International Journal of Computer Vision 68(1), 7–25, 2006

DOI: 10.1007/s11263-005-4839-7

Two-View Multibody Structure from Motion*

RENÉ VIDAL
Center for Imaging Science, Department of Biomedical Engineering, Johns Hopkins University, 308B Clark Hall, 3400 N. Charles St., Baltimore, MD 21218
rvidal@cis.jhu.edu

YI MA
Department of ECE, University of Illinois at Urbana-Champaign, 1406 West Green Street, Urbana, IL 61801
yima@uiuc.edu

STEFANO SOATTO
Computer Science Department, University of California at Los Angeles, 3531 Boelter Hall, Los Angeles, CA 90095
soatto@cs.ucla.edu

SHANKAR SASTRY
Department of EECS, University of California at Berkeley, 237 Cory Hall, Berkeley, CA 94720
sastry@eecs.berkeley.edu

Received May 2002; Accepted February 2005

First online version published in April, 2006

Abstract. We present an algebraic geometric approach to 3-D motion estimation and segmentation of multiple rigid-body motions from noise-free point correspondences in two perspective views. Our approach exploits the algebraic and geometric properties of the so-called *multibody epipolar constraint* and its associated *multibody fundamental matrix*, which are natural generalizations of the epipolar constraint and of the fundamental matrix to multiple motions. We derive a rank constraint on a polynomial embedding of the correspondences, from which one can estimate the number of independent motions as well as linearly solve for the multibody fundamental matrix. We then show how to compute the epipolar lines from the first-order derivatives of the multibody epipolar constraint and the epipoles by solving a plane clustering problem using Generalized PCA (GPCA). Given the epipoles and epipolar lines, the estimation of individual fundamental matrices becomes a linear problem. The clustering of the feature points is then automatically obtained from either the epipoles and epipolar lines or from the individual fundamental matrices. Although our approach is mostly designed for noise-free correspondences, we also test its performance on synthetic and real data with moderate levels of noise.

Keywords: multibody structure from motion, 3-D motion segmentation, multibody epipolar constraint, multibody fundamental matrix, Generalized PCA (GPCA)

*This paper is an extended version of Vidal et al. (2002). Work supported by Hopkins WSE and UIUC ECE startup funds, and by grants NSF CAREER ISS-0447739, ONR N00014-00-1-0621, NSF CAREER IIS-0347456, NSF IIS-0347456, ONR N00014-03-1-0850, ARO DAAD19-99-1-0137 and AFOSRF49620-03-1-0095, N00014-05-1-0836.

1. Introduction

A classic problem in visual motion analysis is to estimate a motion model for a set of 2-D feature points as they move in a video sequence. When the scene is *static*, the problem of fitting a 3-D model compatible with the structure and motion of the scene is well understood (Hartley and Zisserman, 2000; Ma et al., 2003). For instance, it is well-known that two views of a scene are related by the so-called epipolar constraint and that the motion of the camera can be estimated using linear techniques such as the eight-point algorithm (Longuet-Higgins, 1981). However, these techniques can not deal with *dynamic scenes* in which different regions of the image obey different motion models due to depth discontinuities, perspective effects, multiple moving objects, etc.

Motion estimation and segmentation refers to the problem of fitting multiple motion models to the scene, without knowing which feature points are moving according to the same model. Previous work (Feng and Perona, 1998) solves this problem by first clustering the features corresponding to the same motion using K-means or spectral clustering, and then estimating a single motion model for each group. This can also be done in a probabilistic framework (Torr, 1998) in which a maximum-likelihood estimate of the parameters of each motion model is sought by alternating between feature clustering and single-body motion estimation using the Expectation Maximization (EM) algorithm. However, the convergence of EM to the global maximum depends strongly on initialization (Torr et al., 2001).

In order to deal with the initialization problem, recent work has concentrated on the study of the geometry of dynamic scenes. Even for drastically simplified motion models, such as multiple points moving linearly with constant speed (Han and Kanade, 2000; Shashua and Levin, 2001) or in a conic section (Avidan and Shashua, 2000), multiple points moving in a plane (Sturm, 2002), or multiple translating planes (Wolf and Shashua, 2001), the geometric aspects of the problem are non-trivial. More general motion models have only been studied in the case of orthographic cameras observing multiple rigid-body motions (Costeira and Kanade, 1995; Kanatani, 2001; Vidal and Hartley, 2004) and in the case of two perspective cameras observing two rigid-body motions (Wolf and Shashua, 2001).

In this paper, we present an algebraic geometric approach to the estimation and segmentation of an *unknown* number of rigid-body motions from a set of noise-free point correspondences in two perspective views. Rather than alternating between feature clustering and single-body motion estimation, our approach algebraically eliminates the feature clustering stage and directly solves for the motion parameters in an algebraic fashion. This is achieved by fitting a multibody motion model to all the image measurements and then factorizing this model to obtain the individual motion parameters. The final result is a natural generalization of the geometry of the classical two-view structure from motion problem (epipolar constraint, fundamental matrix and eight-point algorithm) to the case of multiple rigid-body motions.

Section 2 studies the geometry and algebra of the multibody structure from motion problem. We introduce the *multibody epipolar constraint* as a geometric relationship between the motion parameters and the image points that is satisfied by all the correspondences, regardless of the body with which they are associated. We show that the multibody epipolar constraint is bilinear on a polynomial embedding of the correspondences and linear on the so-called *multibody fundamental matrix* $\mathcal{F}$, an algebraic structure encoding the parameters of all rigid-body motions. We then study the geometric properties of $\mathcal{F}$ and prove that the embedded epipoles of each independent motion are the intersection of the left null space of $\mathcal{F}$ with the so-called Veronese surface.

Section 3 presents an algebraic geometric algorithm for estimating the number of motions, the motion parameters and the clustering of the correspondences. We first derive a rank constraint on the matrix of embedded correspondences from which one can estimate the number of independent motions n as well as linearly solve for the multibody fundamental matrix $\mathcal{F}$. Given n and $\mathcal{F}$, we show that one can estimate the epipolar line associated with each correspondence from the first-order derivatives of the multibody epipolar constraint at the correspondence. By applying this process to all the correspondences, we obtain a collection of epipolar lines that must intersect at the n epipoles. The estimation of the epipoles is then equivalent to a plane clustering problem, which we solve algebraically using Generalized Principal Component Analysis (GPCA) (Vidal et al., 2005, 2004, 2003). Given the epipoles and epipolar lines, the estimation of individual fundamental matrices becomes a linear problem. The clustering of the feature points is then automatically obtained from either the epipoles and epipolar lines or from the individual fundamental matrices. Since our technique is based

on solving linear systems, taking derivatives of multivariate polynomials, and computing roots of univariate polynomials, 3-D motion segmentation can be solved in closed form if and only if the number of motions is at most four.

Although our 3-D motion segmentation algorithm is mostly designed for noise-free correspondences, in Section 4 we present experiments on synthetic data that evaluate the performance of our approach with respect to moderate levels of noise. We also apply our algorithm to the segmentation of an indoor sequence.

Remark 1. Although the proposed algebraic geometric algorithm is algebraically equivalent to the factorization of symmetric tensors, we avoid the use of tensorial notation throughout the paper, because algorithms based on polynomial algebra are computationally more straightforward and better established. As a consequence, this paper requires little background beyond conventional linear and polynomial algebra.

2. Multibody Epipolar Geometry

In this section, we generalize classical epipolar geometry to the case of multiple rigid-body motions. We introduce the multibody epipolar constraint and the multibody fundamental matrix, and analyze some of their algebraic and geometric properties.

2.1. Two-View Multibody Structure From Motion Problem

Consider a scene containing an *unknown* number n of rigidly moving objects and let $\{g_i(t) \in SE(3)\}_{i=1}^n$ represent their poses at time t. If we assume that the n objects are being observed by a moving perspective camera whose pose at time t is $g_0(t) \in SE(3)$, then the motion of object i relative to the camera between two frames at times t and $t+1$, is given by $(R_i, T_i) = g_i(t+1)g_0(t+1)^{-1}g_0(t)g_i(t)^{-1}$. Let $F_i \in \mathbb{R}^{3\times 3}$ be the fundamental matrix associated with the ith rigid-body motion $(R_i, T_i) \in SE(3)$ for $i = 1, \ldots, n$. That is, $F_i = K_i^{-T}[T_i]_\times R_i K_i^{-1}$, where $K_i \in \mathbb{R}^{3\times 3}$ is the camera calibration matrix, and $[T_i]_\times \in so(3)$ is the skew-symmetric matrix representing the cross product with T_i. We assume that the n rigid-body motions are *independent* from each other, i.e., we assume that the n fundamental matrices $\{F_i\}_{i=1}^n$ are *different* (up to a scale factor).

