MOTION RECOVERY FROM IMAGE SEQUENCES: DISCRETE VIEWPOINT VS. DIFFERENTIAL VIEWPOINT

by

Yi Ma, Jana Košecká, and Shankar Sastry

Memorandum No. UCB/ERL M98/11

12 February 1998

MOTION RECOVERY FROM IMAGE SEQUENCES: DISCRETE VIEWPOINT VS. DIFFERENTIAL VIEWPOINT

by

Yi Ma, Jana Košecká, and Shankar Sastry

Memorandum No. UCB/ERL M98/11

12 February 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Motion Recovery From Image Sequences: Discrete Viewpoint vs. Differential Viewpoint *

Yi Ma† Jana Košecká‡ Shankar Sastry§

Electronics Research Laboratory
University of California at Berkeley
Berkeley, CA 94720-1774
{mayi, janka, sastry}@robotics.eecs.berkeley.edu

February 12, 1998

Abstract

The aim of this paper is to explore intrinsic geometric methods of recovering the three dimensional motion of a moving camera from a sequence of images. Generic similarities between the discrete approach and the differential approach are revealed through a parallel development of their analogous motion estimation theories.

We begin with a brief review of the (discrete) essential matrix approach, showing how to recover the 3D displacement from image correspondences. The space of normalized essential matrices is characterized geometrically: the unit tangent bundle of the rotation group is a double covering of the space of normalized essential matrices. This characterization naturally explains the geometry of the possible number of 3D displacements which can be obtained from the essential matrix.

Second, a differential version of the essential matrix constraint previously explored by [21, 22] is introduced in a isomorphic section. We then present the precise characterization of the space of differential essential matrices, which gives rise to a novel eigenvector-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 SVD-based 3D displacement estimation algorithm from the discrete case.

Finally, 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. Future work will apply these results to cooperate vision with other inertial navigation sensors to recover 3D motion and orientation of mobile robots.

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. Various approaches differ in the types of

*Research is supported by ARO under the MURI grant DAAH04-96-1-0341.

†Address: 211-109 Cory Hall, EECS, UC Berkeley, Berkeley, CA 94720-1772, USA. Tel: (USA) 510-643-2383.

‡Address: 211- 98 Cory Hall, EECS, UC Berkeley, Berkeley, CA 94720-1772, USA. Tel: (USA) 510-643-5806.

§Address: 269M Cory Hall, EECS, UC Berkeley, Berkeley, CA 94720-1772, USA. Tel: (USA) 510-642-1857.

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 of the techniques try to decouple the two problems by estimating the motion first, followed by the structure estimation. So 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 [11]. Some algorithms address the problem of motion and structure (shape) recovery simultaneously either in batch [18] or recursive fashion [11].

The approaches to the motion estimation only, can be partitioned into the discrete and differential methods depending on whether they use as an input set of point correspondences or image velocities. Among the efforts to solve this problem, one of the more appealing approaches is the *essential matrix approach*, proposed by Longuet-Higgins, Huang and Faugeras *et al* in 1980s [7]. 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 of the scene; and are computationally simpler (comparing to most non-intrinsic motion estimation algorithms), using mostly linear algebraic techniques. Tsai and Huang [20] then 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 [19] and later summarized in Maybank [10].

Motivated by recent interests in dynamical motion estimation schemes (Soatto, Frezza and Perona [15]) which usually require smoothness and regularity of the parameter space, the geometric property of the essential matrix space is further explored: the unit tangent bundle of the rotation group, *i.e.* $T_1(SO(3))$, is a double covering of the space of normalized essential matrices.

However, the essential matrix approach based on the Longuet-Higgins constraint only recovers *discrete* 3D displacement. The velocity information can only be approximately obtained from the inverse of the exponential map, as Soatto *et al* did in [15]. In principle, the displacement estimation algorithms obtained by using epipolar constraints work well when the displacement (especially the translation) 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 due to a high sampling rate. In turn, the algorithms become singular due to the small translation and the estimation results become less reliable.

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* [22] with a simplified version given in 1986 [23]; and a first order algorithm was given by Waxman *et al* [8] in 1987. Most of the 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 [3]) or linear subspace methods (Heeger and Jepson [4, 5]). Kanatani [6] 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 Longuet-Higgins 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 of the big advantages of the differential approach is easy to exploit the *nonholonomic constraints* of the mobile base where the camera is mounted. In this paper, we show by example that nonholonomic constraints will reduce the number of dimensions of the motion estimation problem, hence reduce the number of minimum image measurements needed for a unique solution. The motion estimation algorithm can thus be dramatically simplified. The differential approach developed in this paper can also be generalized to uncalibrated camera [21, 2].

Finally, 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 Review of the Discrete Essential Matrix Approach

In this section, we study the problem of recovering 3D displacement of a moving camera from image correspondences (*i.e.* corresponding point features from two images of the same scene taken from different viewpoints) using an intrinsic geometric constraint: *the Longuet-Higgins constraint or the epipolar constraint.* This approach is usually referred to as the *essential matrix approach*, and "*discrete*" is added whenever we need to separate it from the *differential essential matrix approach* to be developed in section 3.

This section gives a geometric review of main results obtained in the study of essential matrices. Although the review is by no means exhaustive, it provides a self closed introduction to the 3D displacement estimation algorithm. Most of these results have already been obtained before by different researchers, but the proofs provided here are more concise and rigorous. As a new result, we show that the unit tangent bundle of the rotation group, *i.e.* $T_1(SO(3))$, is a double covering of the space of normalized essential matrices.

We first introduce some notations 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 $\mathbb{R}^{3\times3}$) by:

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

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

$$p \times q = \hat{p}q. \tag{2}$$

The matrices of rotation by θ radians about y-axis and z-axis are respectively denoted by:

$$R_Y(\theta) = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}, \quad R_Z(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

2.1 Longuet-Higgins Constraint

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

$$SE(3) = \{(p, R) : p \in \mathbb{R}^3, R \in SO(3)\} \tag{4}$$

where $SO(3)$ is the space of 3×3 rotation matrices (unitary matrices with determinant $+1$) on $\mathbb{R}$. 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_o (see Figure 1). By an abuse of notation,

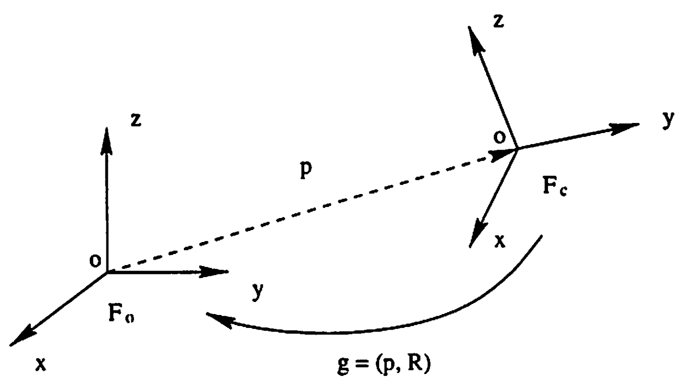

Figure 1: Coordinate frames for specifying rigid body motion of a camera.

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_o.. More precisely, let $q_o, q_c \in \mathbb{R}^3$ be the coordinates of a point q relative to frames F_o and F_c, respectively. Then the coordinate transformation between q_o and q_c is given by:

$$q_o = Rq_c + p. \tag{5}$$

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 $< o, q >$ 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:

$$S^2 = \left\{q \in \mathbb{R}^3 \mid \|q\| = 1\right\}$$

where the norm $\|\cdot\|$ always means 2-norm unless otherwise stated. Then the spherical projection is defined by the map π_s from $\mathbb{R}^3$ to S^2:

$$\begin{aligned} \pi_s : \mathbb{R}^3 &\rightarrow S^2 \\ q &\mapsto \mathbf{q} = \frac{q}{\|q\|}. \end{aligned}$$

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 π_p from $\mathbb{R}^3$ to the projective plane $\mathbb{RP}^2$:

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

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 π_s and π_p by the same letter π. The image of the point q taken by the camera at the initial position then is $\mathbf{q}_o = \pi(q_o)$, and the image of the same point taken at the current position is $\mathbf{q}_c = \pi(q_c)$. It was first obtained by Longuet-Higgins [7] in 1981 that the two corresponding image points $\mathbf{q}_o$ and $\mathbf{q}_c$ have to satisfy a geometric constraint, the so-called *Longuet-Higgins constraint*:

