolar equation*. It can be viewed as a limiting case of the familiar epipolar equation in stereo vision (cf. [1]).

To account for the geometry of the image, it is useful to adopt a separate coordinate frame in the image plane. Let $[m_1, m_2]^T$ be the coordinates of the image of P in the vector basis of this frame. 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

$$\boldsymbol{p} = \boldsymbol{A}\boldsymbol{m}, \tag{4}$$

where $\boldsymbol{A}$ is a 3×3 invertible matrix called the *intrinsic-parameter matrix*. The differential epipolar equation (3) can be restated so as to use the vector $[\boldsymbol{m}^T, \dot{\boldsymbol{m}}^T]^T$ in place of the 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*.

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

$$\boldsymbol{C} = \frac{1}{2}\boldsymbol{A}^T(\widehat{\boldsymbol{v}}\widehat{\boldsymbol{\omega}} + \widehat{\boldsymbol{\omega}}\widehat{\boldsymbol{v}} + \widehat{\boldsymbol{v}}\boldsymbol{B} - \boldsymbol{B}^T\widehat{\boldsymbol{v}})\boldsymbol{A}, \tag{5}$$

$$\boldsymbol{W} = \boldsymbol{A}^T\widehat{\boldsymbol{v}}\boldsymbol{A}. \tag{6}$$

Using (4), (5) and (6), we can write (3) in the form

$$\boldsymbol{m}^T\boldsymbol{W}\dot{\boldsymbol{m}} + \boldsymbol{m}^T\boldsymbol{C}\boldsymbol{m} = 0. \tag{7}$$

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 [8]. In contrast with the above, however, it takes the form of an approximation rather than a strict equality.

In view of (6) 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 $\boldsymbol{W}$ given by the composite ratio $(c_{11} : c_{12} : c_{13} : c_{22} : c_{23} : c_{33} : w_1 : w_2 : w_3)$. Clearly, knowing $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ amounts to knowing $\boldsymbol{C}$ and $\boldsymbol{W}$ to within a common scalar factor. It can be inferred from (5) and (6) that

$$\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w} = 0. \tag{8}$$

This equation shows that $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$ is constrained to a seven-dimensional manifold, a fact already noted in [8].

The differential epipolar equation (7) 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*.

3 Self-calibration with free focal length

We now introduce some camera parameters into our analysis. Let a *free* parameter be one that is unknown and which may vary continuously with time. Assume that the focal length is unknown and free, that pixels are square with unit length, and that the principal point is fixed and known. In this situation, $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & -i_1 \\ 0 & 1 & -i_2 \\ 0 & 0 & -f \end{bmatrix}, \tag{9}$$

where i_1 and i_2 are the coordinates of the known principal point, and f is the unknown focal length. Let $\boldsymbol{\pi}(\boldsymbol{v})$ be the *projective form* of $\boldsymbol{v}$ given by the composite ratio $(v_1 : v_2 : v_3)$. 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 key parameters $\boldsymbol{\omega}$, $\boldsymbol{\pi}(\boldsymbol{v})$, f and $\dot{f}$ in terms of $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

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

$$\delta_1 = -\frac{\omega_1}{f}, \qquad \delta_2 = -\frac{\omega_2}{f}, \qquad \delta_3 = -\omega_3, \qquad \delta_4 = f^2, \qquad \delta_5 = \frac{\dot{f}}{f}. \tag{10}$$

Detailed calculation shows that δ_1, δ_2, and δ_3 satisfy

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

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 $d_1 = 2c_{13} + w_1\delta_3$, $d_2 = 2c_{23} + w_2\delta_3$, and $d_3 = c_{33}$. Further calculation shows that

$$\begin{aligned}
\delta_4 &= \frac{1}{\Gamma}\left(w_1w_3d_1 + w_2w_3d_2 - (w_1^2 + w_2^2)d_3\right), \\
\delta_5 &= \frac{1}{\Gamma}\big((w_1w_2\delta_1 + (w_2^2 + w_3^2)\delta_2)d_1 \\
&\qquad - ((w_1^2 + w_3^2)\delta_1 + w_1w_2\delta_2)d_2 \\
&\qquad + (w_2w_3\delta_1 - w_1w_3\delta_2)d_3\big),
\end{aligned} \tag{12}$$

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 (10), (11) and (12), we obtain

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

By (6),

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

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

4 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 (9). 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 (4) and the equation obtained by differentiating both sides of (4), we determine the values of $\boldsymbol{p}$ and $\dot{\boldsymbol{p}}$. Combining (1), (2), and the equation obtained by differentiating both sides of (2), we find that

$$x_3\big(\dot{f}\boldsymbol{p} - f(\dot{\boldsymbol{p}} + \widehat{\boldsymbol{\omega}}\boldsymbol{p})\big) - \dot{x}_3 f\boldsymbol{p} + f^2\boldsymbol{v} = \boldsymbol{0}. \tag{13}$$

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 (13) 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 (13) 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 (2) 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

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

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

5 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 $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$.

5.1 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

$$\boldsymbol{m}_i^T \boldsymbol{W} \dot{\boldsymbol{m}}_i + \boldsymbol{m}_i^T \boldsymbol{C} \boldsymbol{m}_i = 0 \tag{14}$$

and the non-linear equation (8). 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 (8) is cubic, one or three estimates $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ can be obtained by exploiting these constraints (the complex solutions are discarded).

5.2 Least squares estimator based on algebraic distances

If $n \geq 8$, then the linear homogeneous equations forming system (14) 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 (14) 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

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

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.