Let $\{\mathbf{X}^j \in \mathbb{R}^3\}_{j=1}^N$ be a collection of points in 3-D space lying on the n moving objects. We denote the image of a point $\mathbf{X}^j \in \mathbb{R}^3$ with respect to image frame I_f as $\boldsymbol{x}_f^j \in \mathbb{P}^2$, for $j = 1, \ldots, N$ and $f = 1, 2$. In order to avoid degenerate cases, we will assume that the image points $\{\boldsymbol{x}_f^j\}$ are in general position in 3-D space, i.e., their corresponding 3-D points do not all lie in any critical surface, for example. We will drop the superscript when we refer to a generic image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Also, we will use the homogeneous representation $\boldsymbol{x} = [x, y, z]^T$ to refer to an arbitrary image point in $\mathbb{P}^2$, unless otherwise stated.

In this paper, we are concerned with the following problem.

Problem 1 *(Two-view multibody structure from motion problem.)*

Given a collection of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ corresponding to an unknown number of independently and rigidly moving objects, estimate the number of independent motions n, the fundamental matrices $\{F_i\}_{i=1}^n$, and the object to which each image pair belongs.

2.2. The Multibody Epipolar Constraint

Let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an arbitrary image pair associated with *any* of the n moving objects. Then, there exists a fundamental matrix $F_i \in \mathbb{R}^{3\times 3}$ such that the following *epipolar constraint* (Longuet-Higgins, 1981) is satisfied

$$\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0. \tag{1}$$

Therefore, regardless of the object to which the image pair belongs, the following *multibody epipolar constraint* must be satisfied by the number of independent motions n, the fundamental matrices $\{F_i\}_{i=1}^n$ and the image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$

$$\mathcal{E}(\boldsymbol{x}_1, \boldsymbol{x}_2) \doteq \prod_{i=1}^{n} \left(\boldsymbol{x}_2^T F_i \boldsymbol{x}_1\right) = 0. \tag{2}$$

The multibody epipolar constraint eliminates the problem of clustering the correspondences by taking the product of the epipolar constraints. Although this is not the only way of algebraically eliminating feature clustering, we will see shortly that (2) has the advantage of being a polynomial in $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ with a nice algebraic

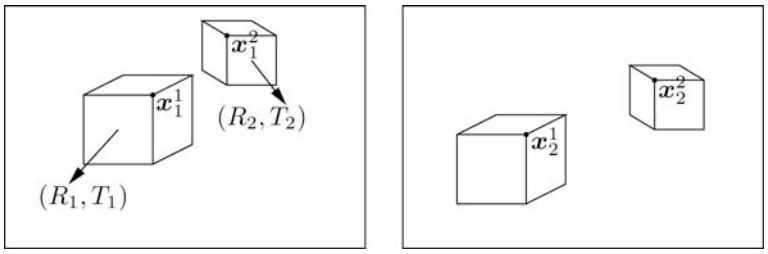

Figure 1. Two views of two independently moving objects with two different rotations and translations: (R_1, T_1) and (R_2, T_2) relative to the camera frame.

structure. For instance, we will show that the multibody epipolar constraint is an irreducible polynomial of minimal degree n, whose coefficients can be easily estimated from data, and whose derivatives encode the motion parameters.

Example 1 (Two rigid-body motions). Imagine the simplest case of a scene containing only two independently moving objects as shown in Fig. 1. In this case, both image pairs $(\boldsymbol{x}_1^1, \boldsymbol{x}_2^1)$ and $(\boldsymbol{x}_1^2, \boldsymbol{x}_2^2)$ satisfy the equation

$$\left(\boldsymbol{x}_2^T F_1 \boldsymbol{x}_1\right)\left(\boldsymbol{x}_2^T F_2 \boldsymbol{x}_1\right) = 0$$

for $F_1 = [T_1]_\times R_1$ and $F_2 = [T_2]_\times R_2$. This equation is no longer bilinear but rather bi-quadratic in the two images $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of any point $\mathbf{X}$ on one of these objects. Furthermore, the equation is no longer linear in F_1 or F_2 but rather bilinear in (F_1, F_2). However, if sufficiently many image pairs $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are given, we can still recover some information about the two fundamental matrices F_1 and F_2 from such equations, in spite of the fact that we do not know the object or motion to which each image pair belongs. An algorithm specifically designed for this special case ($n = 2$) was presented in Wolf and Shashua (2001). In this paper, we provide a general solution for an arbitrary number of motions n.

2.3. The Multibody Fundamental Matrix

The multibody epipolar constraint allows us to convert the multibody structure from motion problem (Problem 1) into one of solving for the number of independent motions n and the fundamental matrices $\{F_i\}_{i=1}^n$ from (2). However, while the epipolar constraint (1) is bilinear in the image points and linear in the fundamental matrix, the multibody epipolar constraint (2) is bi-homogeneous in the image points and multilinear in the fundamental matrices. A standard technique used in algebra to render a nonlinear problem linear is to find an "embedding" that lifts the problem into a higher-dimensional space. To this end, notice that the multibody epipolar constraint defines a homogeneous polynomial of degree n in each of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. For example, if we let $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T$, then Eq. (2) viewed as a function of $\boldsymbol{x}_1$ can be written as a linear combination of the following monomials $\{x_1^n, x_1^{n-1}y_1, x_1^{n-1}z_1, \ldots, z_1^n\}$. It is readily seen that there are a total of

$$M_n \doteq \binom{n+2}{2} = \frac{(n+1)(n+2)}{2} \tag{3}$$

different monomials, thus the dimension of the space of homogeneous polynomials in 3 variables with real coefficients is M_n. Therefore, we can define the following embedding (or lifting) from $\mathbb{R}^3$ into $\mathbb{R}^{M_n}$:

Definition 1 (Veronese map (Harris, 1992)). The *Veronese map* of degree n is defined as $\nu_n : \mathbb{R}^3 \to \mathbb{R}^{\mathrm{M_n}}$

$$\nu_n : [x, y, z]^T \mapsto [\ldots, \boldsymbol{x}^I, \ldots]^T, \tag{4}$$

where $\boldsymbol{x}^I$ is a monomial of the form $x^{n_1}y^{n_2}z^{n_3}$, with $0 \le n_1, n_2, n_3 \le n$, $n_1 + n_2 + n_3 = n$, and $\boldsymbol{x}^I$ ordered in the degree-lexicographic order.

Example 2 (The Veronese map of degree two). The Veronese map of degree $n = 2$ is given by $\nu_2(x, y, z) = [x^2, xy, xz, y^2, yz, z^2]^T \in \mathbb{R}^6$.

Thanks to the Veronese map, we can convert the multibody epipolar constraint (2) into a bilinear

expression in $\nu_n(\boldsymbol{x}_1)$ and $\nu_n(\boldsymbol{x}_2)$ as stated by the following theorem.

Theorem 1 (The bilinear multibody epipolar constraint)**.** *The multibody epipolar constraint (2) can be written in bilinear form as*

$$\nu_n(\boldsymbol{x}_2)^T \mathcal{F} \nu_n(\boldsymbol{x}_1) = 0, \tag{5}$$

where the entries of $\mathcal{F} \in \mathbb{R}^{M_n \times M_n}$ are symmetric multilinear functions of degree n on the entries of the fundamental matrices $\{F_i\}_{i=1}^n$.

Proof: Let $\boldsymbol{\ell}_i = F_i \boldsymbol{x}_1 \in \mathbb{R}^3$, for $i = 1, \ldots, n$. Then, the multibody epipolar constraint $\mathcal{E}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \prod_{i=1}^n (\boldsymbol{x}_2^T \boldsymbol{\ell}_i)$ is a homogeneous polynomial of degree n in $\boldsymbol{x}_2 = [x_2, y_2, z_2]^T$, i.e.,

$$\begin{aligned} \mathcal{E}(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \sum a_{n_1,n_2,n_3} x_2^{n_1} y_2^{n_2} z_2^{n_3} \\ &\doteq \sum a_I \boldsymbol{x}_2^I \doteq \nu_n(\boldsymbol{x}_2)^T \boldsymbol{a}, \end{aligned}$$

where $\boldsymbol{a} \in \mathbb{R}^{M_n}$ is the vector of coefficients. From the properties of polynomial multiplication, each a_I is a symmetric multilinear function of $(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n)$, i.e., it is linear in each $\boldsymbol{\ell}_i$ and $a_I(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n) = a_I(\boldsymbol{\ell}_{\sigma(1)}, \ldots, \boldsymbol{\ell}_{\sigma(n)})$ for all $\sigma \in \mathfrak{S}_n$, where $\mathfrak{S}_n$ is the permutation group of n elements. Since each $\boldsymbol{\ell}_i$ is linear in $\boldsymbol{x}_1$, each a_I is in turn a homogeneous polynomial of degree n in $\boldsymbol{x}_1$, i.e., $\boldsymbol{a}_I = \boldsymbol{f}_I^T \nu_n(\boldsymbol{x}_1)$, where each entry of $\boldsymbol{f}_I \in \mathbb{R}^{M_n}$ is a symmetric multilinear function of the entries of the F_i's. Letting

$$\mathcal{F} \doteq [\boldsymbol{f}_{n,0,0}, \boldsymbol{f}_{n-1,1,0}, \ldots, \boldsymbol{f}_{0,0,n}]^T \in \mathbb{R}^{M_n \times M_n},$$

we obtain

$$\mathcal{E}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \nu_n(\boldsymbol{x}_2)^T \mathcal{F} \nu_n(\boldsymbol{x}_1) = 0.$$

□

We call the matrix $\mathcal{F}$ the *multibody fundamental matrix*, because it is a natural generalization of the fundamental matrix to multiple motions. Since Eq. (5) clearly resembles the bilinear form of the epipolar constraint for a single rigid-body motion, we will refer to both Eqs. (2) and (5) as the *multibody epipolar constraint* from now on.