Theorem 1 (Longuet-Higgins Constraint)
Let the 3D displacement of the frame F_c relative to the frame F_o be given by the rigid body motion $g = (p, R) \in SE(3)$, and let $\mathbf{q}_o, \mathbf{q}_c$ be the images of the same point q taken by the camera at frames F_o and F_c, respectively, then $\mathbf{q}_o, \mathbf{q}_c$ satisfy:

$$\mathbf{q}_c^T R^T \hat{p} \mathbf{q}_o \equiv 0. \tag{6}$$

Proof: From (5), the coordinates of q relative to frames F_o and F_c are related by:

$$q_c = R^T(q_o - p). \tag{7}$$

Since $\mathbf{q}_o, \mathbf{q}_c$ are the images of q_o and q_c, respectively, there exists scalars $\lambda_o, \lambda_c \in \mathbb{R}$ (in the spherical projection case, $\lambda_o = \|q_o\|$ and $\lambda_c = \|q_c\|$; in the perspective projection case, $\lambda_o = q_{o3}$ and $\lambda_c = q_{c3}$) such that:

$$q_o = \lambda_o \mathbf{q}_o, \quad q_c = \lambda_c \mathbf{q}_c. \tag{8}$$

Take the inner product of the vectors in (7) with $R^T(p \times \mathbf{q}_o)$:

$$q_c^T R^T(p \times \mathbf{q}_o) = (q_o - p)^T R R^T (p \times \mathbf{q}_o). \tag{9}$$

Since $p^T(p \times \mathbf{q}_o) = 0$ and $q_o^T(p \times \mathbf{q}_o) = \lambda_o \mathbf{q}_o^T(p \times \mathbf{q}_o) = 0$, the right hand side is zero. We then have:

$$\lambda_c \mathbf{q}_c^T R^T \hat{p} \mathbf{q}_o = 0 \tag{10}$$

If $\lambda_c \neq 0$, we get $\mathbf{q}_c^T R^T \hat{p} \mathbf{q}_o = 0$. ∎

Remark 1 *The Longuet-Higgins constraint (6) is also known as epipolar constraint in the computer vision literature. For the special case when $R = I$, i.e. there is only translation, the epipolar constraint simply represents the coplanar relations held by the three vectors: p, q_o and q_c.*

2.2 Characterization of the Essential Matrix

In Theorem 1 we see that the matrix which has the form $E = R^T\hat{p}$ with $R^T \in SO(3)$ and $\hat{p} \in so(3)$ plays an important role. Such a matrix is called an *essential matrix*; and the set of all essential matrices is called the *essential space*, denoted by $\mathcal{E}$:

$$\mathcal{E} \equiv \{RS \mid R \in SO(3), S \in so(3)\} \subset \mathbb{R}^{3\times 3}. \tag{11}$$

The following theorem, due to Huang and Faugeras, gives a characterization of the essential space.

Theorem 2 (Characterization of the Essential Matrix)
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:

$$\Sigma = diag\{\lambda, \lambda, 0\} \tag{12}$$

for some $\lambda \in \mathbb{R}_+$.

Proof: For the necessity, suppose $E = RS$ with $R \in SO(3)$ and $S \in so(3)$. Then $S = \hat{p}$ for some $p \in \mathbb{R}^3$. Thus $E^TE = -S^2 = -\hat{p}^2$ has three eigenvalues $\{\|p\|^2, \|p\|^2, 0\}$. For the sufficiency, knowing $E = U\Sigma V^T$, one can actually construct a rotation matrix R and a skew matrix S in terms of U, Σ and V such that $E = RS$ (see formula (23) in the following). ∎

We will study the properties of the essential matrix and solve the problem of *recovering the 3D displacement $g = (p, R)$ from a given essential matrix E*. Since the Longuet-Higgins constraint (6) is a homogeneous constraint, the essential matrix E can only be recovered up to a linear scale. As long as the translation is not zero, it is customary to set the norm of the translation vector p to be of unit length. Given the non-zero translation vector $p \in \mathbb{R}^3$, [1] define the space of normalized essential matrices, the *normalized essential space* to be:

$$\mathcal{E}_1 \equiv \{RS \mid R \in SO(3), S = \hat{p}/\|p\| \in so(3)\} \subset \mathbb{R}^{3\times 3}. \tag{13}$$

It follows that the singular values of a normalized essential matrix are $\{1, 1, 0\}$.

Now the 3D displacement estimation problem is reduced to: *given a normalized essential matrix E, find all possible solutions $g = (p, R)$ with $p \in S^2$ and $R \in SO(3)$ such that $R^T\hat{p} = E$.*

2.2.1 Uniqueness of 3D Displacement Recovery from the Essential Matrix

The answer to the question: "how many solutions g exist for a given essential matrix E?" relies on a special property of skew matrices, which we will develop now.

It will be convenient to represent a skew matrix as the product of a *unit* skew matrix and a real (non-negative) scalar. This unit skew matrix is called the associated unit skew matrix for a given

[1] The case that $p = 0$ needs to be treated separately. This is because, $p = 0$ is a singular point in the space $\mathcal{E}$ as a subset of $\mathbb{R}^{3\times 3}$ (due to this, $\mathcal{E}$ is not a regular manifold embedded in $\mathbb{R}^{3\times 3}$).

skew matrix. Given a unit skew matrix $\hat{p} \in so(3)$ with $\|p\| = 1$, and a real number $\theta \in \mathbb{R}$, we write the exponential of $\hat{p}\theta$ as:

$$e^{\hat{p}\theta} = I + \theta\hat{p} + \frac{\theta^2}{2!}\hat{p}^2 + \frac{\theta^3}{3!}\hat{p}^3 + \cdots . \tag{14}$$

It can be shown that the exponential map transforms the unit skew matrix $\hat{p}$ to a rotation matrix, *i.e.* $e^{\hat{p}\theta} \in SO(3)$, which represents the rotation about the axis p by θ radians. Actually, $e^{\hat{p}\theta}$ is given by the *Rodrigues' formula*:

$$e^{\hat{p}\theta} = I + \hat{p}\sin(\theta) + \hat{p}^2(1 - \cos(\theta)). \tag{15}$$

For details, see Murray, Li and Sastry [12].

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

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

$$R\hat{p}R = \hat{p}. \tag{16}$$

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, (16) is rewritten as:

$$e^{\hat{\omega}\theta}\hat{p}e^{\hat{\omega}\theta} = \hat{p}. \tag{17}$$

Applying this equation to ω, we get:

$$e^{\hat{\omega}\theta}\hat{p}e^{\hat{\omega}\theta}\omega = \hat{p}\omega. \tag{18}$$

Since $e^{\hat{\omega}\theta}\omega = \omega$, we obtain:

$$e^{\hat{\omega}\theta}\hat{p}\omega = \hat{p}\omega. \tag{19}$$

Since ω is the only eigenvector associated to the eigenvalue 1 of the matrix $e^{\hat{\omega}\theta}$ and $\hat{p}\omega$ is orthogonal to ω, $\hat{p}\omega$ has to be zero. Thus, ω is equal to p or $-p$. R then has the form $e^{\hat{p}\theta}$, which commutes with $\hat{p}$. Thus from (16), we get:

$$e^{2\hat{p}\theta}\hat{p} = \hat{p}. \tag{20}$$

According to *Rodrigues' formula*, we have:

$$e^{2\hat{p}\theta} = I + \hat{p}\sin(2\theta) + \hat{p}^2(1 - \cos(2\theta)) \tag{21}$$

(20) yields:

$$\hat{p}^2\sin(2\theta) + \hat{p}^3(1 - \cos(2\theta)) = 0. \tag{22}$$

Since $\hat{p}^2$ and $\hat{p}^3$ are linear independent [12], we have $\sin(2\theta) = 1 - \cos(2\theta) = 0$. That is, θ 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 $e^{\hat{p}\pi}S = -S$. ∎

Theorem 3 (Uniqueness of the Displacement Recovery from the Essential Matrix) *There exist exactly two 3D displacements $g = (p, R) \in SE(3)$ corresponding to a non-zero essential matrix $E \in \mathcal{E}$.*

Proof: Since any non-zero essential matrix is a product of a normalized essential matrix and a real scalar, we only need to prove the result for a normalized essential matrix.

For a given normalized essential matrix $E \in \mathcal{E}_1$, the existence of one solution is direct from the definition of the essential matrix. We now prove there are exactly two such displacements using Lemma 1. 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 unit skew matrices and $R_2 R_1^T$ is a rotation matrix, according to Lemma 1, we have either $(p_2, R_2) = (p_1, R_1)$ or $(p_2, R_2) = (-p_1, e^{\hat{p}_1 \pi} R_1)$. ∎

2.2.2 3D Displacement Recovery from the Essential Matrix

Although Theorem 3 claims that there are two displacements corresponding to an essential matrix, it does not give an explicit construction for the solutions. A construction based on the SVD of the essential matrix is given in Soatto *et al* [15]. However, since the SVD of an essential matrix is not unique and the unitary matrices U, V associated with the SVD are in $O(3)$ not necessarily in the rotation group $SO(3)$, the proof given in [15] is not complete. In this section, we show that such construction actually depends on a stronger characterization of the essential matrix than the one given by Huang & Faugeras (Theorem 2). Based on this characterization, we give the proof for an explicit construction of the solutions.

