



## Linear Differential Algorithm for Motion Recovery: A Geometric Approach

YI MA, JANA KOŠECKÁ AND SHANKAR SASTRY

*Electronics Research Laboratory, University of California at Berkeley, Berkeley, CA 94720-1774, USA*

mayi@robotics.eecs.berkeley.edu

janka@robotics.eecs.berkeley.edu

sastry@robotics.eecs.berkeley.edu

**Abstract.** The aim of this paper is to explore a linear geometric algorithm for recovering the three dimensional motion of a moving camera from image velocities. Generic similarities and differences between the discrete approach and the differential approach are clearly revealed through a parallel development of an analogous motion estimation theory previously explored in Vieville, T. and Faugeras, O.D. 1995. In *Proceedings of Fifth International Conference on Computer Vision*, pp. 750–756; Zhuang, X. and Haralick, R.M. 1984. In *Proceedings of the First International Conference on Artificial Intelligence Applications*, pp. 366–375. We present a precise characterization of the space of differential essential matrices, which gives rise to a novel eigenvalue-decomposition-based 3D velocity estimation algorithm from the optical flow measurements. This algorithm gives a unique solution to the motion estimation problem and serves as a differential counterpart of the well-known SVD-based 3D displacement estimation algorithm for the discrete case. Since the proposed algorithm only involves linear algebra techniques, it may be used to provide a fast initial guess for more sophisticated nonlinear algorithms (Ma et al., 1998c. Electronic Research Laboratory Memorandum, UC Berkeley, UCB/ERL(M98/37)). Extensive simulation results are presented for evaluating the performance of our algorithm in terms of bias and sensitivity of the estimates with respect to different noise levels in image velocity measurements.

**Keywords:** differential epipolar constraint, differential essential matrix, optical flow, motion estimation

### 1. Introduction

The problem of estimating structure and motion from image sequences has been studied extensively by the computer vision community in the past decade. The various approaches differ in the kinds of assumptions they make about the projection model, the model of the environment, or the type of algorithms they use for estimating the motion and/or structure. Most techniques try to decouple the two problems by estimating the motion first, followed by the structure estimation. Thus the two are usually viewed as separate problems. In spite of the fact that the robustness of existing algorithms has been studied quite extensively, it has been suggested that the fact that the structure and motion estimation are decoupled typically hinders their

performance (McLauchlan and Murray, 1995). Some algorithms address the problem of motion and structure (shape) recovery simultaneously either in batch (Tomasi and Kanade, 1992) or recursive fashion (McLauchlan and Murray, 1995).

Approaches to motion estimation alone, can be partitioned into the discrete and differential methods depending on whether they use as input a set of point correspondences or image velocities. Among the efforts to solve the motion estimation problem, one of the more appealing approaches is the *essential matrix approach*, proposed by Longuet-Higgins, Huang and Faugeras et al. in 1980's (Longuet-Higgins, 1981). It shows that the relative 3D displacement of a camera can be recovered from an *intrinsic* geometric constraint between two images of the same scene, the so-called

*Longuet-Higgins constraint* (also called the *epipolar or essential constraint*). Estimating 3D motion can therefore be decoupled from estimation of the structure of the 3D scene. This endows the resulting motion estimation algorithms with some advantageous features: they do not need to assume any *a priori* knowledge about the scene; and are computationally simpler (compared to most non-intrinsic motion estimation algorithms), using mostly linear algebra techniques. Tsai and Huang (1984) proved that, given an essential matrix associated with the Longuet-Higgins constraint, there are only two possible 3D displacements. The study of the essential matrix then led to a three-step SVD-based algorithm for recovering the 3D displacement from noisy image correspondences, proposed in 1986 by Toscani and Faugeras (1986) and later summarized in Maybank (1993).

However, the essential matrix approach based on the epipolar constraint recovers only *discrete* 3D displacement. The velocity information can only be obtained approximately from the logarithm map (the inverse of the exponential map), as Soatto et al. did in (Soatto and Perona, 1996). In principle, displacement estimation algorithms obtained by using epipolar constraint work well when the displacement (especially the translation, i.e., the so called base-line) between the two images is relatively large. However, in real-time applications, even if the velocity of the moving camera is not small, the relative displacement between two consecutive images might become small owing to a high frame rate. In turn, the algorithms become singular due to the small translation and the estimation results become less reliable. Further, in applications such as robotic control, an on-board camera, as a feedback sensor, is required not only to provide relative orientation of the robot but also its relative speed (for control purposes).

A differential (or continuous) version of the 3D motion estimation problem is to recover the 3D velocity of the camera from optical flow. This problem has also been explored by many researchers: an algorithm was proposed in 1984 by Zhuang et al. (1984) with a simplified version given in 1986 (Zhuang et al., 1988); and a first order algorithm was given by Waxman et al. (1987) in 1987. Most algorithms start from the basic bilinear constraint relating optical flow to the linear and angular velocities and solve for rotation and translation separately using either numerical optimization techniques (Bruss and Horn, 1983) or linear subspace methods (Heeger and Jepson, 1992; Jepson and Heeger, 1993). Kanatani (1993a) proposed a linear

algorithm reformulating Zhuang's approach in terms of essential parameters and twisted flow. However, in these algorithms, the similarities between the discrete case and the differential case are not fully revealed and exploited.

In this paper, we develop, in parallel to the discrete essential matrix approach, a *differential essential matrix approach* for recovering 3D velocity from optical flow. Based on the differential version of the epipolar constraint, so called *differential essential matrices* are defined. We then give a complete characterization of the space of these matrices and prove that there exists exactly one 3D velocity corresponding to a given differential essential matrix. As a differential counterpart of the three-step SVD-based 3D displacement estimation algorithm, a four-step eigenvector-decomposition-based 3D velocity estimation algorithm is proposed.

One must note that, in this paper, only linear algorithms are studied and compared. It is well-known that linear algorithms are not optimal and give severely biased estimates when the noise level is high. In order to obtain optimal and unbiased estimates, nonlinear search schemes have to be used to solve for maximum likelihood estimates. In the sequel of this paper (Ma et al., 1998c), we have proposed an intrinsic geometric optimization algorithm based on Riemannian optimization techniques on manifolds. However, since nonlinear algorithms are only locally convergent, the linear algorithms studied in this paper can be used to initialize the search process of nonlinear algorithms. Further more, due to their geometric simplicity, clearly understanding the linear algorithms certainly helps in developing and understanding more sophisticated motion estimation schemes. For example, it can be shown that under the same conditions that the linear algorithms have a unique solution the nonlinear algorithms have quadratic rate of convergence (Ma et al., 1998c).

One of the big advantages of the differential approach is easy to exploit the *nonholonomic constraints* of a mobile base where the camera is mounted. In this paper, we show by example that nonholonomic constraints may reduce the number of dimensions of the motion estimation problem, hence reduce the number of minimum image measurements needed for a unique solution. The proposed motion estimation algorithm can thus be dramatically simplified. The differential approach developed in this paper can also be generalized to the case of an uncalibrated camera (Vieville and Faugeras, 1995; Brooks et al., 1995). Finally, in Section 3, simulation results are presented evaluating

the performance of our algorithm in terms of bias and sensitivity of the estimates with respect to the noise in optical flow measurements.

## 2. Differential Essential Matrix Approach

We first introduce some notation which will be frequently used in this paper. Given a vector  $p = (p_1, p_2, p_3)^T \in \mathbb{R}^3$ , we define  $\hat{p} \in so(3)$  (the space of skew symmetric matrices in  $\mathbb{R}^{3 \times 3}$ ) by:

$$\hat{p} = \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix}. \quad (1)$$

It then follows from the definition of cross-product of vectors that, for any two vectors  $p, q \in \mathbb{R}^3$ :  $p \times q = \hat{p}q$ .

Camera motion is modeled as rigid body motion in  $\mathbb{R}^3$ . The displacement of the camera belongs to the special Euclidean group  $SE(3)$ :

$$SE(3) = \{(p, R) : p \in \mathbb{R}^3, R \in SO(3)\} \quad (2)$$

where  $SO(3) \in \mathbb{R}^{3 \times 3}$  is the space of rotation matrices (orthogonal matrices with determinant +1). An element  $g = (p, R)$  in this group is used to represent the 3D translation and orientation (the displacement) of a coordinate frame  $F_c$  attached to the camera relative to an inertial frame which is chosen here as the initial position of the camera frame  $F_0$  (see Fig. 1). By an abuse of notation, the element  $g = (p, R)$  serves as both a specification of the configuration of the camera and as a transformation taking the coordinates of a point from  $F_c$  to  $F_0$ . More precisely, let  $q_0, q_c \in \mathbb{R}^3$  be the coordinates of a point  $q$  relative to frames  $F_0$  and  $F_c$ .

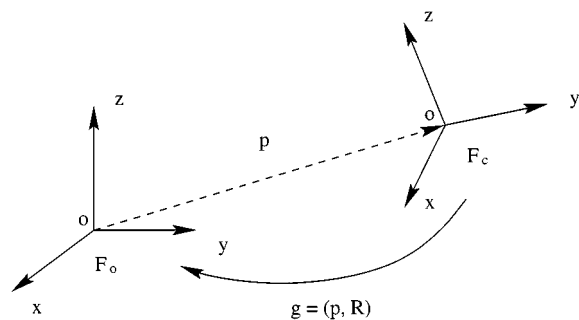


Figure 1. Coordinate frames for specifying rigid body motion of a camera.

respectively. Then the coordinate transformation between  $q_0$  and  $q_c$  is given by:

$$q_0 = Rq_c + p. \quad (3)$$

Assume that the camera frame is chosen such that the optical center of the camera, denoted by  $o$ , is the same as the origin of the frame. Then the image of a point  $q$  in the scene is the point where the ray  $\langle o, q \rangle$  intersects the imaging surface. A sphere or a plane is usually used to model the imaging surface. The model of image formation is then referred as *spherical projection* and *perspective projection*, respectively.

In this paper, we use bold letters to denote quantities associated with the image. The image of a point  $q \in \mathbb{R}^3$  in the scene is then denoted by  $\mathbf{q} \in \mathbb{R}^3$ . For the spherical projection, we simply choose the imaging surface to be the unit sphere:  $S^2 = \{q \in \mathbb{R}^3 \mid \|q\| = 1\}$  where the norm  $\|\cdot\|$  always means 2-norm unless otherwise stated. Then the spherical projection is defined by the map  $\pi_s$  from  $\mathbb{R}^3$  to  $S^2$ :

$$\pi_s : \mathbb{R}^3 \rightarrow S^2, \quad q \mapsto \mathbf{q} = \frac{q}{\|q\|}.$$

For the perspective projection, we choose the imaging surface to be the plane of unit distance away from the optical center. The perspective projection onto this plane is then defined by the map  $\pi_p$  from  $\mathbb{R}^3$  to the projective plane  $\mathbb{RP}^2$ :

$$\pi_p : \mathbb{R}^3 \rightarrow \mathbb{RP}^2, \\ q = (q_1, q_2, q_3)^T \mapsto \mathbf{q} = \begin{pmatrix} q_1 & q_2 & 1 \\ q_3 & q_3 & \end{pmatrix}^T.$$

The essential approach taken in this paper only exploits the intrinsic geometric relations which are preserved by both projection models. Thus, theorems and algorithms to be developed are always true for both cases, unless otherwise stated. By an abuse of notation, we will simply denote both  $\pi_s$  and  $\pi_p$  by the same letter  $\pi$ . The image of the point  $q$  taken by the camera is then  $\mathbf{q} = \pi(q)$ .