Remark 2. The multibody fundamental matrix is a matrix representation of the symmetric tensor product of all the fundamental matrices

$$\sum_{\sigma \in \mathfrak{S}_n} F_{\sigma(1)} \otimes F_{\sigma(2)} \otimes \cdots \otimes F_{\sigma(n)}, \tag{6}$$

with $\mathfrak{S}_n$ the permutation group of n elements and $\otimes$ the tensor product.

Although the multibody fundamental matrix $\mathcal{F}$ seems a complicated mixture of all the individual fundamental matrices $F_1, \ldots, F_n$, we will show in Section 3 that, under some mild conditions (e.g., the F_i's are different), one can still recover all the individual fundamental matrices from $\mathcal{F}$ by looking at the first-order derivatives of the multibody epipolar constraint. Before doing so, we shall further explore some algebraic and geometric properties of the multibody fundamental matrix.

2.4. Null Space of the Multibody Fundamental Matrix

In this section, we study the relationships between the multibody fundamental matrix $\mathcal{F}$ and the epipoles $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ associated with the fundamental matrices $F_1, \ldots, F_n$.[1] First of all, recall that the epipole e_i associated with the ith motion in the second image is defined as the left kernel of the (rank-2) fundamental matrix F_i, that is

$$\boldsymbol{e}_i^T F_i \doteq 0. \tag{7}$$

Therefore, each epipole $\boldsymbol{e}_i$, $i = 1, \ldots, n$, satisfies that

$$\begin{aligned} \forall \boldsymbol{x} \in \mathbb{R}^3, \quad & \left(\boldsymbol{e}_i^T F_1 \boldsymbol{x}\right)\left(\boldsymbol{e}_i^T F_2 \boldsymbol{x}\right) \ldots \left(\boldsymbol{e}_i^T F_n \boldsymbol{x}\right) \\ & = \nu_n(\boldsymbol{e}_i)^T \mathcal{F} \nu_n(\boldsymbol{x}) = 0. \end{aligned} \tag{8}$$

We call the vector $\nu_n(\boldsymbol{e}_i)$ the *embedded epipole* associated with the ith motion. Since $\nu_n(\boldsymbol{x})$ as a vector spans the entire $\mathbb{R}^{M_n}$ when $\boldsymbol{x}$ ranges over $\mathbb{R}^3$ (or $\mathbb{P}^2$),[2] from (8) we have

$$\nu_n(\boldsymbol{e}_i)^T \mathcal{F} = 0 \quad \text{for} \quad i = 1, \ldots, n. \tag{9}$$

Therefore, the embedded epipoles $\{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n$ lie on the left null space of $\mathcal{F}$ while the epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$ lie on the left null space of $\{F_i\}_{i=1}^n$. Hence, the rank of $\mathcal{F}$ is bounded above depending on the number of different epipoles (up to a scale factor) as stated by Lemmas 1 and 2 below.

Lemma 1 (Null space of $\mathcal{F}$ when the epipoles are different). *Let $\mathcal{F}$ be the multibody fundamental matrix generated by the fundamental matrices $F_1, \ldots, F_n$. If the epipoles $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are different (up to a scale factor), then the (left) null space of $\mathcal{F} \in \mathbb{R}^{M_n \times M_n}$ contains at least the n linearly independent vectors*

$$\nu_n(\boldsymbol{e}_i) \in \mathbb{R}^{M_n}, \quad i = 1, \ldots, n. \tag{10}$$

Therefore, the rank of the multibody fundamental matrix $\mathcal{F}$ is bounded above by

$$\text{rank}(\mathcal{F}) \leq (M_n - n). \tag{11}$$

Proof: We only need to show that if the $\boldsymbol{e}_i$'s are different up to a scale factor, then the $\nu_n(\boldsymbol{e}_i)$'s are linearly independent. If we let $\boldsymbol{e}_i = [x_i, y_i, z_i]^T$, for $i = 1, \ldots, n$, then we only need to prove that the rank of the following matrix

$$U \doteq \begin{bmatrix} \nu_n(\boldsymbol{e}_1)^T \\ \nu_n(\boldsymbol{e}_2)^T \\ \vdots \\ \nu_n(\boldsymbol{e}_n)^T \end{bmatrix} = \begin{bmatrix} x_1^n & x_1^{n-1}y_1 & x_1^{n-1}z_1 & \ldots & z_1^n \\ x_2^n & x_2^{n-1}y_2 & x_2^{n-1}z_2 & \ldots & z_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^n & x_n^{n-1}y_n & x_n^{n-1}z_n & \ldots & z_n^n \end{bmatrix} \in \mathbb{R}^{n \times M_n}$$

is exactly n. Since the $\boldsymbol{e}_i$'s are different up to a scale factor, we can assume without loss of generality that so are $\{[x_i, z_i]\}_{i=1}^n$ and that $z_i \neq 0$.[3] Then, after dividing the ith row of U by z_i^n and letting $t_i = x_i/z_i$, we can extract the following Van Der Monde sub-matrix of U

$$V \doteq \begin{bmatrix} t_1^{n-1} & t_1^{n-2} & \ldots & 1 \\ t_2^{n-1} & t_2^{n-2} & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ t_n^{n-1} & t_n^{n-2} & \ldots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Since $\det(V) = \prod_{i<j}(t_i - t_j)$, the Van Der Monde matrix V has rank n if and only if $t_1, \ldots, t_n$ are different. Hence $\text{rank}(U) = \text{rank}(V) = n$. □

Although we know that the linearly independent vectors $\nu_n(\boldsymbol{e}_i)$'s lie on the left null space of $\mathcal{F}$, we do not know if the n-dimensional subspace spanned by them is equal to the left null space of $\mathcal{F}$. That is, we do not know if $\text{rank}(\mathcal{F}) = M_n - n$. We conjecture that this is true when the n epipoles are different up to scale.

Let us now consider the case in which one of the epipoles is repeated. In this case, the null space of $\mathcal{F}$ is actually enlarged by higher-order derivatives of the Veronese map as stated by the following Lemma.

Lemma 2 (Null space of $\mathcal{F}$ when one epipole is repeated). *Let $\mathcal{F}$ be the multibody fundamental matrix generated by the fundamental matrices $F_1, \ldots, F_n$ with epipoles $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$. Let $\boldsymbol{e}_1$ be repeated k times, i.e., $\boldsymbol{e}_1 = \cdots = \boldsymbol{e}_k$ (up to scale), and let the other $n - k$ epipoles be different. Then the rank of the multibody fundamental matrix $\mathcal{F}$ is bounded by*

$$\text{rank}(\mathcal{F}) \leq M_n - M_{k-1} - (n - k). \tag{12}$$

Proof: When $k = 2$, $\boldsymbol{e}_1 = \boldsymbol{e}_2$ is a "repeated root" of $\nu_n(\boldsymbol{x})^T\mathcal{F}$ as a polynomial (matrix) in $\boldsymbol{x} = [x_1, x_2, x_3]^T$. Therefore, its first-order partial derivatives must vanish at $\boldsymbol{x} = \boldsymbol{e}_1$, i.e., $D\nu_n(\boldsymbol{e}_1)^T\mathcal{F} = [\frac{\partial \nu_n(\boldsymbol{e}_1)}{\partial x_1}, \frac{\partial \nu_n(\boldsymbol{e}_1)}{\partial x_2}, \frac{\partial \nu_n(\boldsymbol{e}_1)}{\partial x_3}]^T\mathcal{F} = 0$. This gives $M_1 = 3$ linearly independent vectors in the left null space of $\mathcal{F}$, namely the columns of $D\nu_n(\boldsymbol{e}_1)$, because $D\nu_n(\boldsymbol{x})$ is full rank for all $\boldsymbol{x} \in \mathbb{P}^2$,[4] and $\boldsymbol{e}_1 \neq 0$. Furthermore, we also have that $\forall \boldsymbol{x} \in \mathbb{R}^3, n\nu_n(\boldsymbol{x}) = D\nu_n(\boldsymbol{x})\boldsymbol{x}$. Therefore, the span of the columns of $D\nu_n(\boldsymbol{e}_1)$ contains $\nu_n(\boldsymbol{e}_1)$, but does not contain $\nu_n(\boldsymbol{e}_i)$ for $i = 3, \ldots, n$. Hence $\text{rank}(\mathcal{F}) \leq M_n - M_1 - (n-1) = M_n - 3 - (n-1)$. Now if $k > 2$, all partial derivatives of $\nu_n(\boldsymbol{x})^T\mathcal{F}$ of order less than k must vanish at $\boldsymbol{x} = \boldsymbol{e}_1$. Furthermore, since $n\nu_n(\boldsymbol{x}) = D\nu_n(\boldsymbol{x})\boldsymbol{x}$, all the derivatives of order less than $(k-1)$ at $\boldsymbol{x} = \boldsymbol{e}_1$, including the embedded epipole $\nu_n(\boldsymbol{e}_1)$, must lie in the span of the derivatives of order $(k-1)$. As in the case $k = 2$ one can show that these M_{k-1} partial derivatives are linearly independent and that their span does not contain the other n–k embedded epipoles $\{\nu_n(\boldsymbol{e}_i)\}_{i=k+1}^n$. This gives at least $M_{k-1}+(n-k)$ linearly independent vectors in null($\mathcal{F}$). The readers are referred to Fan and Vidal (2005) for further details. □