For the singular value decomposition (SVD) of a matrix $E = U\Sigma V^T$, the matrices U and V are unitary matrices, *i.e.* $U, V \in O(3)$. The following lemma gives a slightly stronger characterization of the SVD of essential matrices.

In both the next lemma and theorem we will use the fact that, according to the definition of the matrix $R_Z(\theta)$, for the matrix $\Sigma = diag\{\lambda, \lambda, 0\}$, $R_Z(\pm\frac{\pi}{2})\Sigma$ are skew matrices.

Lemma 2 *Consider the SVD of a non-zero essential matrix $E = U\Sigma V^T$. If $V \in SO(3)$, then so is $U \in SO(3)$, and vice versa.*

Proof: Since E is an essential matrix, there exist $R \in SO(3)$ and $S \in so(3)$ such that $RS = E$. Then $RS = U\Sigma V^T$. Multiplying both sides by $VR_Z(+\frac{\pi}{2})U^T$, we get $VR_Z(+\frac{\pi}{2})U^T RS = VR_Z(+\frac{\pi}{2})\Sigma V^T$. Since both S and $VR_Z(+\frac{\pi}{2})\Sigma V^T$ are skew matrices, according to Lemma 1, $VR_Z(+\frac{\pi}{2})U^T R = I$ or $VR_Z(+\frac{\pi}{2})U^T R = e^{S\pi}$. In both cases, taking the determinants of both sides of the equations we get $det(V)det(U^T) = +1$. Thus, $U \in SO(3)$ if and only if $V \in SO(3)$. ∎

Starting from any SVD of an essential matrix $E = U\Sigma V^T$, if $det(V) = -1$, then write $E = (-U)\Sigma(-V)^T$. Since $-V \in SO(3)$, so is $-U \in SO(3)$ according this lemma. Thus this lemma guarantees that the essential matrix E always has a SVD $E = U\Sigma V^T$ such that both unitary matrices U and V are in $SO(3)$. The lemma is important because it validates the construction of the rotation matrix in the following theorem.

Theorem 4 (Displacement Recovery from the Essential Matrix) *Given the singular value decomposition of a non-zero essential matrix E: $E = U\Sigma V^T$ with both*

$U, V \in SO(3)$, then the two 3D displacements (p, R) corresponding to E are given by:

$$\begin{aligned}(R_1^T, \hat{p}_1) &= (UR_Z^T(+\frac{\pi}{2})V^T, VR_Z(+\frac{\pi}{2})\Sigma V^T)\\ (R_2^T, \hat{p}_2) &= (UR_Z^T(-\frac{\pi}{2})V^T, VR_Z(-\frac{\pi}{2})\Sigma V^T).\end{aligned} \quad (23)$$

Proof: Lemma 2 guarantees so defined R_1, R_2 are in $SO(3)$. Since $R_Z(\pm\frac{\pi}{2})\Sigma$ are skew matrices, so are $\hat{p}_1, \hat{p}_2$. It is then direct to check that $R_i^T\hat{p}_i = E$, $i = 1, 2$. ■

Although the matrices U and V are not unique in the SVD of an essential matrix (for example, they may differ by an arbitrary $R_Z(\theta)$), Theorem 3 guarantees that these variations give the same pair of solutions as given in (23).

2.2.3 Characterization of the Normalized Essential Space

In this section, we study the geometric properties of the normalized essential space. Readers who are not familiar with differential geometry may simply skip this section, without loss of continuity.

It is known from differential geometry that the set of tangent vectors (the tangent space) of the rotation group $SO(3)$ at the identity element e, is the set of skew matrices $so(3)$ (see Boothby [1]), that is:

$$T_e(SO(3)) = so(3). \quad (24)$$

Since $SO(3)$ is a Lie group, the tangent space at any point $R \in SO(3)$ is given by the push-forward map:

$$T_R(SO(3)) = R_*(so(3)) = \{RS | S \in so(3)\}. \quad (25)$$

It is direct to check that the push-forward map is an isomorphism between $so(3)$ and $T_R(SO(3))$. Thus, the tangent bundle (for a reference on fiber bundle theory see Steenrod [16]) of the rotation group $SO(3)$, defined to be:

$$T(SO(3)) = \bigcup_{R \in SO(3)} T_R(SO(3)) \quad (26)$$

is equivalent to the space $SO(3) \times so(3)$ (Lie groups have trivial tangent bundles), which, in turn, can be used to represent the configuration space of a rigid body motion (the base point $R \in SO(3)$ represents for the rotation and the tangent vector $\hat{p} \in so(3)$ for the translation). We are interested in the case when the translation vector of the rigid body motion is of unit length. Define the unit tangent bundle of the rotation group $SO(3)$ to be:

$$T_1(SO(3)) = \bigcup_{R \in SO(3)} \{RS | S = \hat{p}, < p, p >= 1\}. \quad (27)$$

where $< \cdot, \cdot >$ is any invariant Riemannian metric on $SO(3)$.[2]

We then have:

[2]Since $SO(3)$ is a compact Lie group, there exists an invariant metric (there are infinitely many of such invariant metrics but they are essentially the same — only differ by constant scalars). Using any such metric, the unit tangent bundle is well-defined.

Theorem 5 (Characterization of the Normalized Essential Space) *The unit tangent bundle of the rotation group $SO(3)$, i.e. $T_1(SO(3))$, is a double covering of the normalized essential space $\mathcal{E}_1$, or equivalently speaking, $\mathcal{E}_1 = T_1(SO(3))/\mathbb{Z}_2$.*

Proof: Consider the projection from $T_1(SO(3))$ to $\mathcal{E}_1$:

$$\begin{aligned} h : T_1(SO(3)) &\rightarrow \mathcal{E}_1 \\ (R, RS) \in T_1(SO(3)) &\mapsto RS \in \mathcal{E}_1. \end{aligned} \tag{28}$$

According to the proof of Theorem 3:

$$\begin{aligned} h^{-1}(RS) &= \{(R, RS),\ (Re^{-S\pi}, Re^{-S\pi}(e^{S\pi}S))\} \\ &= \{(R, RS),\ (Re^{-S\pi}, RS)\}. \end{aligned} \tag{29}$$

Thus, $T_1(SO(3))$ is a double covering of $\mathcal{E}_1$ with the covering map h. ∎

This theorem gives a global characterization for the local coordinates assigned by Soatto *et al* in [15] to the essential space. Since $T_1(SO(3))$ is a 5-dimensional connected compact manifold, we thus also have:

Corollary 1 (Regularity of the Normalized Essential Space) *The normalized essential space $\mathcal{E}_1$ is a 5-dimensional connected compact differentiable manifold embedded in $\mathbb{R}^{3\times 3}$.*

Proof: The compactness is trivial. That $\mathcal{E}_1$ is embedded in $\mathbb{R}^{3\times 3}$ follows from the fact that, if $F : N \rightarrow M$ is a one-to-one immersion and N is compact, then F is an embedding (see Boothby [1]). ∎

This property validates potential motion estimation algorithms which might require certain smoothness or regularity on the space of the parameters to be estimated.

2.3 Algorithm

In this section, we introduce the three-step SVD-based 3D displacement estimation algorithm proposed by Toscani and Faugeras [19]. For a detailed proof of this algorithm see Maybank [10].

Let $E = R^T\hat{p} \in \mathbb{R}^{3\times 3}$ be the essential matrix associated with the Longuet-Higgins constraint (6). The 3×3 matrix E has the form:

$$E = \begin{pmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{pmatrix}. \tag{30}$$

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

$$\mathbf{e} = (e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9)^T. \tag{31}$$

Define a vector $\mathbf{a} \in \mathbb{R}^9$ associated to each pair of image correspondence $\mathbf{q}_o = (x_o, y_o, z_o)^T \in \mathbb{R}^3$, $\mathbf{q}_c = (x_c, y_c, z_c)^T \in \mathbb{R}^3$ to be:

$$\mathbf{a} = (x_c x_o, x_c y_o, x_c z_o, y_c x_o, y_c y_o, y_c z_o, z_c x_o, z_c y_o, z_c z_o)^T. \tag{32}$$

We thus have

$$\mathbf{q}_c^T E \mathbf{q}_o = \mathbf{a}^T \mathbf{e}. \tag{33}$$

Then the Longuet-Higgins constraint can be rewritten as $\mathbf{a}^T\mathbf{e} = 0$. Given a set of (possibly noisy) image correspondences: $\mathbf{q}_o^i, \mathbf{q}_c^i$, $i = 1, \dots, m$ between two images, vectors $\mathbf{a}^i$, $i = 1, \dots, m$ are then defined for each pair $(\mathbf{q}_o^i, \mathbf{q}_c^i)$ using (32). Define the matrix $A \in \mathbb{R}^{m\times 9}$ associated with these measurements to be:

$$A = (\mathbf{a}^1, \mathbf{a}^2, \dots, \mathbf{a}^m)^T. \tag{34}$$

If there is no noise on these measurements, the essential vector $\mathbf{e}$ has to satisfy:

$$A\mathbf{e} = 0. \tag{35}$$

From this equation we see that, in order to have a unique solution for $\mathbf{e}$, the rank of the matrix A has to be eight. Thus, *at least eight pairs of image correspondences are needed for this algorithm to recover the 3D displacement, i.e.* $m \geq 8$.