5.3 Enforcing the cubic constraint

The estimate $(\widehat{\boldsymbol{C}}, \widehat{\boldsymbol{W}})$ developed above may fail to satisfy equation (8). 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 (8), then $\boldsymbol{P}\widehat{\boldsymbol{C}}\boldsymbol{P} = \boldsymbol{0}$. Hence $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{W}}_\rho$ whenever (8) 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.

5.4 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 $J_1(\boldsymbol{C},\boldsymbol{W};\mathcal{S})$. Because of the complicated way in which $\boldsymbol{C}$ and $\boldsymbol{W}$ enter J_2, it is not clear whether the least-square estimate based on J_2 can be given an explicit form. Various techniques might be employed to evolve an approximation of this estimate. One such technique is proposed next.

5.5 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

$$\boldsymbol{R}_{\widehat{\boldsymbol{C}},\widehat{\boldsymbol{W}}} = \left[\begin{array}{ccc} \lambda_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \lambda_n \end{array}\right],$$

$$\lambda_i = \left(\|2\widehat{\boldsymbol{C}}\boldsymbol{m}_i + \widehat{\boldsymbol{W}}\dot{\boldsymbol{m}}_i\|^2 + \|\widehat{\boldsymbol{W}}\boldsymbol{m}_i\|^2\right)^{-1}.$$

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 (8):

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 5.3.
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 5.3.
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.

5.6 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 [7]. 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.

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 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

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

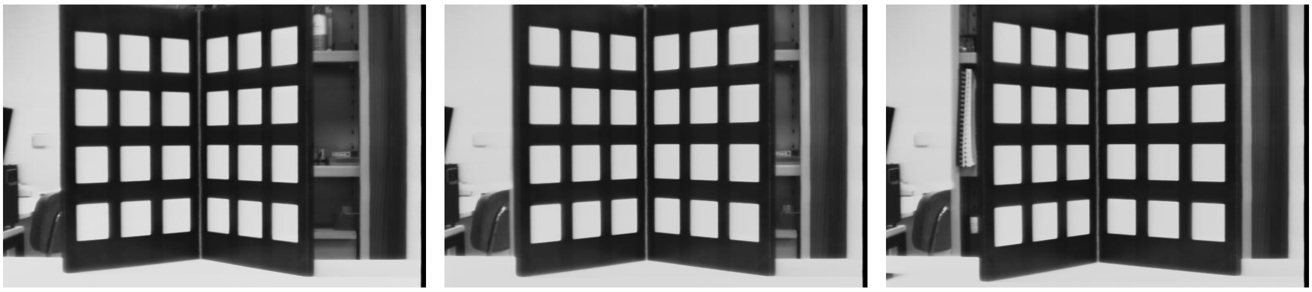

Figure 1: Image sequence of a calibration grid.

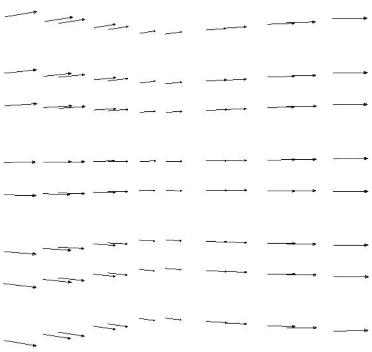

Figure 2: Optical flow.

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

4. Letting $(s_m, k_m) \in \mathcal{S}_0 \times \{1,2,3\}$ be such that
$$M_{s_m,k_m} = \min_{(s,k)\in\mathcal{S}_0\times\{1,2,3\}} M_{s,k},$$
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 $\boldsymbol{\pi}(\boldsymbol{C}, \boldsymbol{W})$.

6 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 1 were captured by a Phillips CCD camera with a 12.5-mm lens. Corners were localised to subpixel accuracy with the use of a corner detector, correspondences between the images were obtained, and the optical flow depicted in Figure 2 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 3

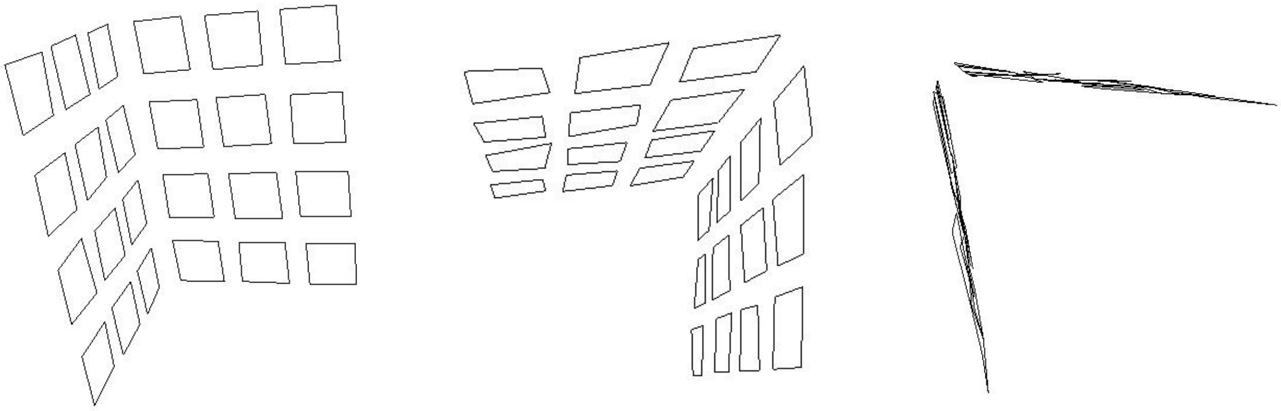

Figure 3: Reconstruction from various views.

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, *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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.