Example 3 (Two repeated epipoles). In the two-body problem, if F_1 and F_2 have the same (left) epipole, then the rank of the two-body fundamental matrix $\mathcal{F}$ is bounded above by $M_2 - M_1 - (2-2) = 6 - 3 = 3$ instead of $M_2 - 2 = 4$.

Since the null space of $\mathcal{F}$ is enlarged by higher-order derivatives of the Veronese map evaluated at repeated

epipoles, in order to identify the embedded epipoles $\nu_n(\boldsymbol{e}_i)$ from the left null space of $\mathcal{F}$ we will need to exploit the algebraic structure of the Veronese map ν_n. Let us denote the image of the real projective space $\mathbb{P}^2$ under ν_n as $\nu_n(\mathbb{P}^2)$.[5] The following theorem establishes a key relationship between the null space of $\mathcal{F}$ and the epipoles of each fundamental matrix.

Theorem 2 (Veronese null space of multibody fundamental matrix)**.** *The intersection of the left null space of the multibody fundamental matrix $\mathcal{F}$,* null($\mathcal{F}$)*, with the Veronese surface $\nu_n(\mathbb{P}^2)$ is exactly*

$$\text{null}(\mathcal{F}) \cap \nu_n(\mathbb{P}^2) = \{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n. \tag{13}$$

Proof: Let $\boldsymbol{x} \in \mathbb{P}^2$ be a vector whose Veronese map is in the left null space of $\mathcal{F}$. We then have

$$\nu_n(\boldsymbol{x})^T \mathcal{F} = 0 \Leftrightarrow \forall \boldsymbol{y} \in \mathbb{P}^2,\ \nu_n(\boldsymbol{x})^T \mathcal{F} \nu_n(\boldsymbol{y}) = 0.$$

Since $\mathcal{F}$ is a multibody fundamental matrix, for this $\boldsymbol{x}$ we have that

$$\forall \boldsymbol{y} \in \mathbb{P}^2, \quad \nu_n(\boldsymbol{x})^T \mathcal{F} \nu_n(\boldsymbol{y}) = \prod_{i=1}^{n} (\boldsymbol{x}^T F_i \boldsymbol{y}) = 0.$$

If $\boldsymbol{x}^T F_i \neq 0$ for all $i = 1, \ldots, n$, then the set of $\boldsymbol{y}$ that satisfy the above equation is simply the union of n 2-dimensional subspaces in $\mathbb{P}^2$, which will never fill the entire space $\mathbb{P}^2$. Hence we must have $\boldsymbol{x}^T F_i = 0$ for some i. Therefore, $\boldsymbol{x}$ must be one of the epipoles. □

The significance of Theorem 2 is that, in spite of the fact that repeated epipoles may enlarge the null space of $\mathcal{F}$, and that we do not know if the dimension of the null space equals n for different epipoles, one may always find the epipoles exactly by intersecting the left null space of $\mathcal{F}$ with the Veronese surface $\nu_n(\mathbb{P}^2)$, as illustrated in Fig. 2.

However, even though Theorem 2 gives a nice conceptual description of the relation between the multibody fundamental matrix $\mathcal{F}$ and the multiple epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$, it does not provide a computational algorithm for estimating the individual epipoles from the intersection of null($\mathcal{F}$) with $\nu_n(\mathbb{P}^2)$. One possible approach, explored in Wolf and Shashua (2001) for $n = 2$ and generalized in Vidal et al. (2002) to $n \geq 2$, consists of determining a vector $v \in \mathbb{R}^n$ such that $Bv \in \nu_n(\mathbb{P}^2)$, where B is a matrix whose columns form a basis for the left null space of $\mathcal{F}$. Finding v, hence the epipoles, is equivalent to solving for the roots of polynomials of degree n in $n-1$ variables. Although this is feasible for $n = 2$ and even for $n = 3$, it is computationally formidable for $n > 3$.

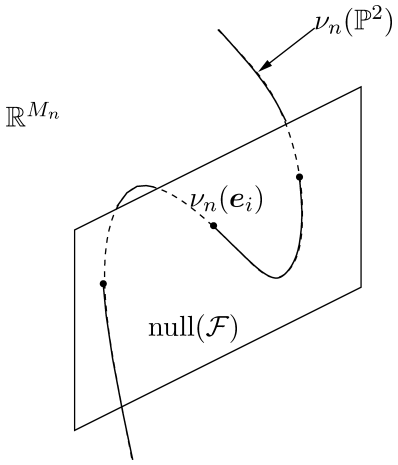

Figure 2. The intersection of $\nu_n(\mathbb{P}^2)$ and null($\mathcal{F}$) is exactly n points representing the Veronese map of the n epipoles, repeated or not.

In the next section, we propose a completely different approach for estimating the epipoles. We first estimate a set of epipolar lines from the derivatives of the multibody epipolar constraint and then cluster the epipolar lines using GPCA. Given the epipoles and epipolar lines, the estimation of the individual fundamental matrices becomes a linear problem.

3. Multibody Motion Estimation and Segmentation

Up until now, we have been mostly concerned with the study of the geometry and algebra of the multibody structure from motion problem. From now on, we will concentrate on the computational aspects of the problem. More specifically, this section proposes an algebraic geometric algorithm for estimating the number of motions, the fundamental matrices and the clustering of the image points. The algorithm proceeds as follows. Section 3.1 shows how to estimate the number of motions n and the multibody fundamental matrix $\mathcal{F}$ from a rank constraint on the embedded correspondences. Section 3.2 shows how to estimate the epipolar line associated with each image pair from the partial derivatives of the multibody epipolar constraint. Section 3.3 shows that the estimation of the epipoles is equivalent

to a plane clustering problem which can be solved in an algebraic fashion using GPCA. Section 3.4 shows how to estimate the individual fundamental matrices from epipoles and epipolar lines and Section 3.5 shows how to cluster the image points from either the epipoles and epipolar lines or from the individual fundamental matrices. Section 3.6 summarizes our multibody structure from motion algorithm and analyzes some of its properties.

3.1. Estimating the Number of Motions n and the Multibody Fundamental Matrix $\mathcal{F}$

Notice that, by definition, the multibody fundamental matrix $\mathcal{F}$ depends explicitly on the number of independent motions n. Therefore, even though the multibody epipolar constraint (5) is *linear* in $\mathcal{F}$, we cannot use it to estimate $\mathcal{F}$ without knowing n in advance. Fortunately, one can use the multibody epipolar constraint to derive a rank constraint on the image measurements from which one can compute n explicitly. Given n, the estimation of $\mathcal{F}$ becomes a *linear* problem. To see this, let us first rewrite the multibody epipolar constraint (5) as

$$(\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1))^T \boldsymbol{f} = 0, \tag{14}$$

where $\boldsymbol{f} \in \mathbb{R}^{M_n^2}$ is the stack of the rows of $\mathcal{F}$ and $\otimes$ represents the Kronecker product. Given a collection of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$, the vector $\boldsymbol{f}$ satisfies the system of linear equations

$$\boldsymbol{V}_n \boldsymbol{f} \doteq \begin{bmatrix} \big(\nu_n\big(\boldsymbol{x}_2^1\big) \otimes \nu_n\big(\boldsymbol{x}_1^1\big)\big)^T \\ \big(\nu_n\big(\boldsymbol{x}_2^2\big) \otimes \nu_n\big(\boldsymbol{x}_1^2\big)\big)^T \\ \vdots \\ \big(\nu_n\big(\boldsymbol{x}_2^N\big) \otimes \nu_n\big(\boldsymbol{x}_1^N\big)\big)^T \end{bmatrix} \boldsymbol{f} = 0. \tag{15}$$

In order to determine $\boldsymbol{f}$ uniquely (up to a scale factor) from (15), we must have that

$$\operatorname{rank}(\boldsymbol{V}_n) = M_n^2 - 1. \tag{16}$$

The above rank constraint on the matrix $\boldsymbol{V}_n$ provides an effective criterion for determining the number of independent motions n from the given image pairs, as stated by the following theorem.