However, since the measurements are usually noisy, there might not exist any solution of $\mathbf{e}$ for $A\mathbf{e} = 0$. In this case, we use the one which minimizes the error function $\|A\mathbf{e}\|^2$.

It is straight forward to know that (Theorem 6.1 of Maybank [10]):

Lemma 3 *If a matrix $A \in \mathbb{R}^{n\times n}$ has the singular value decomposition $A = U\Sigma V^T$ and $c_n(V)$ is the n^{th} column vector of V (the singular vector associated to the smallest singular value σ_n), then $\mathbf{e} = c_n(V)$ minimizes $\|A\mathbf{e}\|^2$ subject to the condition $\|\mathbf{e}\| = 1$.*

The matrix E which is reconstructed from such optimal vector $\mathbf{e}$ is not yet guaranteed to be in the essential space $\mathcal{E}$ yet. Let $\|\cdot\|_f$ stand for the Frobenius norm. The following theorem gives a natural projection onto the essential space.

Theorem 6 (Projection to the Essential Space)
If the SVD of a matrix $F \in \mathbb{R}^{3\times 3}$ is given by $F = U\mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\}V^T$ with $\sigma_1 \geq \sigma_2 > \sigma_3$, then the essential matrix $E \in \mathcal{E}$ which minimizes the Frobenius distance $\|E - F\|_f^2$ is given by $E = U\mathrm{diag}\{\lambda, \lambda, 0\}V^T$ where $\lambda = (\sigma_1 + \sigma_2)/2$.

For the proof of this theorem see Maybank [10].

We are finally ready to give the algorithm.

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

1. **Estimate Essential Vector:**
   For a given set of image correspondences: $(\mathbf{q}_o^i, \mathbf{q}_c^i)$, $i = 1, \dots, m$ $(m \geq 8)$, find the vector $\mathbf{e}$ which minimizes the error function:

   $$V(\mathbf{e}) = \|A\mathbf{e}\|^2 \tag{36}$$

   subject to the condition $\|\mathbf{e}\| = 1$, where A is the matrix associated with image measurements;

2. **Singular Value Decomposition:**
   Recover matrix E from $\mathbf{e}$ and find the singular value decomposition of the matrix E:
   $$E = U diag\{\sigma_1, \sigma_2, \sigma_3\} V^T \tag{37}$$
   where $\sigma_1 \geq \sigma_2 \geq \sigma_3$;

3. **Recover Displacement from the Essential Matrix:**
   Define the diagonal matrix Σ to be:
   $$\Sigma = diag\{1, 1, 0\}. \tag{38}$$
   Then the 3D displacement (p, R) is given by:
   $$R^T = U R_Z^T(\pm\frac{\pi}{2}) V^T, \quad \hat{p} = V R_Z(\pm\frac{\pi}{2}) \Sigma V^T. \tag{39}$$

■

Note in step 2, the matrix E reconstructed directly from $\mathbf{e}$ has Frobenius norm 1. We know normalized essential matrices have Frobenius norm $\sqrt{2}$, instead of 1. However, this causes no trouble for the above algorithm because essential matrices differed by a pure scalar have the same projection on the normalized essential space $\mathcal{E}_1$.

Remark 2 *Note that if $E \in \mathcal{E}_1$ satisfies the Longuet-Higgins constraint, so does $-E \in \mathcal{E}_1$. This introduces the so-called "twisted-pair" ambiguity (Maybank [10]). Thus, totally, we get four ambiguous solutions of the 3D displacement for a given set of image correspondences. However, after imposing the so-called "positive depth constraint", that all the points lie in front of the camera, the ambiguities will be resolved and there is only one best solution.*

Note the essential vector $\mathbf{e}$ is actually parameterized by p and R, *i.e.* we should write $\mathbf{e}(p, R)$. In general, the above algorithm, unfortunately, does not necessarily give the optimal solution (p^*, R^*) which minimizes the originally picked error function $\|A\mathbf{e}(p, R)\|^2$ among all possible pairs (p, R), *i.e.* on $\mathcal{E}_1$. This problem is indeed an optimization problem on $\mathcal{E}_1$, which is not a Euclidean space. Taking the Riemannian metric induced from $T_1(SO(3))$ for $\mathcal{E}_1$, it is then converted to an optimization problem on a Riemannian manifold, a topic which has been studied by Smith, Brockett *et al* [14].

A dynamic version of this algorithm which recursively estimates the essential vector $\mathbf{e}$ using implicit extended Kalman filter, the so-called *essential filter*, has been proposed by Soatto [15]. A big advantage of the dynamic approach is that the algorithm is not as sensitive to noise since it uses a sequence of images instead of only two images to recover the 3D motion.

3 Differential Essential Matrix Approach

The differential case is the infinitesimal version of the discrete case. To reveal the generic similarities between these two cases, we now develop the *differential essential matrix approach* for estimating 3D velocity from optical flow in a parallel way as we did in last section for the discrete essential matrix approach for estimating 3D displacement from image correspondences.

After deriving the differential version of the Longuet-Higgins constraint, the concept of differential essential matrix is defined; we then give a thorough characterization for 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, a four-step eigenvector-decomposition-based 3D velocity estimation algorithm is proposed; at last, we discuss the reasons why the zero-translation case makes all essential constraint based motion estimation algorithms fail to work and suggest possible ways to overcome this trouble.

3.1 Differential Longuet-Higgins Constraint

Suppose the motion of the camera is described by a smooth curve $g(t) = (p(t), R(t)) \in SE(3)$. According to (5), for a point q attached to the inertial frame F_o, its coordinates in the inertial frame and the moving camera frame satisfy:

$$q_o = R(t)q_c(t) + p(t). \tag{40}$$

Differentiating this equation yields:

$$\dot{q}_c = -R^T\dot{R}q_c - R^T\dot{p}. \tag{41}$$

Since $-R^T\dot{R} \in so(3)$ and $-R^T\dot{p} \in \mathbb{R}^3$ (see Murray *et al* [12]), 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:

$$\hat{\omega} = -R^T\dot{R}, \quad v = -R^T\dot{p}. \tag{42}$$

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:

$$\dot{q}_c = \hat{\omega}q_c + v. \tag{43}$$

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 $\mathbf{u}$, $\mathbf{u} = \dot{\mathbf{q}} \in \mathbb{R}^3$.

Theorem 7 (Differential Longuet-Higgins Constraint)
Consider a camera moving with linear velocity v and angular velocity ω with respect to the inertial frame. Then the optical flow $\mathbf{u}$ at an image point $\mathbf{q}$ satisfies:

$$\mathbf{u}^T\hat{v}\mathbf{q} + \mathbf{q}^T\hat{\omega}\hat{v}\mathbf{q} \equiv 0 \tag{44}$$

or in an equivalent form:

$$(\mathbf{u}^T, \mathbf{q}^T)\begin{pmatrix} \hat{v} \\ s \end{pmatrix}\mathbf{q} = 0 \tag{45}$$

where s is a symmetric matrix defined to be $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \in \mathbb{R}^{3\times3}$.

Proof: From the definition of the map π's, there exists a real scalar function $\lambda(t)$ ($\|q(t)\|$ or $q_3(t)$, depending on the projection type) such that:

$$q = \lambda \mathbf{q}. \tag{46}$$

Take the inner product of the vectors in (43) with $(v \times \mathbf{q})$:

$$\dot{q}^T(v \times \mathbf{q}) = (\hat{\omega} q + v)^T (v \times \mathbf{q}) = q^T \hat{\omega}^T \hat{v} \mathbf{q}. \tag{47}$$

Since $\dot{q} = \dot{\lambda}\mathbf{q} + \lambda\dot{\mathbf{q}}$ and $\mathbf{q}^T(v \times \mathbf{q}) = 0$, from (47) we then have:

$$\lambda \dot{\mathbf{q}}^T \hat{v} \mathbf{q} - \lambda \mathbf{q}^T \hat{\omega}^T \hat{v} \mathbf{q} = 0. \tag{48}$$

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

$$\mathbf{u}^T \hat{v} \mathbf{q} + \mathbf{q}^T \hat{\omega} \hat{v} \mathbf{q} \equiv 0 \tag{49}$$

Due to the following Lemma 4, for any skew symmetric matrix $A \in \mathbb{R}^{3\times3}$, $\mathbf{q}^T A \mathbf{q} = 0$. 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:

$$\mathbf{u}^T \hat{v} \mathbf{q} + \mathbf{q}^T s \mathbf{q} \equiv 0. \tag{50}$$

■

The proof indicates that there is some redundancy in the expression of the differential Longuet-Higgins constraint (44). The following lemma shows where this redundancy comes from.