### 2.1. Review of the Discrete Essential Matrix Approach

Before developing the analysis of the differential epipolar constraint which is the main focus of this paper, we first provide a brief review of the epipolar geometry in

the discrete case, also known as the *essential matrix approach*, originally developed by Huang and Faugeras (1989). Let the 3D displacement of the frame  $F_c$  relative to the frame  $F_0$  be given by the rigid body motion  $g = (p, R) \in SE(3)$ , and let  $\mathbf{q}_0, \mathbf{q}_c$  be the images of the same point  $q$  taken by the camera at frames  $F_0$  and  $F_c$ , respectively, then it is well-known that the two image points  $\mathbf{q}_0, \mathbf{q}_c$  satisfy the so called *epipolar constraint*:

$$\mathbf{q}_c^T R^T \hat{p} \mathbf{q}_0 = 0. \quad (4)$$

In this equation we see that the matrix  $E = R^T \hat{p}$  with  $R^T \in SO(3)$  and  $\hat{p} \in so(3)$  contains the unknown motion parameters. A matrix of this form is called an *essential matrix*; and the set of all essential matrices is called the *essential space*, denoted by  $\mathcal{E}$ :

$$\mathcal{E} \equiv \{RS \mid R \in SO(3), S \in so(3)\} \subset \mathbb{R}^{3 \times 3}. \quad (5)$$

Huang and Faugeras (1989) established that a non-zero matrix  $E$  is an essential matrix if and only if the singular value decomposition (SVD) of  $E : E = U \Sigma V^T$  satisfies:

$$\Sigma = \text{diag}\{\lambda, \lambda, 0\} \quad (6)$$

for some  $\lambda \in \mathbb{R}_+$ . In order to answer the question: given an essential matrix  $E \in \mathcal{E}$ , how many pairs  $(p, R)$  exist such that  $R^T \hat{p} = E$ , we first give the following lemma from linear algebra:

**Lemma 1.** *Given any non-zero skew symmetric matrix  $S \in so(3)$ , if, for a rotation matrix  $R \in SO(3)$ ,  $RS$  is also a skew symmetric matrix, then  $R = I$  or  $e^{\hat{p}\pi}$  where  $\hat{p}$  is the unit skew symmetric matrix associated with  $S$ . Further,  $e^{\hat{p}\pi} S = -S$ .*

**Proof:** Without loss of generality, we assume  $S$  is a unit skew symmetric matrix, i.e., there exists a vector  $p \in \mathbb{R}^3$  of unit length such that  $\hat{p} = S$ . Since  $RS$  is also a skew symmetric matrix,  $(RS)^T = -RS$ . This equation gives:

$$R\hat{p}R = \hat{p}. \quad (7)$$

Since  $R$  is a rotation matrix, there exists  $\omega \in \mathbb{R}^3$ ,  $\|\omega\| = 1$  and  $\theta \in \mathbb{R}$  such that  $R = e^{\hat{\omega}\theta}$ . Then, (7) is rewritten as:  $e^{\hat{\omega}\theta} \hat{p} e^{\hat{\omega}\theta} = \hat{p}$ . Applying this equation to  $\omega$ , we get:  $e^{\hat{\omega}\theta} \hat{p} e^{\hat{\omega}\theta} \omega = \hat{p}\omega$ . Since  $e^{\hat{\omega}\theta} \omega = \omega$ , we obtain:  $e^{\hat{\omega}\theta} \hat{p}\omega = \hat{p}\omega$ . Since  $\omega$  is the only eigenvector associated to the eigenvalue 1 of the matrix  $e^{\hat{\omega}\theta}$  and  $\hat{p}\omega$  is orthogonal to  $\omega$ ,  $\hat{p}\omega$  has to be zero. Thus,  $\omega$  is equal

to  $p$  or  $-p$ .  $R$  then has the form  $e^{\hat{p}\theta}$ , which commutes with  $\hat{p}$ . Thus from (7), we get:

$$e^{2\hat{p}\theta} \hat{p} = \hat{p}. \quad (8)$$

According to *Rodrigues' formula* (Murray et al., 1994), we have:

$$e^{2\hat{p}\theta} = I + \hat{p} \sin(2\theta) + \hat{p}^2 (1 - \cos(2\theta)) \quad (9)$$

(8) yields:

$$\hat{p}^2 \sin(2\theta) + \hat{p}^3 (1 - \cos(2\theta)) = 0. \quad (10)$$

Since  $\hat{p}^2$  and  $\hat{p}^3$  are linearly independent (Lemma 2.3 in Murray et al. (1994), we have  $\sin(2\theta) = 1 - \cos(2\theta) = 0$ . That is,  $\theta$  is equal to  $2k\pi$  or  $2k\pi + \pi$ ,  $k \in \mathbb{Z}$ . Therefore,  $R$  is equal to  $I$  or  $e^{\hat{p}\pi}$ . It is direct from the geometric meaning of  $e^{\hat{p}\pi} \hat{p}$  that  $e^{\hat{p}\pi} \hat{p} = -\hat{p}$ , thus  $e^{\hat{p}\pi} S = -S$ .  $\square$

Following this lemma, suppose  $(p_1, R_1) \in SE(3)$  and  $(p_2, R_2) \in SE(3)$  are both solutions for the equation  $R^T \hat{p} = E$ . Then we have  $R_1^T \hat{p}_1 = R_2^T \hat{p}_2$ . It yields  $R_2 R_1^T \hat{p}_1 = \hat{p}_2$ . Since  $\hat{p}_1, \hat{p}_2$  are skew symmetric matrices and  $R_2 R_1^T$  is a rotation matrix, we then have either  $(p_2, R_2) = (p_1, R_1)$  or  $(p_2, R_2) = (-p_1, e^{\hat{p}_1\pi} R_1)$ . Therefore, given an essential matrix  $E$  there are exactly *two* pairs  $(p, R)$  such that  $R^T \hat{p} = E$ . Further, if  $E$  has the SVD:  $E = U \Sigma V^T$  with  $U, V \in SO(3)$ ,<sup>1</sup> the following formulae give the two solutions:

$$\begin{cases} (R_1^T, \hat{p}_1) = \left( UR_Z^T \left( +\frac{\pi}{2} \right) V^T, VR_Z \left( +\frac{\pi}{2} \right) \Sigma V^T \right), \\ (R_2^T, \hat{p}_2) = \left( UR_Z^T \left( -\frac{\pi}{2} \right) V^T, VR_Z \left( -\frac{\pi}{2} \right) \Sigma V^T \right) \end{cases} \quad (11)$$

where  $R_Z(\theta)$  is defined to be the rotation matrix around the  $Z$ -axis by an angle  $\theta$ , i.e.,  $R_Z(\theta) = e^{\hat{e}_3\theta}$  with  $e_3 = (0, 0, 1)^T \in \mathbb{R}^3$ .

Since from the epipolar constraint (4) one can only recover the essential matrix up to an arbitrary scale (in particular, both  $E$  and  $-E$  satisfy the same equation), so in general four solutions  $(p, R)$  will be obtained from image correspondences. Usually, the *positive depth constraint* can be imposed to discard three of the ambiguous solutions. We here omit these well known details and simply summarize the discrete essential matrix approach for motion estimation as the following algorithm (which is essentially the same as

that given in Maybank (1993) and we repeat it here for comparison with the algorithm that we will develop for the differential case:

### Three-Step SVD-Based 3D Displacement Estimation Algorithm:

1. *Estimate the essential matrix:*

For a given set of image correspondences:  $(\mathbf{q}_c^i, \mathbf{q}_0^i)$ ,  $i = 1, \dots, m$  ( $m \geq 8$ ), find the matrix  $E$  which minimizes the error function:

$$V(E) = \sum_{i=1}^m (\mathbf{q}_c^{iT} R^T \hat{p} \mathbf{q}_0^i)^2 \quad (12)$$

subject to the condition  $\|E\| = 1$ ;

2. *Singular value decomposition:*

Recover matrix  $E$  from  $\mathbf{e}$  and find the singular value decomposition of the matrix  $E$ :

$$E = U \text{diag}\{\sigma_1, \sigma_2, \sigma_3\} V^T \quad (13)$$

where  $\sigma_1 \geq \sigma_2 \geq \sigma_3$ ;

3. *Recover displacement from the essential matrix:*

Define the diagonal matrix  $\Sigma$  to be:

$$\Sigma = \text{diag}\{1, 1, 0\}. \quad (14)$$

Then the 3D displacement  $(p, R)$  is given by:

$$R^T = UR_Z^T \left( \pm \frac{\pi}{2} \right) V^T, \quad \hat{p} = VR_Z \left( \pm \frac{\pi}{2} \right) \Sigma V^T. \quad (15)$$

The epipolar geometric relationship between projections of the points and their displacements transfers to the differential case. So, intuitively speaking, the differential case is an infinitesimal version of the discrete case. However, the differential case is by no means simply a ‘‘first order approximation’’ of the discrete case. When differentiation takes place, while structure of the geometry of the discrete case is inherited by the differential case, some degeneracies may occur. Such degeneracies will become clear when we study the differential version of the epipolar constraint. It is also known that it is exactly due to these degeneracies that camera calibration cannot be fully recovered from differential epipolar constraint as opposed to the discrete case (Ma et al., 1998b). Generally speaking, the similarity between these two cases is that methods and geometric intuition used in the discrete case can be

extended to the differential case, even though geometric characterization of the objects is different. One of the main goals of this paper is to clarify the geometric *similarity* and *difference* between the discrete and differential cases. Although the theory will be developed in a calibrated camera framework, the clear geometric nature of this approach has helped us to understand the uncalibrated situation as well—especially the relation between the Kruppa’s equation and its differential version (Ma et al., 1998b).

### 2.2. Differential Epipolar Constraint

We now develop a *differential essential matrix approach* for estimating 3D velocity from optical flow in a parallel way to the discrete essential matrix approach for estimating 3D displacement from image correspondences.

The starting point of this approach is a differential version of the epipolar constraint and associated concept of differential essential matrix. This constraint is bilinear in nature and has been used extensively in the motion estimation from optical flow measurements (Vieville and Faugeras, 1995; Heeger and Jepson, 1992). Here we give a characterization of such matrices and show that there exists exactly one 3D velocity corresponding to a non-zero differential essential matrix; as a differential version of the three-step SVD-based 3D displacement estimation algorithm, we propose a four-step eigenvector-decomposition-based 3D velocity estimation algorithm; finally, we discuss the reasons why the zero-translation case makes all essential constraint based motion estimation algorithms fail and suggest possible ways to overcome this difficulty.

Assume that camera motion is described by a smooth curve  $g(t) = (p(t), R(t)) \in SE(3)$ . According to (3), for a point  $q$  attached to the inertial frame  $F_0$ , its coordinates in the inertial frame and the moving camera frame satisfy:

$$q_0 = R(t)q_c(t) + p(t). \quad (16)$$

Differentiating this equation yields:

$$\dot{q}_c = -R^T \dot{R} q_c - R^T \dot{p}. \quad (17)$$

Since  $-R^T \dot{R} \in so(3)$  and  $-R^T \dot{p} \in \mathbb{R}^3$  (see Murray et al. (1994)), we may define  $\omega = (\omega_1, \omega_2, \omega_3)^T \in \mathbb{R}^3$

and  $v = (v_1, v_2, v_3)^T \in \mathbb{R}^3$  to be:

$$\hat{\omega} = -R^T \dot{R}, \quad v = -R^T \dot{p}. \quad (18)$$

The interpretation of these velocities is:  $-\omega$  is the angular velocity of the camera frame  $F_c$  relative to the inertial frame  $F_i$  and  $-v$  is the velocity of the origin of the camera frame  $F_c$  relative to the inertial frame  $F_i$ . Using the new notation, we get:

$$\dot{q}_c = \hat{\omega} q_c + v. \quad (19)$$

From now on, for convenience we will drop the subscript  $c$  from  $q_c$ . The notation  $q$  then serves both as a point fixed in the frame and its coordinates in the current camera frame  $F_c$ . The image of the point  $q$  taken by the camera is given by the spherical projection:  $\mathbf{q} = \pi(q)$ . Denote the velocity of the image point  $\mathbf{q}$ , the so called optical flow, by  $\mathbf{u}$ , i.e.,  $\mathbf{u} = \dot{\mathbf{q}} \in \mathbb{R}^3$ .

**Theorem 1** (Differential epipolar constraint). *Consider a camera moving with linear velocity  $v$  and angular velocity  $\omega$  with respect to the inertial frame. Then the optical flow  $\mathbf{u}$  at an image point  $\mathbf{q}$  satisfies:*

$$\mathbf{u}^T \hat{v} \mathbf{q} + \mathbf{q}^T \hat{\omega} \hat{v} \mathbf{q} \equiv 0 \quad (20)$$

or in an equivalent form:

$$(\mathbf{u}^T, \mathbf{q}^T) \begin{pmatrix} \hat{v} \\ s \end{pmatrix} \mathbf{q} = 0 \quad (21)$$

where  $s$  is a symmetric matrix defined to be  $s = \frac{1}{2}(\hat{\omega} \hat{v} + \hat{v} \hat{\omega}) \in \mathbb{R}^{3 \times 3}$ .