Theorem 3 (Number of independent motions)**.** *Let $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ be a collection of image pairs corresponding to 3-D points in general configuration and undergoing an unknown number n of independent rigid-body motions with nonzero translation. Let $\boldsymbol{V}_i \in \mathbb{R}^{N \times M_i^2}$ be the matrix defined in (15), but computed using the Veronese map ν_i of degree $i \geq 1$. Then, if the number of image pairs is big enough ($N \geq M_n^2 - 1$ when n is known) and at least 8 points correspond to each motion, we have*

$$\operatorname{rank}(\boldsymbol{V}_i) \begin{cases} > M_i^2 - 1, & \text{if} \quad i < n, \\ = M_i^2 - 1, & \text{if} \quad i = n, \\ < M_i^2 - 1, & \text{if} \quad i > n. \end{cases} \tag{17}$$

Therefore, the number of independent motions n is given by

$$n \doteq \min \left\{ i : \operatorname{rank}(\boldsymbol{V}_i) = M_i^2 - 1 \right\}. \tag{18}$$

Proof: Let Z_i be the set of points $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that satisfy $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0$. Since each fundamental matrix F_i has rank 2, the polynomial $p_i = \boldsymbol{x}_2^T F_i \boldsymbol{x}_1$ is irreducible over the real field $\mathbb{R}$, i.e., it can not be factored as a product of nonconstant polynomials with real coefficients. This implies that any polynomial p in $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that vanishes on the entire set Z_i must be of the form $p = p_i h$, where h is some polynomial. Hence, if $F_1, \ldots, F_n$ are different up to scale, a polynomial that vanishes on the set $\cup_{i=1}^n Z_i$ must be of the form $p = p_1 p_2 \cdots p_n h$ for some h. Therefore, the only polynomial of *minimal* degree that vanishes on $\cup_{i=1}^n Z_i$ is

$$p = p_1 p_2 \ldots p_n = \big(\boldsymbol{x}_2^T F_1 \boldsymbol{x}_1\big)\big(\boldsymbol{x}_2^T F_2 \boldsymbol{x}_1\big) \ldots \big(\boldsymbol{x}_2^T F_n \boldsymbol{x}_1\big).$$

Since the rows of $\boldsymbol{V}_n$ are of the form $(\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1))^T$ and the entries of $\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1)$ are exactly the independent monomials of p (as we will show below), this implies that if the number of data points per motion is at least 8 and $N \geq M_n^2 - 1$, then:

1. There is no polynomial of degree $2i < 2n$ whose coefficients are in the null space of $\boldsymbol{V}_i$, i.e., $\operatorname{rank}(\boldsymbol{V}_i) = M_i^2 > M_i^2 - 1$ for $i < n$.
2. There is a unique polynomial of degree $2n$, namely p, whose coefficients are in the null space of $\boldsymbol{V}_n$, i.e., $\operatorname{rank}(\boldsymbol{V}_n) = M_n^2 - 1$.
3. There is more than one polynomial of degree $2i > 2n$ (one for each independent choice of the $2(i-n)$-degree polynomial h in $p = p_1 p_2 \cdots p_n h$) with

coefficients in the null space of $\boldsymbol{V}_i$, i.e., $\text{rank}(\boldsymbol{V}_i) < M_i^2 - 1$ for $i > n$.

The rest of the proof is to show that the entries of $\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1)$ are exactly the independent monomials in the polynomial p, which we do by induction. Since the claim is obvious for $n = 1$, we assume that it is true for n and prove it for $n+1$. Let $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T$ and $\boldsymbol{x}_2 = [x_2, y_2, z_2]^T$. Then the entries of $\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1)$ are of the form $(x_2^{m_1} y_2^{m_2} z_2^{m_3})(x_1^{n_1} y_1^{n_2} z_1^{n_3})$ with $m_1+m_2+m_3 = n_1+n_2+n_3 = n$, while the entries of $\boldsymbol{x}_2 \otimes \boldsymbol{x}_1$ are of the form $(x_2^{i_1} y_2^{i_2} z_2^{i_3})(x_1^{j_1} y_1^{j_2} z_1^{j_3})$ with $i_1+i_2+i_3 = j_1+j_2+j_3 = 1$. Thus a basis for the product of these monomials is given by the entries of $\nu_{n+1}(\boldsymbol{x}_2) \otimes \nu_{n+1}(\boldsymbol{x}_1)$. □

The significance of Theorem 3 is that the number of independent motions n can now be determined incrementally using Eq. (18). Once the number of motions is known, the multibody fundamental matrix $\mathcal{F}$ is simply the 1-D null space of the corresponding matrix $\boldsymbol{V}_n$, which can be linearly obtained. Nevertheless, in order for this scheme to work, the minimum number of image pairs needed is $N \geq M_n^2 - 1$. For $n = 1, 2, 3, 4$, the minimum N is 8, 35, 99, 225, respectively. If n is large, N grows approximately in the order of $O(n^4)$—a price to pay for working with a linear representation of Problem 1. In Section 3.6 we will discuss many variations to the general scheme that will significantly reduce the number of data points required, especially for large n.

Remark 3 (Estimating n and $\mathcal{F}$ from noisy correspondences). In the presence of a moderate level of noise, if n is known we can still solve for the multibody fundamental matrix $\boldsymbol{f}$ in (15), in a least-squares sense: we let $\boldsymbol{f}$ be the eigenvector of $\boldsymbol{V}_n^T \boldsymbol{V}_n$ associated with its smallest eigenvalue. However, when the number of motions n is unknown we cannot directly estimate it from (18), because the matrix $\boldsymbol{V}_i$ may be full rank for all i. In this case, we compute n from a noisy matrix $\boldsymbol{V}_i$ as

$$n = \arg\min_{i \geq 1} \frac{\sigma_{M_i^2}^2(\boldsymbol{V}_i)}{\sum_{k=1}^{M_i^2-1} \sigma_k^2(\boldsymbol{V}_i)} + \mu\big(M_i^2 - 1\big), \quad (19)$$

where $\sigma_k(\boldsymbol{V}_i)$ is the kth singular value of $\boldsymbol{V}_i$ and μ is a parameter. The above formula for estimating n is motivated by model selection techniques (Kanatani, 2002) in which one minimizes a cost function that consists of a data fitting term and a model complexity term. The data fitting term measures how well the data is approximated by the model—in this case how close the matrix $\boldsymbol{V}_i$ is to dropping rank by one. The model complexity term penalizes choosing models of high complexity – in this case choosing a large rank.

3.2. *Estimating the Epipolar Lines* $\{\boldsymbol{\ell}^j\}_{j=1}^N$

Given a point $\boldsymbol{x}_1$ in the first view, the epipolar lines associated with it in the second view are defined as $\boldsymbol{\ell}_i \doteq F_i \boldsymbol{x}_1 \in \mathbb{P}^2$, $i = 1, \ldots, n$. In the case of a single motion ($n = 1$), we know from the epipolar constraint that there is only *one* epipolar line $\boldsymbol{\ell} = F\boldsymbol{x}_1$ associated with $\boldsymbol{x}_1$ and that this line must pass through the corresponding point in the second view $\boldsymbol{x}_2$, i.e., $\boldsymbol{x}_2^T \boldsymbol{\ell} = 0$. In the case of n motions there are n possible epipolar lines $\{\boldsymbol{\ell}_i\}_{i=1}^n$ associated with a point $\boldsymbol{x}_1$ in the first view, each one corresponding to each one of the n motions. One of these n epipolar lines, say $\boldsymbol{\ell} = \boldsymbol{\ell}_i$, corresponds to the actual motion of $\boldsymbol{x}_1$ and hence it must pass through $\boldsymbol{x}_2$, i.e., there exists and i such that $\boldsymbol{x}_2^T \boldsymbol{\ell}_i = 0$. We will refer to $\boldsymbol{\ell}$ as the epipolar line associated with $(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

We now concentrate on how to determine the n epipolar lines $\{\boldsymbol{\ell}_i\}_{i=1}^n$ associated with $\boldsymbol{x}_1$ and the epipolar line $\boldsymbol{\ell}$ associated with $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ from the multibody fundamental matrix $\mathcal{F}$. To this end, notice that if $\mathcal{F}$ is the multibody fundamental matrix associated with $\{F_i\}_{i=1}^n$, then

$$\begin{aligned} \mathcal{E}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \nu_n(\boldsymbol{x}_2)^T \mathcal{F} \nu_n(\boldsymbol{x}_1) &= \prod_{i=1}^n \big(\boldsymbol{x}_2^T F_i \boldsymbol{x}_1\big) \\ &= \prod_{i=1}^n \big(\boldsymbol{x}_2^T \boldsymbol{\ell}_i\big). \end{aligned} \quad (20)$$