Lemma 4 *Consider matrices $M_1, M_2 \in \mathbb{R}^{3\times3}$. $\mathbf{q}^T M_1 \mathbf{q} = \mathbf{q}^T M_2 \mathbf{q}$ for all $\mathbf{q} \in \mathbb{R}^3$ if and only if $M_1 - M_2$ is a skew matrix, i.e. $M_1 - M_2 \in so(3)$.*

Proof: The sufficiency of $M_1 - M_2$ being a skew matrix is trivial. We only need to prove the necessity. Given $\mathbf{q}^T M_1 \mathbf{q} = \mathbf{q}^T M_2 \mathbf{q}$ for all $\mathbf{q} \in \mathbb{R}^3$, we have, for all $\mathbf{q} \in \mathbb{R}^3$:

$$\mathbf{q}^T (M_1 - M_2) \mathbf{q} = 0. \tag{51}$$

Let $M = M_1 - M_2$. Then:

$$\mathbf{q}^T M \mathbf{q} \equiv 0. \tag{52}$$

Suppose:

$$M = \begin{pmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & m_9 \end{pmatrix}. \tag{53}$$

Substituting $\mathbf{q} = (1,0,0)^T$, $(0,1,0)^T$ and $(0,0,1)^T$ into (52) respectively, we obtain $m_1 = m_5 = m_9 = 0$. Then:

$$M = \begin{pmatrix} 0 & m_2 & m_3 \\ m_4 & 0 & m_6 \\ m_7 & m_8 & 0 \end{pmatrix}. \tag{54}$$

Now substituting $\mathbf{q} = (1,1,0)^T$, $(1,0,1)^T$ and $(0,1,1)^T$ into (52) respectively, we obtain $m_2 = -m_4$, $m_3 = -m_7$ and $m_6 = -m_8$. Thus, $M = M_1 - M_2$ is a skew matrix. ∎

Let us define an equivalence relation on the space $\mathbb{R}^{3\times3}$, the space of 3×3 matrices over $\mathbb{R}$: for $x, y \in \mathbb{R}^{3\times3}$, $x \sim y$ if and only if $x - y \in so(3)$. Denote by $[x] = \{y \in \mathbb{R}^{3\times3} \,|\, y \sim x\}$ the equivalence class of x, and denote by $[X]$ the set $\bigcup_{x\in X}[x]$. The quotient space $\mathbb{R}^{3\times3}/\sim$ can be naturally identified with the space of all 3×3 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 (45).

Using this notation, Theorem 7 can then be re-expressed in the following way:

Corollary 2 *Consider a camera undergoing a smooth rigid body motion with linear velocity v and angular velocity ω. Then the optical flow $\mathbf{u}$ of a image point $\mathbf{q}$ satisfies:*

$$(\mathbf{u}^T, \mathbf{q}^T)\begin{pmatrix} \hat{v} \\ [\hat{\omega}\hat{v}] \end{pmatrix}\mathbf{q} \equiv 0. \tag{55}$$

Because of this redundancy, each equivalence class $[\hat{\omega}\hat{v}]$ can only be recovered up to its symmetric component $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \in [\hat{\omega}\hat{v}]$.

3.2 Characterization of the Differential Essential Matrix

We define the space of 6×3 matrices given by:

$$\mathcal{E}' = \left\{ \begin{pmatrix} \hat{v} \\ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \end{pmatrix} \middle| \omega, v \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{6\times3}. \tag{56}$$

to be the *differential essential space*. A matrix in this space is called a *differential essential matrix*. Note that the differential Longuet-Higgins constraint (45) 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*:

$$\mathcal{E}'_1 = \left\{ \begin{pmatrix} \hat{v} \\ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \end{pmatrix} \middle| \omega \in \mathbb{R}^3, v \in S^2 \right\} \subset \mathbb{R}^{6\times3}. \tag{57}$$

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*:

$$\mathcal{S} = \left\{ \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) \middle| \omega \in \mathbb{R}^3, v \in S^2 \right\} \subset \mathbb{R}^{3\times3}. \tag{58}$$

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 (ω, 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: $\hat{\omega}\hat{v} \in \mathbb{R}^{3\times3}$, 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 unitary matrices U and V are related.

Lemma 5 *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:*

$$Q = -VR_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\}V^T \tag{59}$$

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 $q \in \mathbb{R}^3$,

$$\hat{\omega}\hat{v}q = \omega \times (v \times q). \tag{60}$$

Let $b \in S^2$ be the unit vector perpendicular to both ω 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 ω,then the rest of the proof still holds). Then $\omega = \lambda exp(\hat{b}\theta)v$ (according this definition, θ is the angle between ω and v, and $0 \leq \theta \leq \pi$). It is direct to check that if the matrix V is defined to be:

$$V = (e^{\hat{b}\frac{\pi}{2}}v, b, v) \tag{61}$$

Q has the given form (59).

We now prove the sufficiency. Given a matrix Q which can be decomposed in the form (59), define the unitary matrix $U = -VR_Y(\theta) \in O(3)$. Let the two skew matrices $\hat{\omega}$ and $\hat{v}$ given by the formulae:

$$\hat{\omega} = UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T, \quad \hat{v} = VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \tag{62}$$

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

$$\begin{aligned}
\hat{\omega}\hat{v} &= UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \\
&= UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda(-R_Y^T(\theta))R_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \\
&= Udiag\{\lambda, \lambda\cos(\theta), 0\}V^T \\
&= Q.
\end{aligned} \tag{63}$$

Since ω and v have to be, respectively, the left and the right zero eigenvectors of Q, the reconstruction given in (62) is unique. ∎

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

Theorem 8 (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:

$$\Sigma = diag\{\sigma_1, \sigma_2, \sigma_3\} \tag{64}$$

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 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 5, so are θ and γ. Then according to the lemma

$$\hat{\omega}\hat{v} = -VR_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\}V^T. \tag{65}$$

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

$$s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) = \frac{1}{2}V\left(-R_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\} - diag\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta)\right)V^T. \tag{66}$$

Define the matrix $D(\lambda, \theta) \in \mathbb{R}^{3\times3}$ to be

$$\begin{aligned} D(\lambda, \theta) &= -R_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\} - diag\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta) \\ &= \lambda\begin{pmatrix} -2\cos(\theta) & 0 & \sin(\theta) \\ 0 & -2\cos(\theta) & 0 \\ \sin(\theta) & 0 & 0 \end{pmatrix}. \end{aligned} \tag{67}$$

Directly calculating its eigenvalues and eigenvectors, we obtain that

$$D(\lambda, \theta) = R_Y\left(\frac{\theta}{2} - \frac{\pi}{2}\right) diag\left\{\lambda(1-\cos(\theta)), -2\lambda\cos(\theta), \lambda(-1-\cos(\theta))\right\} R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right). \tag{68}$$

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

$$\left\{\frac{1}{2}\lambda(1-\cos(\theta)), \quad -\lambda\cos(\theta), \quad \frac{1}{2}\lambda(-1-\cos(\theta))\right\}, \tag{69}$$

which satisfy the given conditions.

We now prove the sufficiency. Given $s = V_1 diag\{\sigma_1, \sigma_2, \sigma_3\}V_1^T$ 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 σ_i's have the form given in (69):

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

Define a matrix $V \in SO(3)$ to be $V = V_1 R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right)$. Then $s = \frac{1}{2}VD(\lambda, \theta)V^T$. According to Lemma 5, there exist vectors $v \in S^2$ and $\omega \in \mathbb{R}^3$ such that

$$\hat{\omega}\hat{v} = -VR_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\}V^T. \tag{70}$$

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

Figure 2 gives a geometric interpretation of the three eigenvectors of the special symmetric matrix s for the case when both $\omega, v \in S^2$.

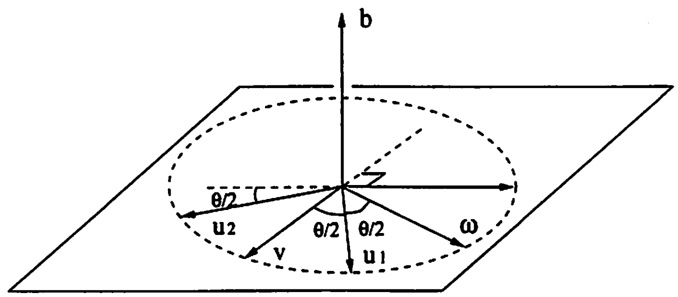

Figure 2: Eigenvectors u_1, u_2, b of a special symmetric matrix: $\frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$.

3.2.1 Uniqueness of 3D Velocity Recovery from the Special Symmetric Matrix

According to the proof of Theorem 8, if we already know the eigenvector decomposition of a special symmetric matrix s, we certainly can find some velocity (ω, 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 $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$.