**Proof:** From the definition of the maps  $\pi$ , there exists a real scalar function  $\lambda(t)$  ( $\|q(t)\|$  or  $q_3(t)$ , depending on the type of projection) such that  $q = \lambda \mathbf{q}$ . Take the inner product of the vectors in (19) with  $(v \times \mathbf{q})$ :

$$\dot{q}^T (v \times \mathbf{q}) = (\hat{\omega} q + v)^T (v \times \mathbf{q}) = q^T \hat{\omega}^T \hat{v} \mathbf{q}. \quad (22)$$

Since  $\dot{q} = \dot{\lambda} \mathbf{q} + \lambda \dot{\mathbf{q}}$  and  $q^T (v \times \mathbf{q}) = 0$ , from (22) we then have:

$$\lambda \dot{\mathbf{q}}^T \hat{v} \mathbf{q} - \lambda \mathbf{q}^T \hat{\omega}^T \hat{v} \mathbf{q} = 0. \quad (23)$$

When  $\lambda \neq 0$ , we obtain a differential version of the epipolar constraint:

$$\mathbf{u}^T \hat{v} \mathbf{q} + \mathbf{q}^T \hat{\omega} \hat{v} \mathbf{q} \equiv 0 \quad (24)$$

Due to the following fact 1, for any skew symmetric matrix  $A \in \mathbb{R}^{3 \times 3}$ ,  $\mathbf{q}^T A \mathbf{q} = 0$ . Since  $\frac{1}{2}(\hat{\omega} \hat{v} - \hat{v} \hat{\omega})$  is a skew symmetric matrix,  $\mathbf{q}^T \frac{1}{2}(\hat{\omega} \hat{v} - \hat{v} \hat{\omega}) \mathbf{q} = \mathbf{q}^T s \mathbf{q} - \mathbf{q}^T \hat{\omega} \hat{v} \mathbf{q} = 0$ . Thus,  $\mathbf{q}^T s \mathbf{q} = \mathbf{q}^T \hat{\omega} \hat{v} \mathbf{q}$ . We then have:

$$\mathbf{u}^T \hat{v} \mathbf{q} + \mathbf{q}^T s \mathbf{q} \equiv 0. \quad (25)$$

□

The proof indicates that there is some redundancy in the expression of the differential epipolar constraint (20). The following fact from linear algebra shows where this redundancy comes from.

**Fact 1.** *Consider matrices  $M_1, M_2 \in \mathbb{R}^{3 \times 3}$ .  $\mathbf{q}^T M_1 \mathbf{q} = \mathbf{q}^T M_2 \mathbf{q}$  for all  $\mathbf{q} \in \mathbb{R}^3$  if and only if  $M_1 - M_2$  is a skew symmetric matrix, i.e.,  $M_1 - M_2 \in so(3)$ .*

Let us define an equivalence relation on the space  $\mathbb{R}^{3 \times 3}$ , the space of  $3 \times 3$  matrices over  $\mathbb{R}$ : for  $x, y \in \mathbb{R}^{3 \times 3}$ ,  $x \sim y$  if and only if  $x - y \in so(3)$ . Denote by  $[x] = \{y \in \mathbb{R}^{3 \times 3} \mid y \sim x\}$  the equivalence class of  $x$ , and denote by  $[X]$  the set  $\bigcup_{x \in X} [x]$ . The quotient space  $\mathbb{R}^{3 \times 3} / \sim$  can be naturally identified with the space of all  $3 \times 3$  symmetric matrices. Especially, we have  $s = \frac{1}{2}(\hat{\omega} \hat{v} + \hat{v} \hat{\omega}) \in [\hat{\omega} \hat{v}]$ , which is the reason why we choose it in the equivalent form (21). Using this notation, Theorem 1 can then be re-expressed in the following way:

**Corollary 1.** *Consider a camera undergoing a smooth rigid body motion with linear velocity  $v$  and angular velocity  $\omega$ . Then the optical flow  $\mathbf{u}$  of a image point  $\mathbf{q}$  satisfies:*

$$(\mathbf{u}^T, \mathbf{q}^T) \begin{pmatrix} \hat{v} \\ [\hat{\omega} \hat{v}] \end{pmatrix} \mathbf{q} \equiv 0. \quad (26)$$

Because of this redundancy, each equivalence class  $[\hat{\omega} \hat{v}]$  can only be recovered up to its symmetric component  $s = \frac{1}{2}(\hat{\omega} \hat{v} + \hat{v} \hat{\omega}) \in [\hat{\omega} \hat{v}]$ . This redundancy is the exact reason why different forms of the differential epipolar constraint exist in the literature (Zhuang and Haralick, 1984; Ponce and Genc, 1998; Vieville and Faugeras, 1995; Maybank, 1993; Brooks et al., in press), and, accordingly, various approaches have been proposed to recover  $\omega$  and  $v$  (see Tian et al. (1996)). It is also the reason why the differential case cannot be simply viewed as a first order approximation of the discrete case—a first order approximation of the essential matrix  $R^T \hat{p}$  is  $\hat{\omega} \hat{v}$ , but this is certainly not what one can directly estimate from the differential epipolar

constraint. Instead, one has to deal with its symmetric part  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . This, in fact, makes the study of the differential case harder than the discrete case (in seek for linear algorithms). Notice that the symmetric matrix  $s$  is the same as the matrix  $K$  defined in Kanatani (1993b). Although the characterization of such matrices has been studied in Kanatani (1993b), our constructive proofs given below will lead to a natural algorithm for recovering  $(\omega, v)$  from  $s$ .

### 2.3. Characterization of the Differential Essential Matrix

We define the space of  $6 \times 3$  matrices given by:

$$\mathcal{E}' = \left\{ \left( \begin{array}{c} \hat{v} \\ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \end{array} \right) \middle| \omega, v \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{6 \times 3}. \quad (27)$$

to be the *differential essential space*. A matrix in this space is called a *differential essential matrix*. Note that the differential epipolar constraint (21) is homogeneous on the linear velocity  $v$ . Thus  $v$  may be recovered only up to a constant scale. Consequently, in motion recovery, we will concern ourselves with matrices belonging to *normalized differential essential space*:

$$\mathcal{E}'_1 = \left\{ \left( \begin{array}{c} \hat{v} \\ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \end{array} \right) \middle| \omega \in \mathbb{R}^3, v \in S^2 \right\} \subset \mathbb{R}^{6 \times 3}. \quad (28)$$

The skew-symmetric part of a differential essential matrix simply corresponds to the velocity  $v$ . The characterization of the (normalized) essential matrix only focuses on the characterization of the symmetric part of the matrix:  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . We call the space of all the matrices of such form the *special symmetric space*:

$$\mathcal{S} = \left\{ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \middle| \omega \in \mathbb{R}^3, v \in S^2 \right\} \subset \mathbb{R}^{3 \times 3}. \quad (29)$$

A matrix in this space is called a *special symmetric matrix*. The motion estimation problem is now reduced to the one of *recovering the velocity*  $(\omega, v)$  with  $\omega \in \mathbb{R}^3$  and  $v \in S^2$  from a given special symmetric matrix  $s$ .

The characterization of special symmetric matrices depends on a characterization of matrices in the form:  $\hat{\omega}\hat{v} \in \mathbb{R}^{3 \times 3}$ , which is given in the following lemma. This lemma will also be used in the next section for showing the uniqueness of the velocity recovery from special symmetric matrices. Like the (discrete) essential matrices, matrices with the form  $\hat{\omega}\hat{v}$  are characterized

by their singular value decomposition (SVD):  $\hat{\omega}\hat{v} = U\Sigma V^T$ ; moreover, the orthogonal matrices  $U$  and  $V$  are related. Define the matrix  $R_Y(\theta)$  to be the rotation around the  $Y$ -axis by an angle  $\theta \in \mathbb{R}$ , i.e.,  $R_Y(\theta) = e^{\hat{e}_2\theta}$  with  $e_2 = (0, 1, 0)^T \in \mathbb{R}^3$ .

**Lemma 2.** *A matrix  $Q \in \mathbb{R}^{3 \times 3}$  has the form  $Q = \hat{\omega}\hat{v}$  with  $\omega \in \mathbb{R}^3$ ,  $v \in S^2$  if and only if  $Q$  has the form:*

$$Q = -VR_Y(\theta) \text{diag}\{\lambda, \lambda \cos(\theta), 0\}V^T \quad (30)$$

for some rotation matrix  $V \in SO(3)$ . Further,  $\lambda = \|\omega\|$  and  $\cos(\theta) = \omega^T v / \lambda$ .

**Proof:** We first prove the necessity. The proof follows from the geometric meaning of  $\hat{\omega}\hat{v}$ : for any vector  $q \in \mathbb{R}^3$ ,

$$\hat{\omega}\hat{v}q = \omega \times (v \times q). \quad (31)$$

Let  $b \in S^2$  be the unit vector perpendicular to both  $\omega$  and  $v$ :  $b = \frac{v \times \omega}{\|v \times \omega\|}$  (if  $v \times \omega = 0$ ,  $b$  is not uniquely defined. In this case, pick any  $b$  orthogonal to  $v$  and  $\omega$ , then the rest of the proof still holds). Then  $\omega = \lambda \exp(\hat{b}\theta)v$  (according this definition,  $\theta$  is the angle between  $\omega$  and  $v$ , and  $0 \leq \theta \leq \pi$ ). It is direct to check that if the matrix  $V$  is defined to be:

$$V = (e^{\hat{b}\frac{\pi}{2}}v, b, v), \quad (32)$$

then  $Q$  has the given form (30).

We now prove the sufficiency. Given a matrix  $Q$  which can be decomposed into the form (30), define the orthogonal matrix  $U = -VR_Y(\theta) \in O(3)$ .<sup>2</sup> Let the two skew symmetric matrices  $\hat{\omega}$  and  $\hat{v}$  given by the formulae:

$$\hat{\omega} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T, \quad \hat{v} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \quad (33)$$

where  $\Sigma_\lambda = \text{diag}\{\lambda, \lambda, 0\}$  and  $\Sigma_1 = \text{diag}\{1, 1, 0\}$ . Then:

$$\begin{aligned} \hat{\omega}\hat{v} &= UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \\ &= UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda (-R_Y^T(\theta))R_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \\ &= U \text{diag}\{\lambda, \lambda \cos(\theta), 0\}V^T \\ &= Q. \end{aligned} \quad (34)$$

Since  $\omega$  and  $v$  have to be, respectively, the left and the right zero eigenvectors of  $Q$ , the reconstruction given in (33) is unique.  $\square$

The following theorem gives a characterization of the special symmetric matrix.

**Theorem 2** (Characterization of the special symmetric matrix). *A matrix  $s \in \mathbb{R}^{3 \times 3}$  is a special symmetric matrix if and only if  $s$  can be diagonalized as  $s = V \Sigma V^T$  with  $V \in SO(3)$  and:*

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \sigma_3\} \quad (35)$$

with  $\sigma_1 \geq 0$ ,  $\sigma_3 \leq 0$  and  $\sigma_2 = \sigma_1 + \sigma_3$ .

**Proof:** We first prove the necessity. Suppose  $s$  is a special symmetric matrix, there exist  $\omega \in \mathbb{R}^3$ ,  $v \in S^2$  such that  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . Since  $s$  is a symmetric matrix, it is diagonalizable, all its eigenvalues are real and all the eigenvectors are orthogonal to each other. It then suffices to check that its eigenvalues satisfy the given conditions.