Therefore, the vector $\tilde{\boldsymbol{\ell}} \doteq \mathcal{F}\nu_n(\boldsymbol{x}_1) \in \mathbb{R}^{M_n}$ represents the coefficients of the homogeneous polynomial in $\boldsymbol{x}$

$$q_n(\boldsymbol{x}) \doteq (\boldsymbol{x}^T \boldsymbol{\ell}_1)(\boldsymbol{x}^T \boldsymbol{\ell}_2) \ldots (\boldsymbol{x}^T \boldsymbol{\ell}_n) = \nu_n(\boldsymbol{x})^T \tilde{\boldsymbol{\ell}}. \quad (21)$$

We call the vector $\tilde{\boldsymbol{\ell}}$ the *multibody epipolar line* associated with $\boldsymbol{x}_1$. Notice that $\tilde{\boldsymbol{\ell}}$ is a vector representation of the symmetric tensor product of all the epipolar lines $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n$ and is in general *not* the Veronese map (or lifting) $\nu_n(\boldsymbol{\ell}_i)$ of any particular epipolar line $\boldsymbol{\ell}_i$, $i = 1, \ldots, n$. From (21) we notice that in order to estimate the individual epipolar lines $\{\boldsymbol{\ell}_i\}_{i=1}^n$ associated with any image point $\boldsymbol{x}_1$ from the multibody epipolar line $\tilde{\boldsymbol{\ell}}$, we can factorize the homogeneous polynomial of

degree n, $q_n(\boldsymbol{x})$, into a product of n homogeneous polynomials of degree one $\{(\boldsymbol{\ell}_i^T \boldsymbol{x})\}_{i=1}^n$. We showed in Vidal et al. (2003) that this polynomial factorization problem has a unique solution (up to a scale for each factor) and that is algebraically equivalent to solving for the roots of a polynomial of degree n in *one* variable, plus solving a linear system in n variables. We shall assume that this polynomial factorization technique is available to us from now on, and refer interested readers to Vidal et al. (2003) for further details. Given such a factorization of $\tilde{\boldsymbol{\ell}}$ into the n epipolar lines $\{\boldsymbol{\ell}_i\}_{i=1}^n$ associated with $\boldsymbol{x}_1$, we can compute the epipolar line $\boldsymbol{\ell}$ associated with $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ as the vector $\boldsymbol{\ell}_i$ that minimizes $(\boldsymbol{x}_2^T \boldsymbol{\ell}_i)^2$ for $i = 1, \ldots, n$.

We can interpret the factorization of the multibody epipolar line $\tilde{\boldsymbol{\ell}} = \mathcal{F}\nu_n(\boldsymbol{x}_1)$ as a generalization of the conventional "epipolar transfer" to multiple motions. In essence, the multibody fundamental matrix $\mathcal{F}$ allows us to "transfer" a point $\boldsymbol{x}_1$ in the first image to a set of epipolar lines in the second image, the same way a fundamental matrix maps a point in the first image to an epipolar line in the second image. We illustrate the *multibody epipolar transfer* process with the following sequence of maps

$$\boldsymbol{x}_1 \overset{\text{Veronese}}{\longmapsto} \nu_n(\boldsymbol{x}_1) \overset{\substack{\text{Epipolar}\\ \text{Transfer}}}{\longmapsto} \mathcal{F}\nu_n(\boldsymbol{x}_1) \overset{\substack{\text{Polynomial}\\ \text{Factorization}}}{\longmapsto} \{\boldsymbol{\ell}_i\}_{i=1}^n,$$

as shown geometrically in Fig. 3.

There is however a simpler and more elegant way of computing the epipolar line $\boldsymbol{\ell}$ associated with an image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Notice from Eq. (20) that the partial derivative of the multibody epipolar constraint with respect to $\boldsymbol{x}_2$ is given by

$$\frac{\partial}{\partial \boldsymbol{x}_2}(\nu_n(\boldsymbol{x}_2)^T \mathcal{F}\nu_n(\boldsymbol{x}_1)) = \sum_{i=1}^{n} \prod_{\ell \neq i} \big(\boldsymbol{x}_2^T F_\ell \boldsymbol{x}_1\big)(F_i \boldsymbol{x}_1). \tag{22}$$

Therefore, if the image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ corresponds to motion i, i.e., if $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0$, then

$$\frac{\partial}{\partial \boldsymbol{x}_2}(\nu_n(\boldsymbol{x}_2)^T \mathcal{F}\nu_n(\boldsymbol{x}_1)) = \prod_{\ell \neq i} \big(\boldsymbol{x}_2^T F_\ell x_1\big)(F_i \boldsymbol{x}_1) \sim F_i \boldsymbol{x}_1 = \boldsymbol{\ell}_i. \tag{23}$$

In other words, the partial derivative of the multibody epipolar constraint with respect to $\boldsymbol{x}_2$ evaluated at $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is proportional to the epipolar line associated with $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in the second view. Similarly, the partial derivative of the multibody epipolar constraint with respect to $\boldsymbol{x}_1$ evaluated at $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is proportional to the epipolar line associated with $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in the first view. Therefore, given a set of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$, we can obtain its corresponding collection of N epipolar lines in the first and second views, $\{\boldsymbol{\ell}_1^j\}_{j=1}^N$ and $\{\boldsymbol{\ell}_2^j\}_{j=1}^N$, from the first-order derivatives of the multibody epipolar constraint. We illustrate this multibody epipolar transfer process with the following diagram

$$\begin{array}{ccccccc} & & & & & & \boldsymbol{x}_2 \\ & & & & & & \downarrow \\ \boldsymbol{x}_1 & \overset{\text{Veronese}}{\longmapsto} & \nu_n(\boldsymbol{x}_1) & \overset{\substack{\text{Epipolar}\\ \text{Transfer}}}{\longmapsto} & \mathcal{F}\nu_n(\boldsymbol{x}_1) & \overset{\substack{\text{Polynomial}\\ \text{Differentiation}}}{\longmapsto} & \boldsymbol{\ell}. \end{array}$$

The only case in which this process fails is when a particular image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ belongs to two or more motions, i.e., if $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = \boldsymbol{x}_2^T F_j \boldsymbol{x}_1 = 0$ for some $i \neq j = 1, .., n$, because in this case the multibody epipolar constraint has a repeated factor, hence its derivative at the image pair is zero. We summarize our discussion so far with the following theorem.

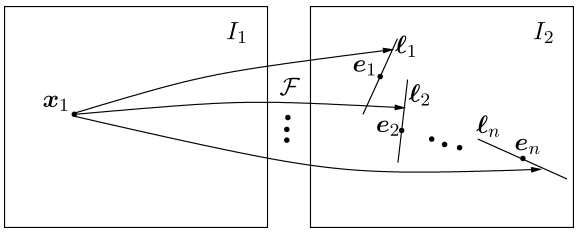

Figure 3. The multibody fundamental matrix $\mathcal{F}$ maps each point $\boldsymbol{x}_1$ in the first image to n epipolar lines $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n$ which pass through the n epipoles $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ respectively. One of these epipolar lines must pass through $\boldsymbol{x}_2$.

Theorem 4 (Epipolar lines from the multibody fundamental matrix)**.** *Let $\mathcal{F} \in \mathbb{R}^{M_n \times M_n}$ be a multibody fundamental matrix generated by n different fundamental matrices $\{F_i\}_{i=1}^n$. Also let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an image pair associated with only one of the motions, i.e., $\boldsymbol{x}_2^T(F_i - F_j)\boldsymbol{x}_1 \neq 0$ for $i \neq j = 1, \ldots, n$. Then one can compute the epipolar line $\boldsymbol{\ell}_1$ associated with the image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in the first view as*

$$\boldsymbol{\ell}_1 \sim \frac{\partial}{\partial \boldsymbol{x}_1}(\nu_n(\boldsymbol{x}_2)^T \mathcal{F}\nu_n(\boldsymbol{x}_1)), \tag{24}$$

and the epipolar line $\boldsymbol{\ell}_2$ associated with the image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in the second view as

$$\boldsymbol{\ell}_2 \sim \frac{\partial}{\partial \boldsymbol{x}_2}(\nu_n(\boldsymbol{x}_2)^T \mathcal{F} \nu_n(\boldsymbol{x}_1)). \tag{25}$$

Remark 4 (Computing derivatives of homogeneous polynomials). In order to obtain the epipolar lines, we need to compute the derivatives of a homogeneous polynomial of degree n, $q_n(\boldsymbol{x}) = \boldsymbol{c}^T \nu_n(\boldsymbol{x})$, with known $\boldsymbol{c}$. A simple calculation shows that $\frac{\partial q_n(\boldsymbol{x})}{\partial x_k} = \boldsymbol{c}^T \frac{\partial \nu_n(\boldsymbol{x})}{\partial x_k} = \boldsymbol{c}^T E_{nk} \nu_{n-1}(x)$, with $E_{nk} \in \mathbb{N}^{M_n \times M_{n-1}}$ a constant matrix containing some of the exponents of $\nu_n(\boldsymbol{x})$. Therefore, the computation of the derivatives is done algebraically, i.e., it does not involve taking derivatives of the (possibly noisy) data.