Theorem 9 (Uniqueness of the Velocity Recovery from the Special Symmetric Matrix) *There exist exactly four 3D velocities (ω, 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 (ω_1, v_1) and (ω_2, v_2) are both solutions for $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$, we have:

$$\hat{v}_1\hat{\omega}_1 + \hat{\omega}_1\hat{v}_1 = \hat{v}_2\hat{\omega}_2 + \hat{\omega}_2\hat{v}_2. \tag{71}$$

From Lemma 5, we may write:

$$\begin{aligned} \hat{\omega}_1\hat{v}_1 &= -V_1 R_Y(\theta_1) diag\{\lambda_1, \lambda_1\cos(\theta_1), 0\} V_1^T \\ \hat{\omega}_2\hat{v}_2 &= -V_2 R_Y(\theta_2) diag\{\lambda_2, \lambda_2\cos(\theta_2), 0\} V_2^T. \end{aligned} \tag{72}$$

Let $W = V_1^T V_2 \in SO(3)$, then from (71):

$$D(\lambda_1, \theta_1) = W D(\lambda_2, \theta_2) W^T. \tag{73}$$

Since both sides of (73) have the same eigenvalues, according to (68), we have:

$$\lambda_1 = \lambda_2, \quad \theta_2 = \theta_1. \tag{74}$$

We then can denote both θ_1 and θ_2 by θ. It is direct to check that the only possible rotation matrix W which satisfies (73) is given by $I_{3\times 3}$ or:

$$\begin{pmatrix} -\cos(\theta) & 0 & \sin(\theta) \\ 0 & -1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \quad or \quad \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & -1 & 0 \\ -\sin(\theta) & 0 & -\cos(\theta) \end{pmatrix}. \tag{75}$$

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 5, if (ω, v) is one solution and $\hat{\omega}\hat{v} = U diag\{\lambda, \lambda\cos(\theta), 0\} V^T$, then all the solutions are given by:

$$\begin{aligned} \hat{\omega} &= U R_Z(\pm\frac{\pi}{2}) \Sigma_\lambda U^T, \quad \hat{v} = V R_Z(\pm\frac{\pi}{2}) \Sigma_1 V^T; \\ \hat{\omega} &= V R_Z(\pm\frac{\pi}{2}) \Sigma_\lambda V^T, \quad \hat{v} = U R_Z(\pm\frac{\pi}{2}) \Sigma_1 U^T \end{aligned} \tag{76}$$

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

3.2.2 3D Velocity Recovery from Differential Essential Matrix

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

Theorem 10 (Uniqueness of the Velocity Recovery from Differential Essential Matrix) *There exists only one 3D velocity (ω, 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 twist-pair ambiguity (see Maybank [10]), which is caused by a 180° rotation of the camera around the translation direction, is avoided.

It is clear that the normalized differential essential space $\mathcal{E}_1'$ is a 5-dimensional differentiable submanifold embedded in $\mathbb{R}^{6\times3}$. Further considering the symmetric and anti-symmetric structures in the differential essential matrix, the embedding space can be naturally reduced from $\mathbb{R}^{6\times3}$ to $\mathbb{R}^9$. This property is useful when using estimation schemes which require some regularity of the parameter space (for example, the dynamic estimation scheme proposed by Soatto *et al* [15]).

3.3 Algorithm

Based on the previous study on the differential essential matrix, in this section, we propose an 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 Longuet-Higgins constraint (45). Since the submatrix $\hat{v}$ is skew symmetric and s is symmetric, they have the following forms:

$$v = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix}, \quad s = \begin{pmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{pmatrix}. \tag{77}$$

Similar to the discrete case, define the (differential) *essential vector* $\mathbf{e} \in \mathbb{R}^9$ to be:

$$\mathbf{e} = (v_1, v_2, v_3, s_1, s_2, s_3, s_4, s_5, s_6)^T. \tag{78}$$

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[3]:

$$\mathbf{a} = (u_3 y - u_2 z, u_1 z - u_3 x, u_2 x - u_1 y, x^2, 2xy, 2xz, y^2, 2yz, z^2)^T. \tag{79}$$

The differential Longuet-Higgins constraint (45) can be then rewritten as:

$$\mathbf{a}^T \mathbf{e} = 0. \tag{80}$$

[3]For perspective projection, $z = 1$ and $u_3 = 0$ thus the expression for $\mathbf{a}$ can be simplified.

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

$$A = (\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^m)^T \tag{81}$$

where $\mathbf{a}^i$ are defined for each pair $(\mathbf{q}^i, \mathbf{u}^i)$ using (79). In the absence of noise, the essential vector $\mathbf{e}$ has to satisfy:

$$A\mathbf{e} = 0. \tag{82}$$

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 [10]).

When the measurements are noisy, there might be no solution of $\mathbf{e}$ for $A\mathbf{e} = 0$. As in the discrete case, we choose the solution which minimizes the error function $\|A\mathbf{e}\|^2$. This can be mechanized using Lemma 3.

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 [22, 23], such s, with the linear velocity v obtained from the skew-symmetric part, is directly used to calculate the angular velocity ω. This is a 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 ω.

We thus propose a different approach: first extract the special symmetric component from the first-hand symmetric matrix s; then recover the four possible solutions for the 3D velocity using the results obtained in Theorem 9; 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 $\mathcal{S}$.

Theorem 11 (Projection to the Special Symmetric Space)
If a symmetric matrix $F \in \mathbb{R}^{3\times 3}$ is diagonalized as $F = V diag\{\lambda_1, \lambda_2, \lambda_3\}V^T$ 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 $E = V diag\{\sigma_1, \sigma_2, \sigma_2\}V^T$ with:

$$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \quad \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \quad \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}. \tag{83}$$

Proof: Define $\mathcal{S}_\Sigma$ to be the subspace of $\mathcal{S}$ whose elements have the same eigenvalues: $\Sigma = diag\{\sigma_1, \sigma_2, \sigma_3\}$. 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 $\Sigma_\lambda = diag\{\lambda_1, \lambda_2, \lambda_3\}$. We now prove this theorem by two steps.

Step One: 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:

$$\begin{aligned} \|E - F\|_f^2 &= \|V_1\Sigma V_1^T - V\Sigma_\lambda V^T\|_f^2 \\ &= \|\Sigma_\lambda - V^T V_1 \Sigma V_1^T V\|_f^2. \end{aligned} \tag{84}$$

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

$$W = \begin{pmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{pmatrix}. \tag{85}$$

Then:

$$\begin{aligned} \|E - F\|_f^2 &= \|\Sigma_\lambda - W\Sigma W^T\|_f^2 \\ &= tr(\Sigma_\lambda^2) - 2tr(W\Sigma W^T \Sigma_\lambda) + tr(\Sigma^2). \end{aligned} \tag{86}$$

Substituting (85) into the second term, and using the fact that $\sigma_2 = \sigma_1 + \sigma_3$ and W is a rotation matrix, we get:

$$\begin{aligned} tr(W\Sigma W^T \Sigma_\lambda) &= \sigma_1(\lambda_1(1 - w_3^2) + \lambda_2(1 - w_6^2) + \lambda_3(1 - w_9^2)) \\ &+ \sigma_3(\lambda_1(1 - w_1^2) + \lambda_2(1 - w_4^2) + \lambda_3(1 - w_7^2)). \end{aligned} \tag{87}$$

Minimizing $\|E - F\|_f^2$ is equivalent to maximizing $tr(W\Sigma W^T \Sigma_\lambda)$. From (87), $tr(W\Sigma W^T \Sigma_\lambda)$ 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 Two: 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:

$$\|E - F\|_f^2 = (\lambda_1 - \sigma_1)^2 + (\lambda_2 - \sigma_2)^2 + (\lambda_3 - \sigma_3)^2 \tag{88}$$

subject to the constraint:

$$\sigma_2 = \sigma_1 + \sigma_3. \tag{89}$$

The formula (83) for $\sigma_1, \sigma_2, \sigma_3$ are directly obtained from solving this minimization problem. ∎

Remark 3 *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 eigenvector-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: $(\mathbf{q}^i, \mathbf{u}^i)$, $i = 1, \ldots, m$, find the vector $\mathbf{e}$ which minimizes the error function:

   $$V(\mathbf{e}) = \|A\mathbf{e}\|^2 \tag{90}$$

   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 $s \in \mathbb{R}^{3\times3}$ from the remaining six entries.[4] Find the eigenvalue decomposition of the symmetric matrix s:

$$s = V_1 diag\{\lambda_1, \lambda_2, \lambda_3\} V_1^T \tag{91}$$

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 $s = V_1 diag\{\sigma_1, \sigma_2, \sigma_3\} V_1^T$ with:

$$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \quad \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \quad \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}; \tag{92}$$

3. **Recover Velocity from the Special Symmetric Matrix:**
Define:

$$\begin{aligned} \lambda &= \sigma_1 - \sigma_3, \quad \lambda \geq 0 \\ \theta &= \arccos(-\sigma_2/\lambda), \quad \theta \in [0, \pi]. \end{aligned} \tag{93}$$

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

$$\begin{aligned} \hat{\omega} &= UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T, \quad \hat{v} = VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \\ \hat{\omega} &= VR_Z(\pm\frac{\pi}{2})\Sigma_\lambda V^T, \quad \hat{v} = UR_Z(\pm\frac{\pi}{2})\Sigma_1 U^T \end{aligned} \tag{94}$$

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

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 (ω^*, v^*) which satisfies:

$$v^{*T} v_0 = \max_i v_i^T v_0. \tag{95}$$

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

$$\omega = \omega^*, \quad v = v_0. \tag{96}$$

■

Both v_0 and v^* contain recovered information about the linear velocity. However, experimental results show that, statistically, within the tested noise levels (next section), v_0 always yields a better estimate than v^* . We thus simply choose v_0 as the estimate. Nonetheless, one can find statistical correlations between v_0 and v^* (experimentally or analytically) and obtain better estimate, 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 [6] and an algorithm was proposed to remove such bias from Zhuang's algorithm [22].