Let the unit vector  $b$  and the rotation matrix  $V$  be the same as in the proof of Lemma 2, so are  $\theta$  and  $\gamma$ . Then according to the lemma, we have:

$$\hat{\omega}\hat{v} = -VR_Y(\theta) \text{diag}\{\lambda, \lambda \cos(\theta), 0\}V^T. \quad (36)$$

Since  $(\hat{\omega}\hat{v})^T = \hat{v}\hat{\omega}$ , it yields:

$$\begin{aligned} s &= \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \\ &= \frac{1}{2}V(-R_Y(\theta) \text{diag}\{\lambda, \lambda \cos(\theta), 0\} \\ &\quad - \text{diag}\{\lambda, \lambda \cos(\theta), 0\}R_Y^T(\theta))V^T. \end{aligned} \quad (37)$$

Define the matrix  $D(\lambda, \theta) \in \mathbb{R}^{3 \times 3}$  to be:

$$\begin{aligned} D(\lambda, \theta) &= -R_Y(\theta) \text{diag}\{\lambda, \lambda \cos(\theta), 0\} \\ &\quad - \text{diag}\{\lambda, \lambda \cos(\theta), 0\}R_Y^T(\theta) \\ &= \lambda \begin{pmatrix} -2 \cos(\theta) & 0 & \sin(\theta) \\ 0 & -2 \cos(\theta) & 0 \\ \sin(\theta) & 0 & 0 \end{pmatrix}. \end{aligned} \quad (38)$$

Directly calculating its eigenvalues and eigenvectors, we obtain that:

$$\begin{aligned} D(\lambda, \theta) &= R_Y\left(\frac{\theta}{2} - \frac{\pi}{2}\right) \\ &\quad \times \text{diag}\{\lambda(1 - \cos(\theta)), -2\lambda \cos(\theta), \\ &\quad \lambda(-1 - \cos(\theta))\}R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right). \end{aligned} \quad (39)$$

Thus  $s = \frac{1}{2}VD(\lambda, \theta)V^T$  has eigenvalues:

$$\left\{ \frac{1}{2}\lambda(1 - \cos(\theta)), -\lambda \cos(\theta), \frac{1}{2}\lambda(-1 - \cos(\theta)) \right\}, \quad (40)$$

which satisfy the given conditions.

We now prove the sufficiency. Given  $s = V_1 \text{diag}\{\sigma_1, \sigma_2, \sigma_3\}V_1^T$  with  $\sigma_1 \geq 0$ ,  $\sigma_3 \leq 0$  and  $\sigma_2 = \sigma_1 + \sigma_3$  and  $V_1^T \in SO(3)$ , these three eigenvalues uniquely determine  $\lambda, \theta \in \mathbb{R}$  such that the  $\sigma_i$ 's have the form given in (40):

$$\begin{cases} \lambda = \sigma_1 - \sigma_3, & \lambda \geq 0 \\ \theta = \arccos(-\sigma_2/\lambda), & \theta \in [0, \pi] \end{cases}$$

Define a matrix  $V \in SO(3)$  to be  $V = V_1 R_Y^T(\frac{\theta}{2} - \frac{\pi}{2})$ . Then  $s = \frac{1}{2}VD(\lambda, \theta)V^T$ . According to Lemma 2, there exist vectors  $v \in S^2$  and  $\omega \in \mathbb{R}^3$  such that:

$$\hat{\omega}\hat{v} = -VR_Y(\theta) \text{diag}\{\lambda, \lambda \cos(\theta), 0\}V^T. \quad (41)$$

Therefore,  $\frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) = \frac{1}{2}VD(\lambda, \theta)V^T = s$ .  $\square$

Figure 2 gives a geometric interpretation of the three eigenvectors of the special symmetric matrix  $s$  for the case when both  $\omega, v$  are of unit length. Theorem 2 was given as an exercise problem in Kanatani (1993a) but it has never been really exploited in the literature for designing algorithms. For that purpose, the constructive proof given above is more important since it gives an explicit decomposition of the special symmetric matrix  $s$ , which will be studied in more detail next.

According to the proof of the sufficiency of Theorem 2, if we already know the eigenvector decomposition of a special symmetric matrix  $s$ , we certainly can find at least one solution  $(\omega, v)$  such that  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . This section discusses the uniqueness of such reconstruction, i.e., how many solutions exist for  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ .



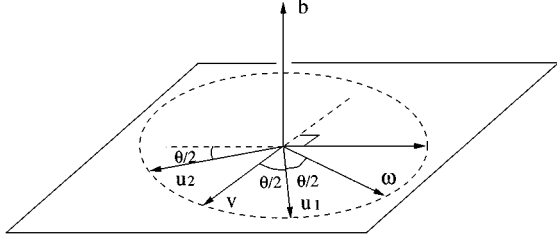


Figure 2. Vectors  $u_1, u_2, b$  are the three eigenvectors of a special symmetric matrix  $\frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . In particular,  $b$  is the normal vector to the plane spanned by  $\omega$  and  $v$ , and  $u_1, u_2$  are both in this plane.  $u_1$  is the average of  $\omega$  and  $v$ .  $u_2$  is orthogonal to both  $b$  and  $u_1$ .

**Theorem 3** (Uniqueness of the velocity recovery from the special symmetric matrix). *There exist exactly four 3D velocities  $(\omega, v)$  with  $\omega \in \mathbb{R}^3$  and  $v \in S^2$  corresponding to a non-zero special symmetric matrix  $s \in \mathcal{S}$ .*

**Proof:** Suppose  $(\omega_1, v_1)$  and  $(\omega_2, v_2)$  are both solutions for  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ . Then we have:

$$\hat{v}_1\hat{\omega}_1 + \hat{\omega}_1\hat{v}_1 = \hat{v}_2\hat{\omega}_2 + \hat{\omega}_2\hat{v}_2. \quad (42)$$

From Lemma 2, we may write:

$$\begin{cases} \hat{\omega}_1\hat{v}_1 = -V_1 R_Y(\theta_1) \text{diag}\{\lambda_1, \lambda_1 \cos(\theta_1), 0\} V_1^T \\ \hat{\omega}_2\hat{v}_2 = -V_2 R_Y(\theta_2) \text{diag}\{\lambda_2, \lambda_2 \cos(\theta_2), 0\} V_2^T. \end{cases} \quad (43)$$

Let  $W = V_1^T V_2 \in SO(3)$ , then from (42):

$$D(\lambda_1, \theta_1) = W D(\lambda_2, \theta_2) W^T. \quad (44)$$

Since both sides of (44) have the same eigenvalues, according to (39), we have:

$$\lambda_1 = \lambda_2, \quad \theta_2 = \theta_1. \quad (45)$$

We then can denote both  $\theta_1$  and  $\theta_2$  by  $\theta$ . It is direct to check that the only possible rotation matrix  $W$  which satisfies (44) is given by  $I_{3 \times 3}$  or:

$$\begin{pmatrix} -\cos(\theta) & 0 & \sin(\theta) \\ 0 & -1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & -1 & 0 \\ -\sin(\theta) & 0 & -\cos(\theta) \end{pmatrix}. \quad (46)$$

From the geometric meaning of  $V_1$  and  $V_2$ , all the cases give either  $\hat{\omega}_1\hat{v}_1 = \hat{\omega}_2\hat{v}_2$  or  $\hat{\omega}_1\hat{v}_1 = \hat{v}_2\hat{\omega}_2$ . Thus, according to the proof of Lemma 2, if  $(\omega, v)$  is one solution and  $\hat{\omega}\hat{v} = U \text{diag}\{\lambda, \lambda \cos(\theta), 0\} V^T$ , then all the solutions are given by:

$$\begin{cases} \hat{\omega} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T, & \hat{v} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T; \\ \hat{\omega} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda V^T, & \hat{v} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 U^T \end{cases} \quad (47)$$

where  $\Sigma_\lambda = \text{diag}\{\lambda, \lambda, 0\}$  and  $\Sigma_1 = \text{diag}\{1, 1, 0\}$ .  $\square$

Given a non-zero differential essential matrix  $E \in \mathcal{E}'$ , according to (47) its special symmetric part gives four possible solutions for the 3D velocity  $(\omega, v)$ . However, in general only one of them has the same linear velocity  $v$  as the skew symmetric part of  $E$  does. We thus have:

**Theorem 4** (Uniqueness of the velocity recovery from differential essential matrix). *There exists only one 3D velocity  $(\omega, v)$  with  $\omega \in \mathbb{R}^3$  and  $v \in \mathbb{R}^3$  corresponding to a non-zero differential essential matrix  $E \in \mathcal{E}'$ .*

In the discrete case, there are two 3D displacements corresponding to an essential matrix. However, the velocity corresponding to a differential essential matrix is unique. This is because, in the differential case, the twisted-pair ambiguity (see Maybank (1993)), which is caused by a  $180^\circ$  rotation of the camera around the translation direction, is avoided.

## 2.4. Algorithm

Based on the preceding study of the differential essential matrix, we propose an new algorithm which recovers the 3D velocity of the camera from a set of (possibly noisy) optical flows.

Let  $E = \begin{pmatrix} \hat{v} \\ s \end{pmatrix} \in \mathcal{E}'_1$  with  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$  be the essential matrix associated with the differential epipolar constraint (21). Since the submatrix  $\hat{v}$  is skew symmetric and  $s$  is symmetric, they have the following form:

$$v = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}, \quad s = \begin{pmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{pmatrix}. \quad (48)$$

Define the (differential) *essential vector*  $\mathbf{e} \in \mathbb{R}^9$  to be:

$$\mathbf{e} = (v_1, v_2, v_3, s_1, s_2, s_3, s_4, s_5, s_6)^T. \quad (49)$$

Define a vector  $\mathbf{a} \in \mathbb{R}^9$  associated to optical flow  $(\mathbf{q}, \mathbf{u})$  with  $\mathbf{q} = (x, y, z)^T \in \mathbb{R}^3$ ,  $\mathbf{u} = (u_1, u_2, u_3)^T \in \mathbb{R}^3$  to be<sup>3</sup>:

$$\mathbf{a} = (u_3y - u_2z, u_1z - u_3x, u_2x - u_1y, x^2, 2xy, 2xz, y^2, 2yz, z^2)^T. \quad (50)$$

The differential epipolar constraint (21) can be then rewritten as:

$$\mathbf{a}^T \mathbf{e} = 0. \quad (51)$$

Given a set of (possibly noisy) optical flow vectors:  $(\mathbf{q}^i, \mathbf{u}^i)$ ,  $i = 1, \dots, m$  generated by the same motion, define a matrix  $A \in \mathbb{R}^{m \times 9}$  associated to these measurements to be:

$$A = (\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^m)^T \quad (52)$$

where  $\mathbf{a}^i$  are defined for each pair  $(\mathbf{q}^i, \mathbf{u}^i)$  using (50). In the absence of noise, the essential vector  $\mathbf{e}$  has to satisfy:

$$A\mathbf{e} = 0. \quad (53)$$

In order for this equation to have a unique solution for  $\mathbf{e}$ , the rank of the matrix  $A$  has to be eight. Thus, *for this algorithm, in general, the optical flow vectors of at least eight points are needed to recover the 3D velocity, i.e.,  $m \geq 8$* , although the minimum number of optical flows needed is 5 (see Maybank (1993)). When the measurements are noisy, there might be no solution of  $\mathbf{e}$  for  $A\mathbf{e} = 0$ . As in the discrete case (Maybank, 1993), we choose the solution which minimizes the error function  $\|A\mathbf{e}\|^2$ .

Since the differential essential vector  $\mathbf{e}$  is recovered from noisy measurements, the symmetric part  $s$  of  $E$  directly recovered from  $\mathbf{e}$  is not necessarily a special symmetric matrix. Thus one can not directly use the previously derived results for special symmetric matrices to recover the 3D velocity. In the algorithms proposed in Zhuang and Haralick (1984) and Zhuang et al. (1988), such  $s$ , with the linear velocity  $v$  obtained from the skew-symmetric part, is directly used to calculate the angular velocity  $\omega$ . This is an over-determined problem since three variables are to be determined from six independent equations; on the other hand, erroneous  $v$

introduces further error in the estimation of the angular velocity  $\omega$ .

We thus propose a different approach: first extract the special symmetric component from the symmetric matrix  $s$  directly estimated from the differential epipolar constraint; then recover the four possible solutions for the 3D velocity using the results obtained in Theorem 3; finally choose the one which has the closest linear velocity to the one given by the skew-symmetric part of  $E$ . In order to extract the special symmetric component out of a symmetric matrix, we need a projection from the space of all symmetric matrices to the special symmetric space  $\mathcal{S}$ , i.e., a differential version of the projection of a matrix to the essential manifold  $\mathcal{E}$  given in Maybank (1993).