3.3. Estimating the Epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$

Given a set of epipolar lines, we now describe how to compute the epipoles. Recall that the (left) epipole associated with each (rank-2) fundamental matrix $F_i \in \mathbb{R}^{3\times 3}$ is defined as the vector $\boldsymbol{e}_i \in \mathbb{P}^2$ lying in the (left) null space of F_i, i.e., $\boldsymbol{e}_i^T F_i = 0$. Now let $\boldsymbol{\ell} \in \mathbb{P}^2$ be an arbitrary epipolar line associated with some image point in the first frame. Since $\boldsymbol{\ell}$ must pass through one of the epipoles (see Fig. 4), there exists an i such that $\boldsymbol{e}_i^T \boldsymbol{\ell} = 0$. Therefore, every epipolar line $\boldsymbol{\ell}$ has to satisfy the following polynomial constraint

$$p_n(\boldsymbol{\ell}) \doteq \left(\boldsymbol{e}_1^T \boldsymbol{\ell}\right)\left(\boldsymbol{e}_2^T \boldsymbol{\ell}\right) \ldots \left(\boldsymbol{e}_n^T \boldsymbol{\ell}\right) = \tilde{\boldsymbol{e}}^T \nu_n(\boldsymbol{\ell}) = 0, \tag{26}$$

regardless of the motion with which it is associated. We call the vector $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_n}$ the *multibody epipole* associated with the n motions. As before, $\tilde{\boldsymbol{e}}$ is a vector representation of the symmetric tensor product of the individual epipoles $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ and it is in general different from any of the embedded epipoles $\nu_n(\boldsymbol{e}_i)$, $i = 1, \ldots, n$.

Given a collection $\{\boldsymbol{\ell}^j\}_{j=1}^N$ of $N \geq M_n - 1$ epipolar lines computed by either polynomial factorization or polynomial differentiation, we can obtain the multibody epipole $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_n}$ as the solution to the linear system

$$P_n \tilde{\boldsymbol{e}} \doteq \begin{bmatrix} \nu_n(\boldsymbol{\ell}^1)^T \\ \nu_n(\boldsymbol{\ell}^2)^T \\ \vdots \\ \nu_n(\boldsymbol{\ell}^N)^T \end{bmatrix} \tilde{\boldsymbol{e}} = 0. \tag{27}$$

In order for Eq. (27) to have a unique solution (up to a scale factor), we will need to replace n by the number of different epipoles (up to a scale factor) as stated by the following theorem.

Theorem 5 (Number of different epipoles). *Let $\{\boldsymbol{\ell}^j\}_{j=1}^N$ be a set of epipolar lines corresponding to 3-D points in general configuration and undergoing n independent rigid-body motions with nonzero translation (relative to the camera). Let $P_i \in \mathbb{R}^{N \times M_i}$ be the matrix of embedded epipolar lines defined in (27), but computed using the Veronese map ν_i of degree $i \geq 1$. If the number of lines is big enough ($N \geq M_n - 1$ when n is known) and at least 2 lines pass through each epipole, then*

$$\operatorname{rank}(P_i) \begin{cases} > M_i - 1, & \text{if } i < n_e, \\ = M_i - 1, & \text{if } i = n_e, \\ < M_i - 1, & \text{if } i > n_e. \end{cases} \tag{28}$$

Therefore, the number of different epipoles $n_e \leq n$ is given by

$$n_e \doteq \min\{i : \operatorname{rank}(P_i) = M_i - 1\}. \tag{29}$$

Proof: See Vidal et al. (2005). □

Similarly to Remark 3, in the presence of noisy correspondences we may estimate the number of different epipoles as

$$n_e = \arg\min_{i \geq 1} \frac{\sigma_{M_i}^2(P_i)}{\sum_{k=1}^{M_i - 1} \sigma_k^2(P_i)} + \mu(M_i - 1), \tag{30}$$

where $\sigma_k(P_i)$ is the kth singular value of P_i and μ is a parameter. Once the number of different epipoles has been computed, the multibody epipole $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_{n_e}}$ can be obtained in a least-squares sense as the eigenvector of $P_{n_e}^T P_{n_e}$ associated with its smallest eigenvalue.

Once $\tilde{\boldsymbol{e}}$ has been computed, the rest of the problem is to compute the individual epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$. As illustrated in Fig. 4, each epipole $\boldsymbol{e}_i$ corresponds to the intersection of the epipolar lines associated with the ith motion. Since at this point we do not know yet which epipolar lines correspond to which motion, we cannot directly estimate the epipoles. However, we know that for each epipolar line $\boldsymbol{\ell}$ there exists an epipole $\boldsymbol{e}_i$ such

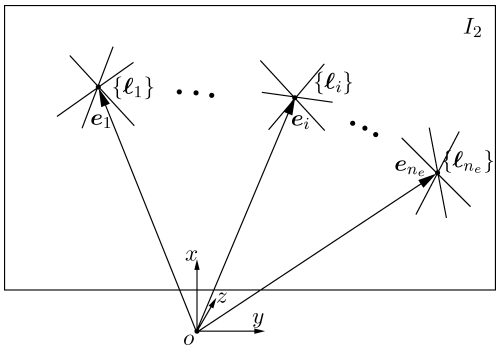

Figure 4. When n objects move independently, the epipolar lines in the second view associated with each image point in the first view form $n_e \leq n$ groups intersecting respectively at n_e different epipoles (up to scale) in the second view.

that $\boldsymbol{e}_i^T \boldsymbol{\ell} = 0$. Therefore, we have

$$\frac{\partial}{\partial \boldsymbol{\ell}}\big(p_{n_e}(\boldsymbol{\ell})\big) = \sum_{i=1}^{n_e} \prod_{\ell \neq i} \big(\boldsymbol{e}_\ell^T \boldsymbol{\ell}\big)\boldsymbol{e}_i = \prod_{\ell \neq i} \big(\boldsymbol{e}_\ell^T \boldsymbol{\ell}\big)\boldsymbol{e}_i \sim \boldsymbol{e}_i. \tag{31}$$

In other words, the partial derivative of $p_{n_e}(\boldsymbol{\ell})$ evaluated at an epipolar line $\boldsymbol{\ell}$ is proportional to the epipole associated with that epipolar line.

By applying (31) to the set of N epipolar lines $\{\boldsymbol{\ell}^j\}_{j=1}^N$ in the second view, we can compute their corresponding N left epipoles. In the absence of noise, only n_e out of the N epipoles are different up to scale, hence one may automatically obtain the n_e epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$ and the clustering of the epipolar lines. In the presence of noise, however, all the N epipoles will be different. Therefore, instead of computing one epipole for each one of the N epipolar lines, we can obtain the n_e epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$ directly by evaluating (31) at n_e lines $\{\boldsymbol{l}_i\}_{i=1}^{n_e}$ passing through the n_e epipoles. In principle, one could choose those n_e lines from the N epipolar lines, however we do not know which epipolar lines pass through which epipoles. We therefore choose the n_e lines $\{\boldsymbol{l}_i\}_{i=1}^{n_e}$ as follows. Let $\mathcal{L}_1 \in \mathbb{P}^2$ and $\mathcal{L}_2 \in \mathbb{P}^2$ be two randomly chosen lines in the plane. In order for the line $\mathcal{L}_1 + t\mathcal{L}_2$ with $t \in \mathbb{R}$ to pass through one of the n_e epipoles we must have that $p_{n_e}(\mathcal{L}_1 + t\mathcal{L}_2) = 0$. Therefore, we can choose n_e lines $\{\boldsymbol{l}_i\}_{i=1}^{n_e}$ passing through the n_e epipoles as

$$\boldsymbol{l}_i = \mathcal{L}_1 + t_i \mathcal{L}_2 \quad i = 1, \ldots n_e, \tag{32}$$

where $\{t_i \in \mathbb{R}\}_{i=1}^{n_e}$ are the n_e roots of the n_eth-degree univariate polynomial $p_{n_e}(\mathcal{L}_1 + t\mathcal{L}_2)$. We have shown the following result.