[4] In order to guarantee v_0 to be of unit length, one needs to "re-normalize" $\mathbf{e}$, *i.e.* multiply $\mathbf{e}$ by a scalar such that the vector determined by the first three entries is of unit length.

Remark 4 *Since both $E, -E \in \mathcal{E}_1'$ satisfy the same set of differential Longuet-Higgins 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 5 *By the way of comparison to the Heeger and Jepson's algorithm [4], note that the equation (82) may be rewritten to highlight the dependence on optical flow as:*

$$[A_1(\mathbf{u}) \mid A_2]\mathbf{e} = 0 \tag{97}$$

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 ω. Our method simultaneously estimates $v \in \mathbb{R}^3, s \in \mathbb{R}^6$. We make a simulation comparison of these two algorithms in section 4. It should be noted that the noise performance of our algorithm can be substantially improved by replacing the estimate of Lemma 2 for $\mathbf{e}$ as the smallest right singular vector of A, by $\mathbf{e}$ as the smallest right **structured** *singular vector of A ([13]) accounting for the noise in only the first 3 columns of A. This will be implemented in future.*

Due to the same arguments for the discrete case, this algorithm is not optimal in the sense that the recovered velocity does not necessarily minimizes the originally picked error function $\|A\mathbf{e}(\omega, v)\|^2$ on $\mathcal{E}_1'$. However, this algorithm only uses linear algebra techniques and is simpler than a one which tries to optimize on the submanifold $\mathcal{E}_1'$.

3.4 Difficulties with Zero Translational Velocity

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 Longuet-Higgins 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 $\mathbf{q}_c = R^T\mathbf{q}_o$, it is straight forward to check that, for all $p \in S^2$, we have:

$$\mathbf{q}_c^T R^T \hat{p} \mathbf{q}_o \equiv 0. \tag{98}$$

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 Longuet-Higgins constraints). In the differential case, we have a similar situation:

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

$$(\mathbf{u}^T, \mathbf{q}^T)\begin{pmatrix} \hat{v} \\ [\hat{\omega}\hat{v}] \end{pmatrix}\mathbf{q} = 0. \tag{99}$$

Proof:

$$\begin{aligned} & \mathbf{u} = \hat{\omega}\mathbf{q} \quad \text{since } \mathbf{u} \text{ is obtained from rotation } \omega \\ \Leftrightarrow \quad & \mathbf{u}^T(v \times \mathbf{q}) = -\mathbf{q}^T\hat{\omega}(v \times \mathbf{q}) \quad \text{for all } v \in S^2 \\ \Leftrightarrow \quad & (\mathbf{u}^T, \mathbf{q}^T)\begin{pmatrix} \hat{v} \\ [\hat{\omega}\hat{v}] \end{pmatrix}\mathbf{q} = 0. \end{aligned} \tag{100}$$

■

This lemma implies that the velocity estimation algorithm proposed in the previous section will have trouble when the linear velocity v is zero. There are infinite many pairs of (ω, v) satisfying the same set of differential Longuet-Higgins constraints.

However, it is shown by Soatto *et al* [15] 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 the scale or accuracy of estimation of 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 ω is still possible using dynamic estimation schemes.

4 Applications and Experimental Results

In this section, we discuss some applications of the proposed motion estimation algorithm. Simulation results are given for evaluating the performance of the algorithm.

4.1 Exploiting Nonholonomic Constraints

Usually, the kinematics of the mobile base on where the computer vision is mounted satisfy some so-called nonholonomic constraints. Roughly speaking, these constraints confine the infinitesimal 3D motion but not the global motion of the mobile base (as an example of studying vision with nonholonomic constraints see Ma, Košecká and Sastry [9]). A big advantage of the differential essential approach over the discrete one is that it can make use of these nonholonomic constraints and much simplify the motion estimation algorithms. We show this by an example of a kinematic aircraft.

Example: Kinematic Model of an Aircraft

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 *homogeneous representation* for g (for a good reference on homogeneous representation see Murray, Li and Sastry [12]), the kinematic equations of the aircraft motion are given by:

$$\dot{g} = g \begin{pmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & 0 \\ -\omega_2 & \omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{101}$$

where ω_1 stands for pitch rate, ω_2 for roll rate, ω_3 for yaw rate and v_1 the velocity of the aircraft.

Then the 3D velocity (ω, v) in the differential Longuet-Higgins constraint (45) has the form:

$$\omega = (\omega_1, \omega_2, \omega_3)^T, \quad v = (v_1, 0, 0)^T. \tag{102}$$

For the algorithm given in section 3.3, 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:

$$\mathbf{e} = (v_1, s_2, s_3, s_4)^T. \tag{103}$$

Then the measurement vector $\mathbf{a} \in \mathbb{R}^4$ is to be:

$$\mathbf{a} = (u_3 y - u_2 z, 2xy, 2xz, y^2 + z^2)^T. \tag{104}$$

The differential Longuet-Higgins constraint can then be rewritten as:

$$\mathbf{a}^T \mathbf{e} = 0. \tag{105}$$

If we define the matrix A as in (81), the matrix $A^T A$ is a 4×4 matrix rather than a 9×9 one. For estimating the velocity (ω, 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 3.3 are 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.

4.2 Camera Self-calibration

So far, we have assumed that the camera is ideal, *i.e.* the formation of the image is through ideal perspective or spherical projections of unit focal length. For a more general camera model, the relation between an actual image and the ideal image is given by a linear transformation specified by the so-called *intrinsic-parameter matrix* of the camera:

$$A = \begin{pmatrix} s_1 & 0 & -s_1 i_1 \\ 0 & s_2 & -s_1 i_2 \\ 0 & 0 & -f \end{pmatrix}. \tag{106}$$

For a general treatment of the camera self-calibration in the differential case, one should refer to Vieville and Faugeras [21] and Brooks *et al* [2]. Here we assume A is time-invariant. Then, it can be shown that the differential Longuet-Higgins constraint now becomes

$$\mathbf{u}^T A^T \hat{v} A \mathbf{q} + \frac{1}{2} \mathbf{q}^T A^T (\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) A \mathbf{q} = 0. \tag{107}$$

If define a *differential fundamental matrix* to be

$$F = \begin{pmatrix} A^T \hat{v} A \\ \frac{1}{2} A^T (\hat{\omega}\hat{v} + \hat{v}\hat{\omega}) A \end{pmatrix}, \tag{108}$$

then the Longuet-Higgins constraint can be rewritten as

$$(\mathbf{u}^T, \mathbf{q}^T) F \mathbf{q} = 0. \tag{109}$$

The problem we are interested here is *how to recover, from the fundamental matrix, the camera intrinsic-parameter matrix A and the camera ego-motion v and ω.*

Note that the fundamental matrix F has an asymmetric part and symmetric part, as the essential matrix does. Such matrices have at most eight degrees of freedom. Actually, it is shown by Brooks *et al* [2] that the set of all F is only a 7-dimensional submanifold in $\mathbb{R}^9$, *i.e.* it has 7 degrees of freedom. However, F (with normalized linear velocity v) is a function of 10 free parameters, 5 for motion and 5 for camera intrinsic parameters. Thus, the essential constraint is not adequate for determining all the camera motion and intrinsic parameters in the most general case. If one wants to recover all 5 ego-motion parameters, maximally, only 2 extra intrinsic parameters can be recovered at the same time.

This seems to be a little frustrating. However, it actually simplifies the problem we are trying to solve – the most general case is not solvable anyway, one has to consider the self-calibration problem for much simplified camera models (with less than two free intrinsic parameters). Some interesting and solvable cases have been studied by Brooks *et al* in [2].

4.3 Experimental Results

We carried out initial simulations in order to study the performance of our algorithm. We chose to evaluate it in terms of bias and sensitivity of the estimate 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 512x512 pixels.