**Theorem 5** (Projection to the special symmetric space).

*If a symmetric matrix  $F \in \mathbb{R}^{3 \times 3}$  is diagonalized as  $F = V \text{diag}\{\lambda_1, \lambda_2, \lambda_3\} V^T$  with  $V \in SO(3)$ ,  $\lambda_1 \geq 0$ ,  $\lambda_3 \leq 0$  and  $\lambda_1 \geq \lambda_2 \geq \lambda_3$ , then the special symmetric matrix  $E \in \mathcal{S}$  which minimizes the error  $\|E - F\|_f^2$  is given by  $E = V \text{diag}\{\sigma_1, \sigma_2, \sigma_2\} V^T$  with:*

$$\begin{aligned} \sigma_1 &= \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, & \sigma_2 &= \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \\ \sigma_3 &= \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}. \end{aligned} \quad (54)$$

**Proof:** Define  $\mathcal{S}_\Sigma$  to be the subspace of  $\mathcal{S}$  whose elements have the same eigenvalues:  $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \sigma_3\}$ . Thus every matrix  $E \in \mathcal{S}_\Sigma$  has the form  $E = V_1 \Sigma V_1^T$  for some  $V_1 \in SO(3)$ . To simplify the notation, define  $\Sigma_\lambda = \text{diag}\{\lambda_1, \lambda_2, \lambda_3\}$ . We now prove this theorem by two steps.

*Step 1:* We prove that the special symmetric matrix  $E \in \mathcal{S}_\Sigma$  which minimizes the error  $\|E - F\|_f^2$  is given by  $E = V \Sigma V^T$ . Since  $E \in \mathcal{S}_\Sigma$  has the form  $E = V_1 \Sigma V_1^T$ , we get:

$$\begin{aligned} \|E - F\|_f^2 &= \|V_1 \Sigma V_1^T - V \Sigma_\lambda V^T\|_f^2 \\ &= \|\Sigma_\lambda - V^T V_1 \Sigma V_1^T V\|_f^2. \end{aligned} \quad (55)$$

Define  $W = V^T V_1 \in SO(3)$  and  $W$  has the form:

$$W = \begin{pmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{pmatrix}. \quad (56)$$

Then:

$$\begin{aligned} & \|E - F\|_f^2 \\ &= \|\Sigma_\lambda - W\Sigma W^T\|_f^2 \\ &= \text{tr}(\Sigma_\lambda^2) - 2\text{tr}(W\Sigma W^T\Sigma_\lambda) + \text{tr}(\Sigma^2). \end{aligned} \quad (57)$$

Substituting (56) into the second term, and using the fact that  $\sigma_2 = \sigma_1 + \sigma_3$  and  $W$  is a rotation matrix, we get:

$$\begin{aligned} & \text{tr}(W\Sigma W^T\Sigma_\lambda) \\ &= \sigma_1(\lambda_1(1 - w_3^2) + \lambda_2(1 - w_6^2) + \lambda_3(1 - w_9^2)) \\ & \quad + \sigma_3(\lambda_1(1 - w_1^2) + \lambda_2(1 - w_4^2) \\ & \quad + \lambda_3(1 - w_7^2)). \end{aligned} \quad (58)$$

Minimizing  $\|E - F\|_f^2$  is equivalent to maximizing  $\text{tr}(W\Sigma W^T\Sigma_\lambda)$ . From (58),  $\text{tr}(W\Sigma W^T\Sigma_\lambda)$  is maximized if and only if  $w_3 = w_6 = 0$ ,  $w_9^2 = 1$ ,  $w_4 = w_7 = 0$  and  $w_1^2 = 1$ . Since  $W$  is a rotation matrix, we also have  $w_2 = w_8 = 0$  and  $w_5^2 = 1$ . All possible  $W$  give a unique matrix in  $\mathcal{S}_\Sigma$  which minimizes  $\|E - F\|_f^2$ :  $E = V\Sigma V^T$ .

*Step 2:* From step one, we only need to minimize the error function over the matrices which have the form  $V\Sigma V^T \in \mathcal{S}$ . The optimization problem is then converted to one of minimizing the error function:

$$\begin{aligned} \|E - F\|_f^2 &= (\lambda_1 - \sigma_1)^2 \\ & \quad + (\lambda_2 - \sigma_2)^2 + (\lambda_3 - \sigma_3)^2 \end{aligned} \quad (59)$$

subject to the constraint:

$$\sigma_2 = \sigma_1 + \sigma_3. \quad (60)$$

The formula (54) for  $\sigma_1, \sigma_2, \sigma_3$  are directly obtained from solving this minimization problem.  $\square$

*Remark 1.* For symmetric matrices which do not satisfy conditions  $\lambda_1 \geq 0$  or  $\lambda_3 \leq 0$ , one may simply choose  $\lambda'_1 = \max(\lambda_1, 0)$  or  $\lambda'_3 = \min(\lambda_3, 0)$ .

We then have an eigenvalue-decomposition based algorithm for estimating 3D velocity from optical flow.

#### Four-Step 3D Velocity Estimation Algorithm:

1. *Estimate essential vector:*

For a given set of optical flows:  $(\mathbf{q}^i, \mathbf{u}^i)$ ,  $i =$

$1, \dots, m$ , find the vector  $\mathbf{e}$  which minimizes the error function:

$$V(\mathbf{e}) = \|\mathbf{Ae}\|^2 \quad (61)$$

subject to the condition  $\|\mathbf{e}\| = 1$ ;

2. *Recover the special symmetric matrix:*

Recover the vector  $v_0 \in S^2$  from the first three entries of  $\mathbf{e}$  and the symmetric matrix  $s \in \mathbb{R}^{3 \times 3}$  from the remaining six entries.<sup>4</sup> Find the eigenvalue decomposition of the symmetric matrix  $s$ :

$$s = V_1 \text{diag}\{\lambda_1, \lambda_2, \lambda_3\} V_1^T \quad (62)$$

with  $\lambda_1 \geq \lambda_2 \geq \lambda_3$ . Project the symmetric matrix  $s$  onto the special symmetric space  $\mathcal{S}$ . We then have the new  $s = V_1 \text{diag}\{\sigma_1, \sigma_2, \sigma_3\} V_1^T$  with:

$$\begin{aligned} \sigma_1 &= \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, & \sigma_2 &= \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \\ \sigma_3 &= \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}; \end{aligned} \quad (63)$$

3. *Recover velocity from the special symmetric matrix:* Define:

$$\begin{cases} \lambda = \sigma_1 - \sigma_3, & \lambda \geq 0, \\ \theta = \arccos\left(\frac{-\sigma_2}{\lambda}\right), & \theta \in [0, \pi]. \end{cases} \quad (64)$$

Let  $V = V_1 R_Y^T(\frac{\theta}{2} - \frac{\pi}{2}) \in SO(3)$  and  $U = -VR_Y(\theta) \in O(3)$ . Then the four possible 3D velocities corresponding to the special symmetric matrix  $s$  are given by:

$$\begin{cases} \hat{\omega} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda U^T, & \hat{v} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 V^T \\ \hat{\omega} = VR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_\lambda V^T, & \hat{v} = UR_Z\left(\pm\frac{\pi}{2}\right)\Sigma_1 U^T \end{cases} \quad (65)$$

where  $\Sigma_\lambda = \text{diag}\{\lambda, \lambda, 0\}$  and  $\Sigma_1 = \text{diag}\{1, 1, 0\}$ ;

4. *Recover velocity from the differential essential matrix:*

From the four velocities recovered from the special symmetric matrix  $s$  in step 3, choose the pair  $(\omega^*, v^*)$  which satisfies:

$$v^{*T} v_0 = \max_i v_i^T v_0. \quad (66)$$

Then the estimated 3D velocity  $(\omega, v)$  with  $\omega \in \mathbb{R}^3$  and  $v \in S^2$  is given by:

$$\omega = \omega^*, \quad v = v_0. \quad (67)$$

Both  $v_0$  and  $v^*$  are estimates of the linear velocity. However, experimental results show that, statistically, within the tested noise levels (see next section),  $v_0$  yields a better estimate than  $v^*$ . Here, thus, we simply choose  $v_0$  as the estimate. Nonetheless, one can find statistical correlations between  $v_0$  and  $v^*$  (experimentally or analytically) and obtain better estimates for  $v$ , using both  $v_0$  and  $v^*$ . Another potential way to improve this algorithm is to study the systematic bias introduced by the least square method in step 1. A similar problem has been studied by Kanatani (1993a) and an algorithm was proposed to remove such bias from Zhuang’s algorithm (Zhuang and Haralick, 1984).

*Remark 2.* Since both  $E, -E \in \mathcal{E}'_1$  satisfy the same set of differential epipolar constraints, both  $(\omega, \pm v)$  are possible solutions for the given set of optical flows. However, as in the discrete case, one can get rid of the ambiguous solution by adding the “positive depth constraint”.

*Remark 3.* By the way of comparison to Heeger and Jepson’s algorithm (Heeger and Jepson, 1992), note that the Eq. (53) may be rewritten to highlight the dependence on optical flow as:  $[A_1(\mathbf{u}) \mid A_2]\mathbf{e} = 0$  where  $A_1(\mathbf{u}) \in \mathbb{R}^{m \times 3}$  is a linear function of the measured optical flow and  $A_2 \in \mathbb{R}^{m \times 6}$  is a function of the image points alone. Heeger and Jepson compute a left null space to the matrix  $A_2$  ( $C \in \mathbb{R}^{(m-6) \times m}$ ) and solve the equation:  $CA_1(\mathbf{u})v = 0$  for  $v$  alone. Then they use  $v$  to obtain  $\omega$ . Our method simultaneously estimates  $v \in \mathbb{R}^3, s \in \mathbb{R}^6$ . We make a detailed simulation comparison of these two algorithms in Section 4.

One should note that this linear algorithm is not optimal in the sense that the recovered velocity does not necessarily minimize the originally picked error function  $\|A\mathbf{e}(\omega, v)\|^2$  on  $\mathcal{E}'_1$  (see next section for a more detailed discussion). However, this algorithm only uses linear algebra techniques and is particularly simpler than a one which tries to optimize on the manifold  $\mathcal{E}'_1$  (Ma et al., 1998c).

One potential problem with the (discrete or differential) essential approaches is that the motion estimation schemes are all based on the assumption that the

translation is not zero. In this section, we study what makes the epipolar constraint fail to work in the zero-translation case.

For the discrete case, if two images are obtained from rotation alone i.e.,  $p = 0$  and  $\mathbf{q}_c = R^T \mathbf{q}_0$ , it is straightforward to check that, for all  $p \in S^2$ , we have:

$$\mathbf{q}_c^T R^T \hat{p} \mathbf{q}_0 \equiv 0. \quad (68)$$

Thus, theoretically, the estimation schemes working on the normalized essential space  $\mathcal{E}_1$  will fail to converge (since there are infinite many pairs of  $(R, p)$  satisfying the same set of epipolar constraints). In the differential case, we have a similar situation:

**Theorem 6.** *An optical flow field  $(\mathbf{q}, \mathbf{u})$  is obtained from a pure rotation with the angular velocity  $\omega$  if and only if for all vectors  $v \in S^2$*

$$(\mathbf{u}^T, \mathbf{q}^T) \begin{pmatrix} \hat{v} \\ [\hat{\omega} \hat{v}] \end{pmatrix} \mathbf{q} = 0. \quad (69)$$

**Proof:**  $\mathbf{u} = \hat{\omega} \mathbf{q}$  since  $\mathbf{u}$  is obtained from rotation  $\omega \Leftrightarrow \mathbf{u}^T(v \times \mathbf{q}) = -\mathbf{q}^T \hat{\omega}(v \times \mathbf{q})$  for all  $v \in S^2 \Leftrightarrow (\mathbf{u}^T, \mathbf{q}^T)(\hat{v}[\hat{\omega} \hat{v}])\mathbf{q} = 0$ .  $\square$

This theorem implies that the velocity estimation algorithm proposed in the previous section will have trouble when the linear velocity  $v$  is zero, since there are infinite many pairs of  $(\omega, v)$  satisfying the same set of differential epipolar constraints. However, it is shown by Soatto and Perona (1996) that, in the dynamical estimation approach, one can actually make use of the noise in the measurements to obtain correct estimate of the rotational component  $R$  regardless of the accuracy of the estimate for the translation vector  $p$ . The same should hold also in the differential case. That is, even in the zero-translation case, the recovery of the angular velocity  $\omega$  is still possible using dynamic estimation schemes. Study of such schemes is beyond the scope of this paper and will be addressed in our future research work.

**Example: Kinematic model of an aircraft.** This example shows how to utilize *nonholonomic constraints* (see Murray et al. (1994)) to simplify the proposed linear motion estimation algorithm in the differential case. Let  $g(t) \in SE(3)$  represents the position and orientation of an aircraft relative to the spatial frame, the inputs  $\omega_1, \omega_2, \omega_3 \in \mathbb{R}$  stand for the rates of the rotation about the axes of the aircraft and  $v_1 \in \mathbb{R}$  the velocity of the

aircraft. Using the standard *homogeneous representation* for  $g$  (see Murray et al. (1994)), the kinematic equations of the aircraft motion are given by:

$$\dot{g} = g \begin{pmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & 0 \\ -\omega_2 & \omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (70)$$

where  $\omega_1$  stands for pitch rate,  $\omega_2$  for roll rate,  $\omega_3$  for yaw rate and  $v_1$  the velocity of the aircraft. Then the 3D velocity  $(\omega, v)$  in the differential epipolar constraint (21) has the form:  $\omega = (\omega_1, \omega_2, \omega_3)^T$ ,  $v = (v_1, 0, 0)^T$ . For the algorithm given in Section 2.4, this adds extra constraints on the symmetric matrix  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$ :  $s_1 = s_5 = 0$  and  $s_4 = s_6$ . Then there are only four different essential parameters left to determine and we can re-define the essential parameter vector  $\mathbf{e} \in \mathbb{R}^4$  to be:  $\mathbf{e} = (v_1, s_2, s_3, s_4)^T$ . Then the measurement vector  $\mathbf{a} \in \mathbb{R}^4$  is to be:  $\mathbf{a} = (u_3y - u_2z, 2xy, 2xz, y^2 + z^2)^T$ . The differential epipolar constraint can then be rewritten as:

$$\mathbf{a}^T \mathbf{e} = 0. \quad (71)$$

If we define the matrix  $A$  from  $\mathbf{a}$  like in (52), the matrix  $A^T A$  is a  $4 \times 4$  matrix rather than a  $9 \times 9$  one. For estimating the velocity  $(\omega, v)$ , the dimensions of the problem is then reduced from 9 to 4. In this special case, the minimum number of optical flow measurements needed to guarantee a unique solution of  $\mathbf{e}$  is reduced to 3 instead of 8. Further more, the symmetric matrix  $s$  recovered from  $\mathbf{e}$  is automatically in the special symmetric space  $\mathcal{S}$  and the remaining steps of the algorithm given in Section 2.4 can be thus dramatically simplified. From this simplified algorithm, the angular velocity  $\omega = (\omega_1, \omega_2, \omega_3)^T$  can be fully recovered from the images. The velocity information can be then used for controlling the aircraft.

### 3. Experimental Results

We have carried out some initial simulations in order to study the performance of our algorithm. We chose to evaluate it in terms of bias and sensitivity of the estimates with respect to the noise in the optical flow measurements. Preliminary simulations were carried out with perfect data which was corrupted by zero-mean Gaussian noise where the standard deviation was specified in terms of pixel size and was independent of

velocity. The image size was considered to be  $512 \times 512$  pixels. Our algorithm has been implemented in Matlab and the simulations have been performed using example sets proposed by Tian et al. (1996), in their paper on comparison of the egomotion estimation from optical flow.<sup>5</sup> The motion estimation was performed by observing the motion of a random cloud of points placed in front of the camera. Depth range of the points varied from  $a$  to  $b$  ( $>a$ ) units of the focal length  $f$ , which was considered to be unity. For example, if the focal length is 8 mm and  $a = 100$  and  $b = 400$ , the point depth varies from 0.8 m to 3.2 m in front of the camera. This setup makes the simulation depend only on the parameter  $c = (b - a)/a$ , called *depth variation parameter*. The results presented below are for a fixed field of view (FOV) of 60 degrees unless otherwise stated.

#### 3.1. Comparing to Subspace Methods

Each simulation consisted of 500 trials for 50 randomly sampled points in a given depth variation  $[a, b] = [100, 400]$  with a fixed noise level and ratio between the image velocity due to translation and rotation for the point in the middle of the random cloud. Figures 3 and 4 compare our algorithm with Heeger and Jepson's linear subspace algorithm (Heeger and Jepson, 1992). The presented results demonstrate the performance of the algorithm while rotating around  $X$ -axis with rate of  $1^\circ$  per frame and translating along  $Y$ -axis with translation to rotation ratio of 1 and 5 respectively (for the point at the center of the random cloud). The first stage of our analysis was performed using benchmarks proposed by Tian et al. (1996). The bias is expressed as an angle between the average estimate out of all trials (for a given setting of parameters) and the true direction of translation and/or rotation. The sensitivity was computed as a standard deviation of the distribution of angles between each estimated vector and the average vector in case of translation and as a standard deviation of angular differences in case of rotation.

We further evaluated the algorithm by varying the direction of translation and rotation. At the noise level of 0.9 pixel and translation/rotation ratio 1, for different combination of translation and rotation axis, the bias of these two algorithm are shown in Fig. 5. From the simulation results, we observe that:

1. In terms of translational bias and sensitivity, the subspace method (Heeger and Jepson, 1992) and our

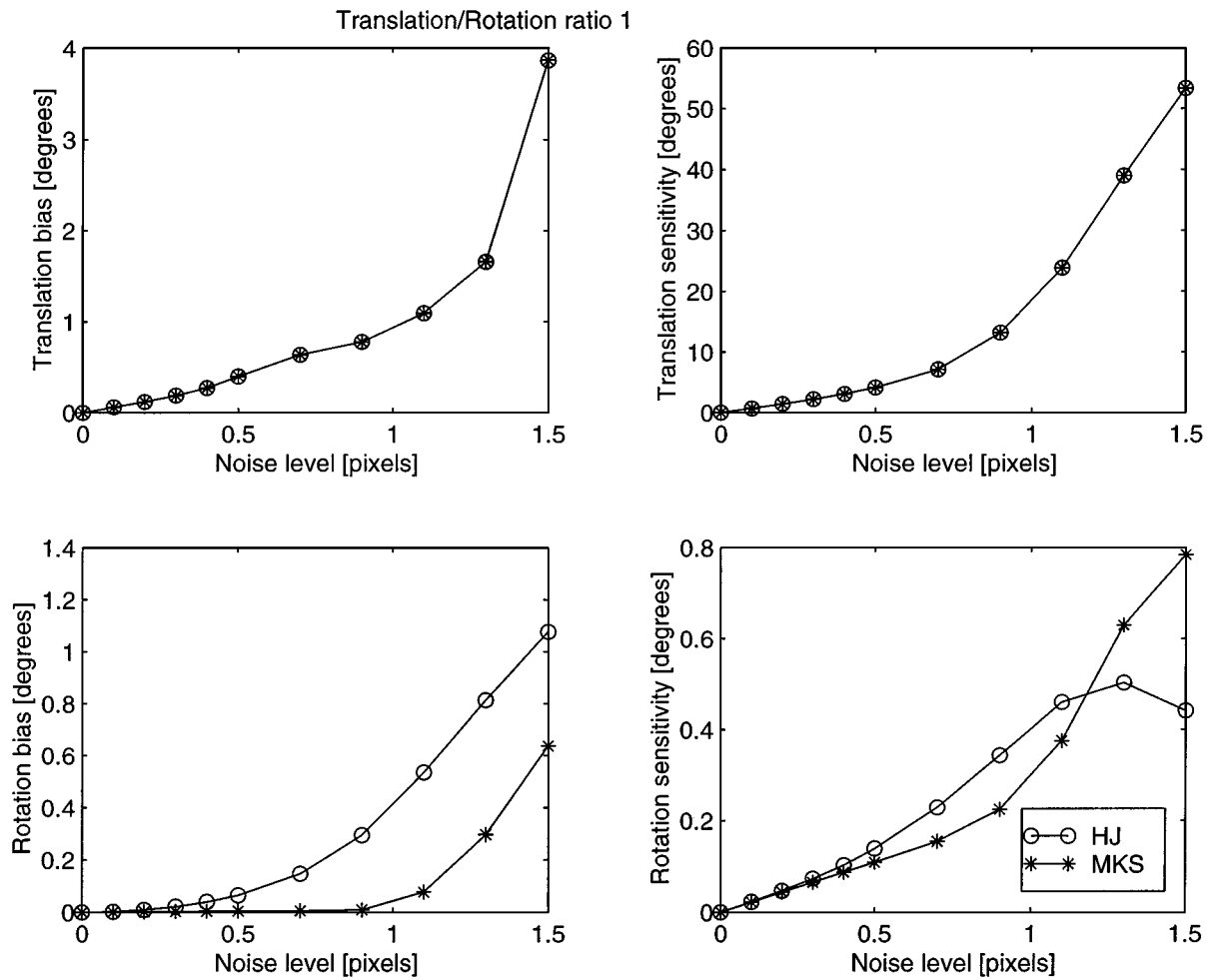


Figure 3. Bias for each noise level was estimated by running 500 trials and computing the average translation and rotation. The ratio between the magnitude of linear and angular velocities is 1.

algorithm have exactly the same performance at all noise levels.

2. The choice of the rotation axis does not influence the translation estimates at all for both algorithms. It does not generally influence the rotation estimates for the subspace method either but indeed influences our algorithm. This is because the decomposition of the special symmetric matrix  $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$  is numerically less accurate when  $\omega$  and  $v$  coincide with each other.
3. Both algorithms give much better estimates when translation along Z-axis is present. This is consistent to the sensitivity analysis done in Daniilidis and Nagel (1990). In the case of translation in X-Y plane, our algorithm gives better rotation estimates

than the subspace method (Heeger and Jepson, 1992), especially when the noise levels are high.

This is due to the fact that in our algorithm the rotation is estimated simultaneously with the translation, so that its bias is only due to the bias of the initially estimated differential essential matrix obtained by linear least squares techniques. This is in contrast to the rotation estimate used by the subspace method (Heeger and Jepson, 1992), which uses another least-squares estimation by substituting an already biased translational estimate to compute the rotation. Increasing the ratio between the magnitude of translational and rotational velocities, the performance of both algorithms improves, especially the translation estimates.

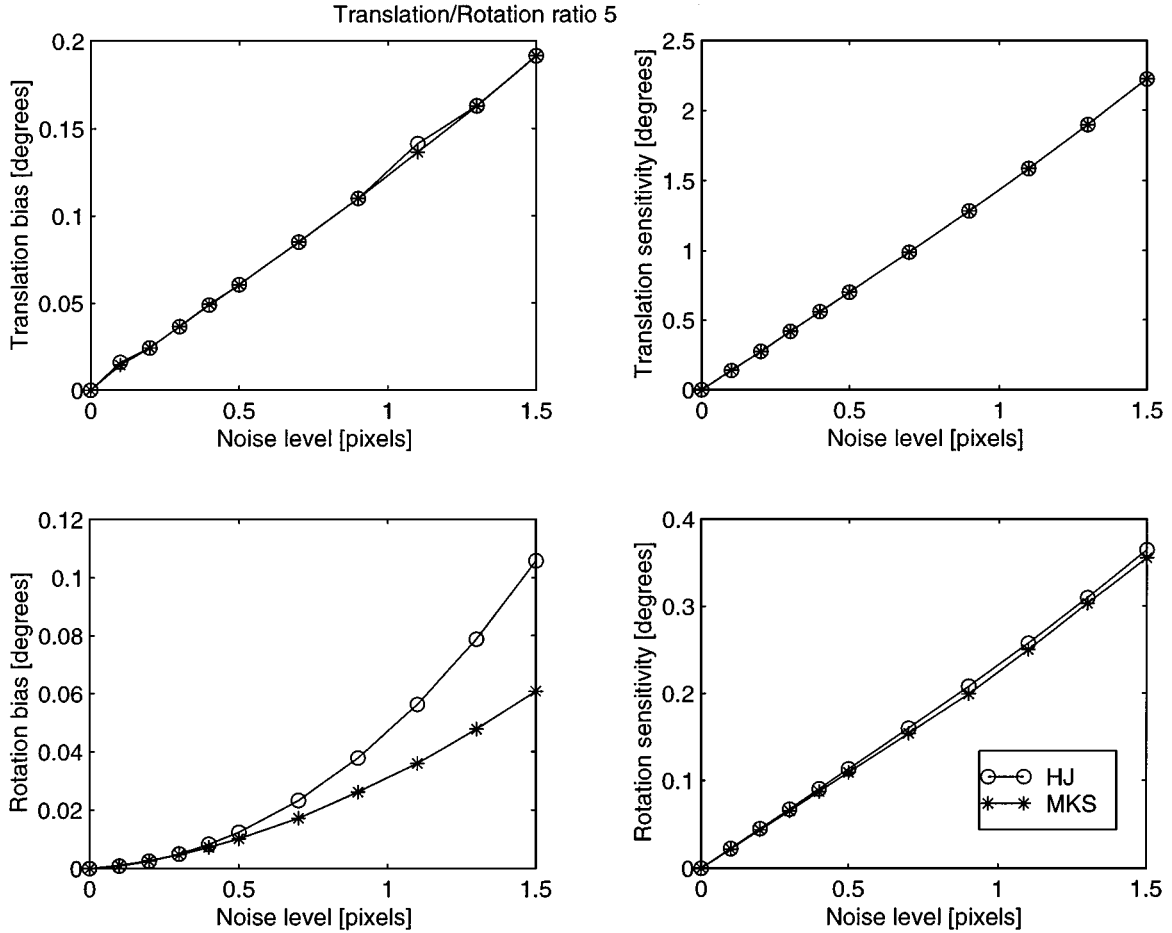


Figure 4. Bias for each noise level was estimated by running 500 trials and computing the average translation and rotation. The ratio between the magnitude of linear and angular velocities is 5.