Our algorithm has been implemented in Matlab and the simulations have been performed using example sets proposed by [17] in their paper on comparison of the egomotion estimation from optical flow[5]. 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 2 to 8 units of the focal length, which was considered to be unity. The results presented below are for fixed field of view (FOV) of 60 degrees. Each simulation consisted of 500 trials with a fixed noise level, FOV 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. The presented results demonstrate the performance of the algorithm while translating along X-axis and rotating around Z-axis with rate of 23^o per frame. The analysis of the obtained results of the motion estimation algorithm was performed using benchmarks proposed by [17]. The bias is expressed as an angle between the average estimate out of all trails (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 and their relative speed. The choice of the rotation axis did not influence the translation estimates. In the case of the rotation estimate our algorithm is slightly better compared to Heeger and Jepson's algorithm. This is due to the fact that in our case the rotation is estimated simultaneously with the translation so its bias is only due to the bias of the initially estimated differential essential matrix obtained by linear least squares techniques. This is in contrary to the rotation estimate used by Jepson and Heeger's algorithm which uses another least-squares estimation by substituting already biased translational estimate to compute the rotation. The translational estimates are very similar. Increasing the ratio between magnitudes of translational and angular velocities improves the bias and sensitivity of both algorithms.

The evaluation of the results and more extensive simulations are currently underway. We believe that through thorough understanding of the source of translational bias we can obtain even better

[5]We would like to thank the authors in [17] for making the code for simulations of various algorithms and evaluation of their results available on the web.

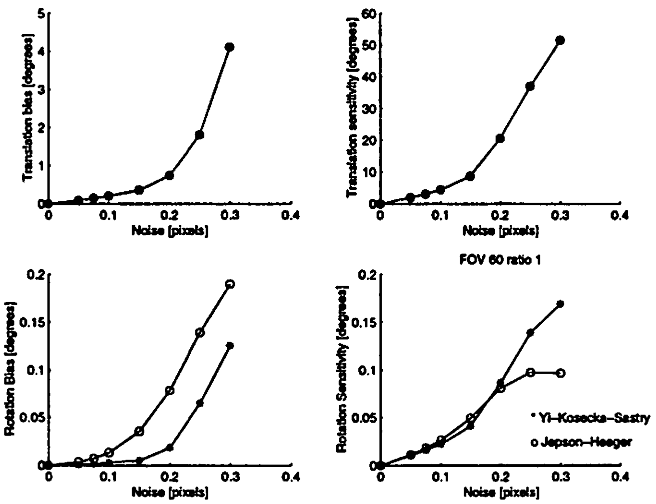

Figure 3: Bias for each noise level was estimated by running 500 trails and computing the average translation and rotation. The ratio between the magnitude of linear and angular velocity is 1.

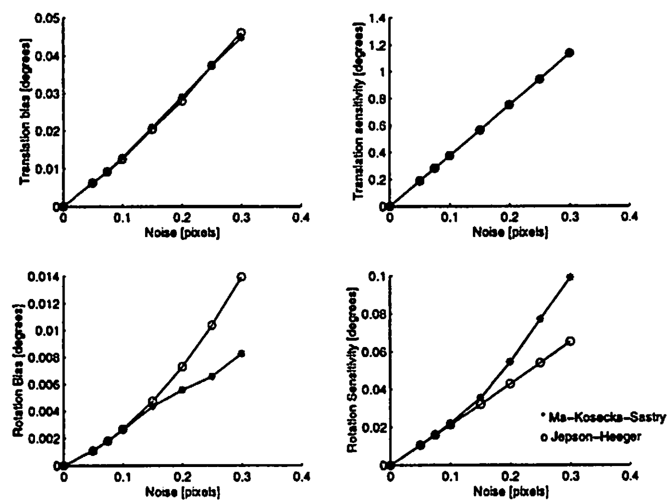

Figure 4: Bias for each noise level was estimated by running 500 trails and computing the average translation and rotation. The ratio between the magnitude of linear and angular velocity is 10.

performance by utilizing additional information about linear velocity, which is embedded in the symmetric part of the differential essential matrix. In the current simulations translation was estimated only from v_0 skew symmetric part of $\mathbf{e}$.

5 Discussions and Future Work

This paper presents a unified view of the problem of egomotion estimation using discrete and differential Longuet-Higgins constraint. In both (discrete and differential) settings we provide a geometric characterization of the space of essential matrices and differential essential elements. This 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 egomotion estimation algorithm, which is a natural counterpart of the three-step SVD based algorithm in developed for the discrete case by [19].

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 *et al* [15] for the discrete case. The same ideas certainly apply to our algorithm.

In applications to robotics, a big advantage of the differential approach over the discrete one is that it can make use of nonholonomic constraints (*i.e.* constraints that confine the infinitesimal motion of the mobile base but not the global motion) and simplify the motion estimation algorithms. An example study of vision guided nonholonomic system can be found in [9]. In this paper, we have assumed that the camera is ideal. This 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 [21, 2]. 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 [2], that the space of such fundamental matrices is a 7-dimensional algebraic variety in $\mathbb{R}^{3\times3}$. Thus, besides five motion parameters, only two extra intrinsic parameters can be recovered.

Due to the geometric clarity of the motion esimation algorithms, it is promising to merge a vision system using these algorithms with INS (inertial navigation sensors, such as gyroscopes) and GPS (global positioning system) to recover 3D motion and orientation of autonomous mobile robots. It will highly improve the robustness and accuracy of the overall system.

Acknowledgments

The authors would like to thank M. Cenk Cavusoglu and Joseph Yan at the Berkeley Intelligent Machines & Robotics Laboratory for their helpful discussions on some of the proofs in this paper.

References

[1] William M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry.* Academic Press, second edition, 1986.

[2] Michael J. Brooks, Wojciech Chojnacki, and Luis Baumela. Determining the ego-motion of an uncalibrated camera from instantaneous optical flow. *in press*, 1997.

[3] A. R. Bruss and B. K. Horn. Passive navigation. *Computer Graphics and Image Processing*, 21:3–20, 1983.

[4] D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117, 1992.

[5] A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering translation direction. *Spatial Vision in Humans and Robots, Cambridge Univ. Press*, pages 39–62, 1993.

[6] K. Kanatani. 3d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282, 1993.

[7] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[8] Waxman A. M., Kamgar-Parsi B., and Subbarao M. Closed form solutions to image flow equations for 3d structure and motion. *International Journal of Computer Vision 1*, pages 239–258, 1987.

[9] Yi Ma, Jana Košecká, and Shankar Sastry. Vision guided navigation for a nonholonomic mobile robot. *Electronic Research Laboratory Memorandum, UC Berkeley*, UCB/ERL(M97/42), June 1997.

[10] Stephen Maybank. *Theory of Reconstruction from Image Motion.* Springer Series in Information Sciences. Springer-Verlag, 1993.

[11] Philip F. McLauchlan and David W. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proceeding of Fifth International Conference on Computer Vision*, pages 314–320, Cambridge, MA, USA, 1995. IEEE Comput. Soc. Press.

[12] Richard M. Murray, Zexiang Li, and Shankar S. Sastry. *A Mathematical Introduction to Robotic Manipulation.* CRC press Inc., 1994.

[13] Andrew Kelly Packard. What's new with μ : structured uncertainty in multivariable control. *PhD thesis, EECS, UC Berkeley*, 1988.

[14] Steven Thomas Smith. Geometric optimization methods for adaptive filtering. *PhD thesis, Division of Applied Sciences, Harvard University*, May 1993.

[15] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393–413, March 1996.

[16] Norman Steenrod. *The Topology of Fiber Bundles.* Princeton Mathematical Series. Princeton University Press, 1951.

[17] T. Y. Tian, C. Tomasi, and D. J. Heeger. Comparison of approaches to egomotion computation. In *Proceedings of 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 315–20, Los Alamitos, CA, USA, 1996. IEEE Comput. Soc. Press.

[18] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography. *Intl. Journal of Computer Vision*, 9(2):137–154, 1992.

[19] G. Toscani and O. D. Faugeras. Structure and motion from two noisy perspective images. *Proceedings of IEEE Conference on Robotics and Automation*, pages 221–227, 1986.

[20] Roger Y. Tsai and Thomas S. Huang. 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, January 1984.

[21] T. Vieville and O. D. Faugeras. Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. *Proceedings of Fifth International Conference on Computer Vision*, pages 750–756, June 1995.

[22] Xinhua Zhuang and R. M. Haralick. Rigid body motion and optical flow image. *Proceedings of the First International Conference on Artificial Intelligence Applications*, pages 366–375, 1984.

[23] Xinhua Zhuang, Thomas S. Huang, and Narendra Ahuja. A simplified linear optic flow-motion algorithm. *Computer Vision, Graphics and Image Processing*, 42:334–344, 1988.