### 3.2. Bias Analysis: Relation with Nonlinear Algorithms

A disadvantage of any linear algorithm is that it tries to directly minimize the epipolar constraint, i.e., the objective function:

$$V(\omega, v) = \sum_{i=1}^N (\mathbf{u}^{iT} \hat{v} \mathbf{q}^i + \mathbf{q}^{iT} \hat{\omega} \hat{v} \mathbf{q}^i)^2. \quad (72)$$

But this is not the likelihood function of  $\omega$  and  $v$  for commonly used noise models of the optical flow. Consequently, estimates given by linear algorithms are usually not close to maximum *a posteriori* (MAP) or minimum mean square estimates (MMSE). In general, this is the source of bias for linear algorithms. In case of perspective projection, a commonly used noise model is:

$$\tilde{\mathbf{u}}^i = \mathbf{u}^i + \mathbf{n}^i \quad (73)$$

where  $\tilde{\mathbf{u}}^i$ 's are corrupted optical flows and  $\mathbf{n}^i$ 's are independent Gaussian noises with identical covariance matrix  $K = \text{diag}\{\sigma^2, \sigma^2, 0\}$ . Substitute  $\tilde{\mathbf{u}}^i$  into the epipolar constraint and we obtain

$$\tilde{\mathbf{u}}^{iT} \hat{v} \mathbf{q}^i + \mathbf{q}^{iT} \hat{\omega} \hat{v} \mathbf{q}^i = \mathbf{n}^{iT} \hat{v} \mathbf{q}^i. \quad (74)$$

The random variable  $\mathbf{n}^{iT} \hat{v} \mathbf{q}^i$  is Gaussian with distribution  $N(0, \|\hat{e}_3 \hat{v} \mathbf{q}^i\|^2)$  where  $\mathbf{e}_3 = (0, 0, 1)^T$ . Consequently, the variance in general depends on the location of the image point  $\mathbf{q}^i$ . Assuming that the *a priori* distributions of  $\omega, v$  are uniform, the MAP estimates are then given by the  $\omega^*, v^*$  which minimize the objective function:

$$V(\omega, v) = \sum_{i=1}^N \frac{(\mathbf{u}^{iT} \hat{v} \mathbf{q}^i + \mathbf{q}^{iT} \hat{\omega} \hat{v} \mathbf{q}^i)^2}{\|\hat{e}_3 \hat{v} \mathbf{q}^i\|^2}. \quad (75)$$

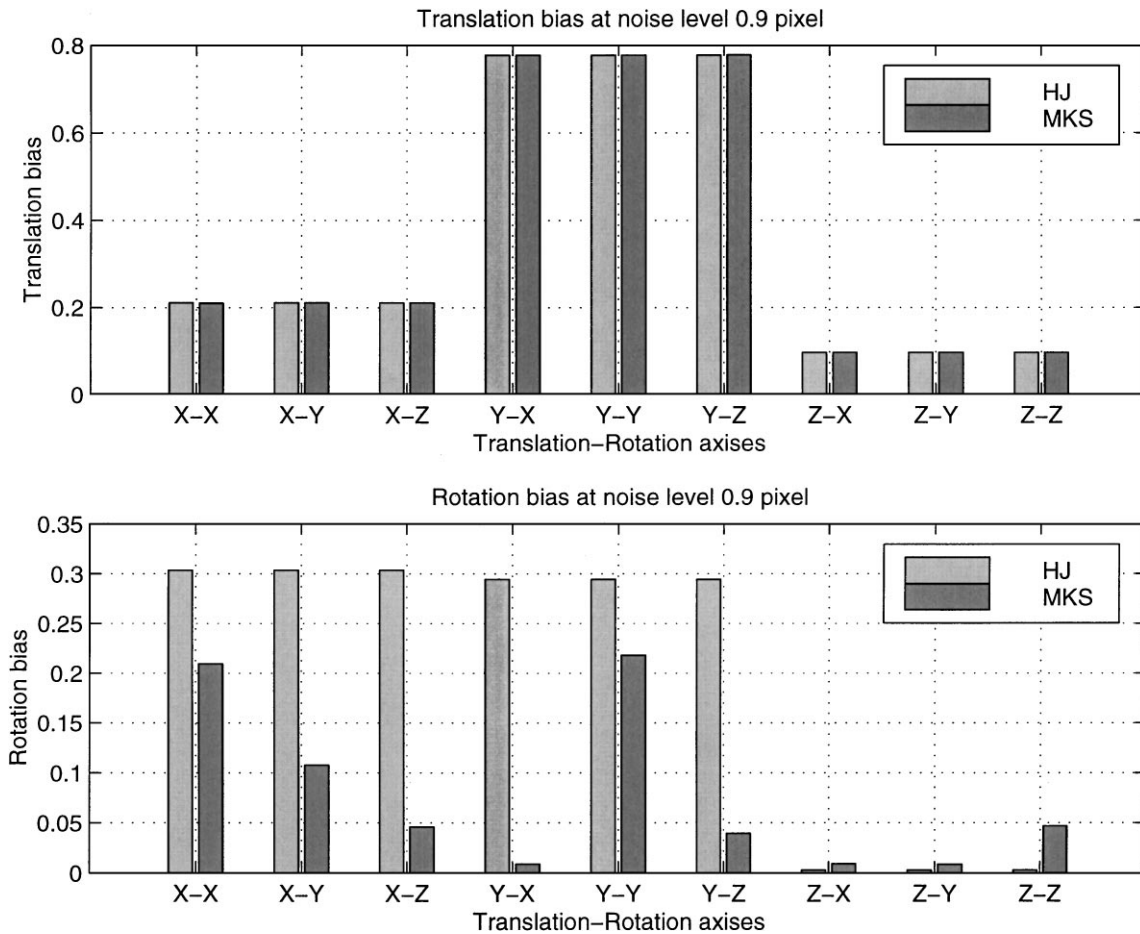


Figure 5. Bias dependency on combination of translation and rotation axes. For example, “X-Y” means the translation direction is in X-axis and rotation axis is the Y-axis. Bias for each combination of axes was estimated by running 500 trials at the noise level 0.9 pixel. The ratio between the magnitude of linear and angular velocities is 1.

This objective function is also referred as *normalized epipolar constraints*. A desirable property of MAP estimates is that they are asymptotically unbiased, i.e., when the number of sample points  $N$  is large, the bias is reduced dramatically. The asymptotic unbiasedness of this motion estimates has been studied extensively by Soatto et al. (submitted), where they have provided a concise and rigorous proof of the asymptotic unbiasedness. However, such a topic is beyond the scope of this paper and we refer interested readers to relevant sections in that paper.

It is important to notice that if the translation  $v$  is in the X-Y plane, i.e.,  $v = (v_1, v_2, 0)^T$ ,  $\|\hat{e}_3 \hat{v} \mathbf{q}^i\| = \|v\|$  is independent of the image point  $\mathbf{q}^i$ . The normalized version (75) is therefore equivalent to the unnormalized version (72). Consequently, the normalization will not

have much effect on the obtained estimates. Note that this is exactly the case when our algorithm performs better than the subspace method. For the case that the Z-axis translation is present, the performance of both algorithms can be further improved by solving the normalized version (especially when a large number of sample points are available). This is, in general, a nonlinear optimization problem and is beyond the scope of this paper. However, since nonlinear algorithms usually only guarantee local convergence, a good linear algorithm may provide a good initial state for a nonlinear search scheme. Simply to demonstrate the effect of normalization, we have run our linear algorithm for the normalized objective function (75) using the actual translation velocity in the denominator to normalize the epipolar constraint and compare the results with those



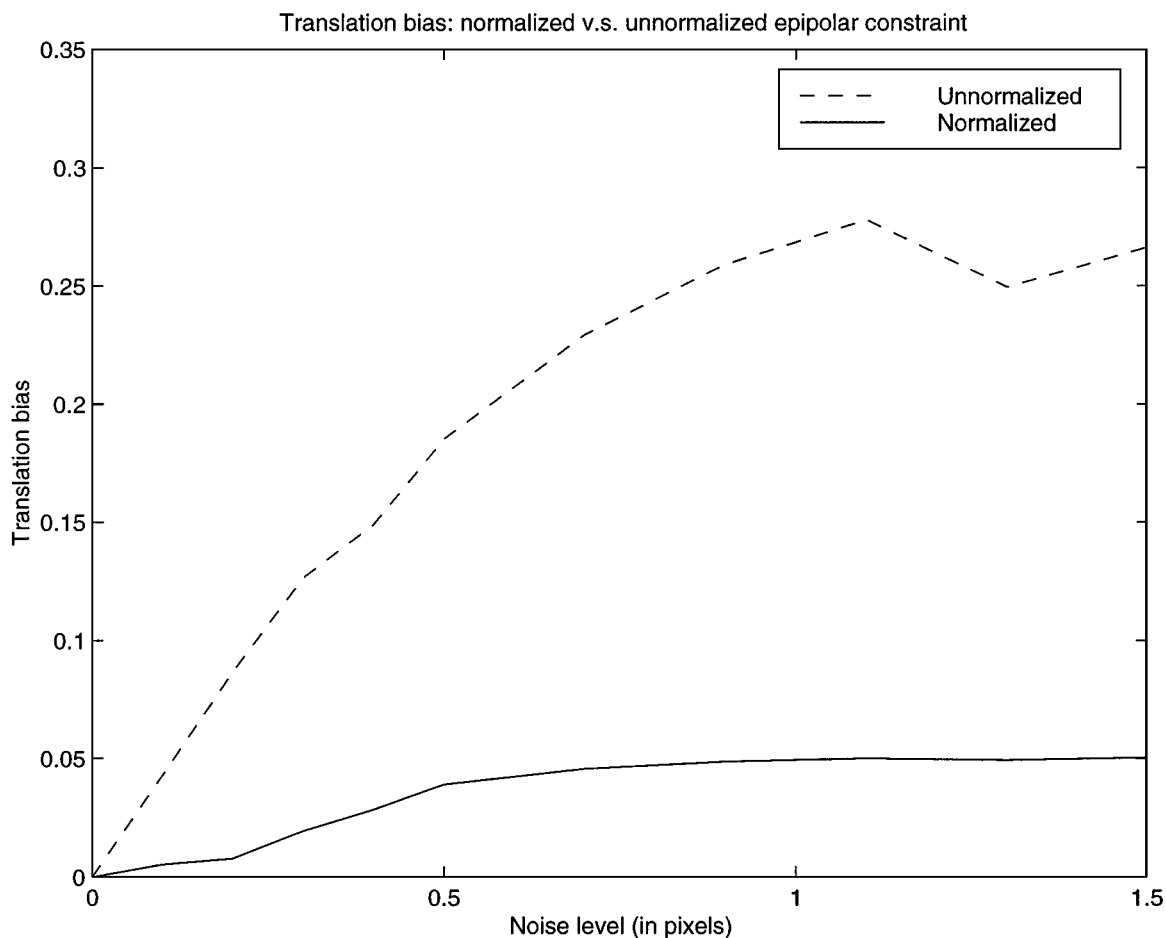


Figure 6. Translation bias of using normalized and unnormalized epipolar constraints. Bias for each noise level is estimated by running 50 trials. Both rotation and translation is along the Z-axis and the ratio between the magnitude of linear and angular velocities is 1.

from the unnormalized one (a loyal implementation of the nonlinear optimization scheme can be found in Ma et al. (1998c)). Since the advantage of MAP estimate is only observed asymptotically, we here pick a large number of sample points  $N = 500$ . Because, up to  $N = 500$ , the improvement in rotation estimate is less significant, we only plot the translation bias with respect to different noise levels in Fig. 6. Both the translation direction and rotation axis are along the Z-axis.

### 3.3. Sensitivity to the Depth Variation Parameter $c$

Simulations also show that the depth variation parameter  $c = (b - a)/a$  is another important measure of the performance of the linear algorithm. We ran our

algorithm at noise levels 0.3, 0.6 and 0.9 pixel with respect to depth variation parameter  $c = 2.5, 3.0, 3.5, 4.0$  for a cloud of 200 points. Translation and rotation biases are plotted in Fig. 7. One should notice that at high noise levels, the bias increases almost exponentially when the depth variation parameter  $c$  decreases. When the depth variation parameter is small and noise level is high, the proposed algorithm gives less robust estimates. Especially, estimates of the bias and sensitivity become less stable. In order to demonstrate the true monotonic relation between the sensitivity and noise level, a larger number of trials have to be performed. In this case, the fact that the points in the scene are (almost) coplanar has to be incorporated in the designing of motion estimation algorithms (Longuet-Higgins 1986; Subbarao and Waxman, 1985).

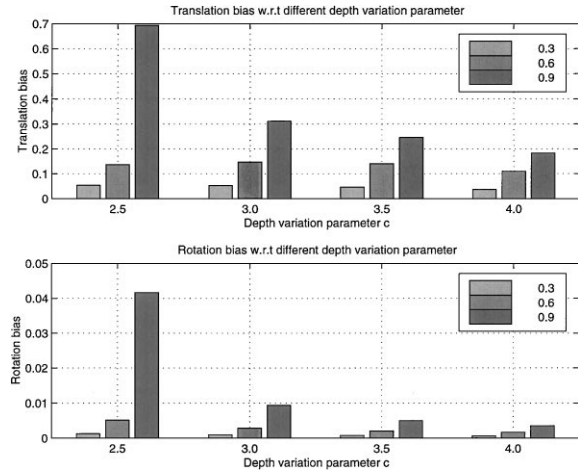


Figure 7. Translation bias and rotation bias with respect to different depth variation parameter  $c$ . Bias for each noise level and depth variation parameter is estimated by running 500 trials. Translation is along the  $X$ -axis and rotation axis is the  $Z$ -axis and the ratio between the magnitude of linear and angular velocities is 1.

### 3.4. Translation estimates $v_0$ versus $v^*$

Further evaluation of the results and more extensive simulations are currently underway. We believe that thoroughly understanding the source of translational bias, we can obtain even better performance by utilizing additional information about the linear velocity which is embedded in the special symmetric part of the differential essential matrix, i.e.,  $v^*$  (see step 4 of the algorithm in the preceding section). In the above simulations, the linear velocity  $v$  was estimated only from the  $v_0$ , the skew symmetric part of the differential essential matrix. Figure 8 demonstrates that  $v_0$  is in general a much better estimate than  $v^*$ . A more detailed analysis of the statistical correlation between  $v_0$  and  $v^*$  is currently under investigation.

## 4. Discussions and Future Work

This paper presents a unified (linear) approach for the problem of egomotion estimation using discrete and differential epipolar constraints. In either the discrete or differential setting, a geometric characterization is given for the space of essential matrices or differential essential matrices. Such characterization gives a natural geometric interpretation for the number of possible solutions to the motion estimation problem. In addition, in the differential case, understanding of the space of differential essential matrices leads to a new

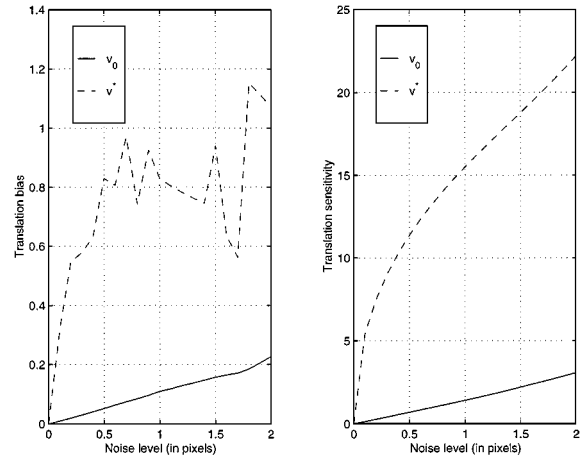


Figure 8. Bias and sensitivity of the translation estimates  $v_0$  from the skew symmetric part and  $v^*$  from the special symmetric part of the differential essential matrix. Bias and sensitivity for each noise level are estimated by running 200 trials for a cloud of 50 points. Both translation and rotation are along the  $X$ -axis and the ratio between the magnitude of linear and angular velocities is 5.

egomotion estimation algorithm, which is a natural counterpart of the well-known three-step SVD based algorithm developed for the discrete case by Toscani and Faugeras (1986).

In order to exploit temporal coherence of motion and improve algorithm's robustness, a dynamic (recursive) motion estimation scheme, which uses implicit extended Kalman filter for estimating the essential parameters, has been proposed by Soatto and Perona (1996) for the discrete case. The same ideas certainly apply to our algorithm.

Although only linear algorithms are studied in this paper, the understanding of the geometric characterization of the essential matrix spaces leads to natural geometric nonlinear search algorithms based on optimization techniques on Riemannian manifolds. Those intrinsic nonlinear algorithms will be presented in the sequel of this paper (Ma et al., 1998c). The problem of 3D structure reconstruction is not discussed in this paper. Like the motion estimation, this subject has also been studied extensively in the computer vision literature. A unified approach has been proposed in a sequel paper (Ma et al., 1998a).

In this paper, we have assumed that the camera is calibrated. Our approach can be extended to uncalibrated camera case, where the motion estimation and camera self-calibration problem can be solved simultaneously, using the differential essential constraint (Vieville and Faugeras, 1995, Brooks et al., in press). In

this case, the essential matrix is replaced by the fundamental matrix which captures both motion information and camera intrinsic parameters. It is shown in Brooks et al. (in press) and Ma et al. (1998b), that besides five motion parameters only two extra intrinsic parameters can be recovered.

## Acknowledgments

We would like to thank M. Cavusoglu and J. Yan at the Berkeley Intelligent Machines & Robotics Laboratory for their helpful discussions on some of the proofs in this paper. We also thank professor Weinstein of the Mathematics Department at UC Berkeley for his insightful comments on the geometry of the space of essential matrices. Research supported in part by ARO under the MURI grant DAAH04-96-1-0341.

## Notes

1. An essential matrix always has a SVD such that  $U, V \in SO(3)$ .
2.  $O(3)$  represents the space of all orthogonal matrices (of determinant  $\pm 1$ .)
3. For perspective projection,  $z = 1$  and  $u_3 = 0$  thus the expression for  $\mathbf{a}$  can be simplified.
4. In order to guarantee  $v_0$  to be of unit length, one needs to “renormalize”  $\mathbf{e}$ , i.e., multiply  $\mathbf{e}$  by a scalar such that the vector determined by the first three entries is of unit length.
5. We would like to thank the authors in Tian et al. (1996) for making the code for simulations of various algorithms and evaluation of their results available on the web.

## References

- Brooks, M.J., Chojnacki, W., and Baumela, L. (in press). Determining the ego-motion of an uncalibrated camera from instantaneous optical flow.
- Bruss, A.R. and Horn, B.K. 1983. Passive navigation. *Computer Graphics and Image Processing*, 21:3–20.
- Daniilidis, K. and Nagel, H.-H. 1990. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303.
- Heeger, D.J. and Jepson, A.D. 1992. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117.
- Huang, T. and Faugeras, O. 1989. Some properties of the E matrix in two-view motion estimation. *IEEE PAMI*, 11(12):1310–12.
- Jepson, A.D. and Heeger, D.J. 1993. Linear subspace methods for recovering translation direction. In *Spatial Vision in Humans and Robots*, Cambridge Univ. Press, pp. 39–62.
- Kanatani, K. 1993a. 3D interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282.
- Kanatani, K. 1993b. *Geometric Computation for Machine Vision*. Oxford Science Publications.
- Longuet-Higgins, H.C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.
- Longuet-Higgins, H.C. 1986. The reconstruction of a plane surface from two perspective projections. In *Proceedings of Royal Society of London*, volume 227 of B, pp. 399–410.
- Ma, Y., Košecák, J., and Sastry, S. 1998a. Euclidean structure and motion from image sequences. Electronic Research Laboratory Memorandum, UC Berkeley, UCB/ERL (M98/38).
- Ma, Y., Košecák, J., and Sastry, S. 1998b. A mathematical theory of camera self-calibration. Electronic Research Laboratory Memorandum, UC Berkeley, UCB/ERL(M98/64).
- Ma, Y., Košecák, J., and Sastry, S. 1998c. Optimal motion from image sequences: A Riemannian viewpoint. Electronic Research Laboratory Memorandum, UC Berkeley, UCB/ERL(M98/37).
- Maybank, S. 1993. *Theory of Reconstruction from Image Motion*. Springer Series in Information Sciences. Springer-Verlag.
- McLauchlan, P.F. and Murray, D.W. 1995. A unifying framework for structure and motion recovery from image sequences. In *Proceeding of Fifth International Conference on Computer Vision*, Cambridge, MA, USA, IEEE Comput. Soc. Press, pp. 314–320.
- Murray, R.M., Li, Z., and Sastry, S.S. 1994. *A Mathematical Introduction to Robotic Manipulation*. CRC Press Inc.
- Ponce, J. and Genc, Y. 1998. Epipolar geometry and linear subspace methods: a new approach to weak calibration. *International Journal of Computer Vision*, 28(3):223–243.
- Soatto, S., Chiuso, A., and Brockett, R. (submitted). Optimal structure from motion: Local ambiguities and global estimates. *International Journal of Computer Vision*.
- Soatto, S. and Perona, P. 1996. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393–413.
- Subbarao, M. and Waxman, A.M. 1985. On the uniqueness of image flow solutions for planar surfaces in motion. In *Third IEEE Workshop on Computer Vision: Representation and Control*, pp. 129–140.
- Tian, T.Y., Tomasi, C., and Heeger, D.J. 1996. Comparison of approaches to egomotion computation. In *Proceedings of 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, IEEE Comput. Soc. Press, pp. 315–320.
- Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography. *Intl. Journal of Computer Vision*, 9(2):137–154.
- Toscani, G. and Faugeras, O.D. 1986. Structure and motion from two noisy perspective images. *Proceedings of IEEE Conference on Robotics and Automation*, pp. 221–227.
- Tsai, R.Y. and Huang, T.S. 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(1):13–27.
- Vieville, T. and Faugeras, O.D. 1995. Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. In *Proceedings of Fifth International Conference on Computer Vision*, pp. 750–756.
- Waxman, A.M., Kamgar-Parsi, B., and Subbarao, M. 1987. Closed form solutions to image flow equations for 3D structure and motion. *International Journal of Computer Vision*, 1: pp. 239–258.
- Zhuang, X. and Haralick, R.M. 1984. Rigid body motion and optical flow image. In *Proceedings of the First International Conference on Artificial Intelligence Applications*, pp. 366–375.
- Zhuang, X., Huang, T.S., and Ahuja, N. 1988. A simplified linear optic flow motion algorithm. *Computer Vision, Graphics and Image Processing*, 42:334–344.