Theorem 6 (Epipoles from the multibody epipole)**.** *Let n_e and $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_{n_e}}$ be the number of different epipoles and the multibody epipole, respectively. Also let $\mathcal{L}_1 \in \mathbb{P}^2$ and $\mathcal{L}_2 \in \mathbb{P}^2$ be two randomly chosen lines in the plane. The individual epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$ can be obtained from the derivatives of $p_{n_e}(\boldsymbol{\ell}) = \tilde{\boldsymbol{e}}^T \nu_{n_e}(\boldsymbol{\ell})$ as*

$$\boldsymbol{e}_i = \left.\frac{D p_{n_e}(\boldsymbol{\ell})}{\|D p_{n_e}(\boldsymbol{\ell})\|}\right|_{\boldsymbol{\ell}=\boldsymbol{l}_i} \qquad i = 1, \ldots, n_e, \tag{33}$$

where $\{l_i = \mathcal{L}_1 + t_i \mathcal{L}_2\}_{i=1}^{n_e}$ and $\{t_i \in \mathbb{R}\}_{i=1}^{n_e}$ are the n_e roots of the univariate polynomial $p_{n_e}(\mathcal{L}_1 + t\mathcal{L}_2)$.

Remark 5 (Estimation of epipoles using GPCA). Since each epipolar line must pass through one of the epipoles, if we work in homogeneous coordinates then we may interpret each epipolar line as a vector $\boldsymbol{\ell} \in \mathbb{R}^3$ lying in one out of n_e planes with normal vectors $\{\boldsymbol{e}_i \in \mathbb{P}^2\}_{i=1}^{n_e}$. Therefore, the problem of estimating the n_e epipoles from N epipolar lines is equivalent to the problem of clustering N data points lying on n_e planes in $\mathbb{R}^3$. The GPCA algorithm of Vidal et al. (2004) gives an algebraic solution to the more general problem of clustering data lying on a collection of subspaces. Theorem 6 is in essence a special case of GPCA.

Remark 6 (Estimation of epipoles in the presence of noise). Notice that Theorem 6 is also applicable when the correspondences are corrupted with a moderate level of noise. As we alluded to earlier, one can estimate the multibody epipole in a least-squares sense and the derivatives of $p_{n_e}(\boldsymbol{\ell})$ as described in Remark 4. The roots of $p_{n_e}(\mathcal{L}_1 + t\mathcal{L}_2)$ can also be computed, except that one may obtain complex roots when two epipoles are similar or when the noise level is high. We refer the interested reader to Vidal et al. (2004) for variations of the GPCA algorithm that choose the lines $\{\boldsymbol{l}_i\}_{i=1}^{n_e}$ in a way that does not require solving for the roots of a univariate polynomial.

Remark 7 (Left and right epipoles). Notice that one may compute the set of right epipoles, $\{\boldsymbol{e}_{i1}\}_{i=1}^{n}$ such

that $F_i \boldsymbol{e}_{i1} = 0$, in a similar fashion. We can just apply Theorem 6 to the multibody epipole $\tilde{\boldsymbol{e}}$ obtained by solving the linear system in (27) from a set of epipolar lines in the first (rather than the second) view.

3.4. Estimating the Individual Fundamental Matrices $\{F_i\}_{i=1}^n$

Given the epipolar lines $\{\boldsymbol{\ell}^j\}_{j=1}^N$ and the epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$, we now show how to recover each one of the individual fundamental matrices $\{F_i\}_{i=1}^n$. To avoid degenerate cases, we will assume that all the epipoles are different up to scale,[6] i.e.,

$$n_e = n. \tag{34}$$

Since at this point both epipolar lines and epipoles are known, we may cluster the epipolar lines and/or the correspondences according to the motion they belong. For instance, we can assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ and epipolar line $\boldsymbol{\ell}^j$ to group i if

$$i = \arg\min_{\ell=1,\ldots,n} \left(\boldsymbol{e}_\ell^T \boldsymbol{\ell}^j\right)^2. \tag{35}$$

Once the epipolar lines and/or the correspondences have been clustered, we can compute a fundamental matrix F_i for each group by using either of the following procedures:

1. *Fundamental matrices from eight-point algorithm*: apply the eight-point algorithm to the image pairs in group i, where $i = 1, \ldots, n$.
2. *Fundamental matrices from epipolar lines*: If the image point $\boldsymbol{x}_1^j$ belongs to group i, then we must have $\boldsymbol{\ell}^j \sim F_i \boldsymbol{x}_1^j$. Therefore, we can compute fundamental matrix F_i by solving the set of equations

$$[\boldsymbol{\ell}^j]_\times F_i \boldsymbol{x}_1^j = 0 \quad \text{for all } j \text{ in group } i = 1, \ldots, n. \tag{36}$$

Notice that both of these procedures can also be applied in the case of noisy correspondences by solving for the individual fundamental matrices in a least-squares sense.

3.5. Clustering the Feature Points

Feature clustering refers to the problem of assigning each image pair $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$, to the motion it corresponds. In the previous section we already showed how to solve this problem by using epipoles and epipolar lines. More specifically, we assigned an image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to group i if

$$i = \arg\min_{\ell=1,\ldots,n} \left(\boldsymbol{e}_\ell^T \boldsymbol{\ell}^j\right)^2. \tag{37}$$

Alternatively, one can also cluster the feature points by using the already computed fundamental matrices. For example, we can assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to group i if

$$i = \arg\min_{\ell=1,\ldots,n} \left(\boldsymbol{x}_2^{jT} F_\ell \boldsymbol{x}_1^j\right)^2. \tag{38}$$

In the presence of noise, the square of the epipolar constraint is only an algebraic way of measuring how close an image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is to satisfying the epipolar constraint. We therefore assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to the group i that minimizes the square of the Sampson error (Hartley and Zisserman, 2000), also known as the normalized epipolar constraint (Ma et al., 2001), i.e.,

$$i = \arg\min_{\ell=1,\ldots,n} \left(E_S(F_\ell) \doteq \frac{\left(\boldsymbol{x}_2^{jT} F_\ell \boldsymbol{x}_1^j\right)^2}{\left\|[e_3]_\times F_\ell^T \boldsymbol{x}_2^j\right\|^2 + \left\|[e_3]_\times F_\ell \boldsymbol{x}_1^j\right\|^2} \right), \tag{39}$$

where $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$.

3.6. Two-View Multibody Structure from Motion Algorithm

We are now ready to present an algebraic geometric algorithm for multibody motion estimation and segmentation from two perspective views. Given a collection of $N \geq M_n^2 - 1$ image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$, Algorithm 1 determines the number of independent motions n, the individual fundamental matrices $\{F_i\}_{i=1}^n$ and the clustering of the image pairs. Therefore, Algorithm 1 is a natural generalization of the eight-point algorithm to the case of multiple rigid-body motions.

Remark 8 (Comments about the algorithm).

1. **Algebraic solvability.** The only nonlinear part of Algorithm 1 is to solve for the roots of a univariate polynomial of degree n in Step 5. Therefore, the multibody structure from motion problem is algebraically solvable (i.e., there is a closed form solution) if and only if the number of motions is such that $n \leq 4$ (see (Lang, 1993)). When $n \geq 5$, the above algorithm must rely on a numerical solution for the roots of those polynomials.
2. **Repeated epipoles.** If two individual fundamental matrices share the same (left) epipoles, we cannot segment the epipolar lines as described in Step 6 of Algorithm 1. In this case, one can consider the right epipoles (in the first image frame) instead, since it is extremely rare that two motions give rise to the same left and right epipoles.[7]
3. **Special motions and structures.** Algorithm 1 works for independent motions with nonzero translation and different epipoles and for 3-D points in general configuration. Future research is needed for special motions such as pure rotation or repeated left and right epipoles, and for special structures, e.g., planar objects. We refer the interested reader to Vidal and Ma (2004) for the study of some of these cases, such as pure rotation and planar structures, by using homographies rather than fundamental matrices to model each rigid-body motion.
4. **Minimum number of feature points.** A drawback of Algorithm 1 is that it needs a lot of image pairs in order to compute the multibody fundamental matrix, which often makes it impractical for large n. However, one can significantly reduce the data requirements by incorporating partial knowledge about the motion or the segmentation of the objects with minor changes in the general algorithm. We now discuss a few of such possible variations.

   (a) **Multiple linearly moving objects.** In many practical situations, the motion of the objects can be well approximated by a linear motion, i.e., there is only translation but no rotation. In this case, the epipolar constraint reduces to $\boldsymbol{x}_2^T[\boldsymbol{e}_i]_\times \boldsymbol{x}_1 = 0$ or $\boldsymbol{e}_i^T[\boldsymbol{x}_2]_\times \boldsymbol{x}_1 = 0$, where $\boldsymbol{e}_i \in \mathbb{P}^2$ represents the epipole associated with the ith motion, $i = 1, \ldots, n$. Therefore, the vector $\boldsymbol{\ell} = [\boldsymbol{x}_2]_\times \boldsymbol{x}_1 \in \mathbb{P}^2$ is an epipolar line satisfying the polynomial $p_n(\boldsymbol{\ell}) = (\boldsymbol{e}_1^T\boldsymbol{\ell})(\boldsymbol{e}_2^T\boldsymbol{\ell})\ldots(\boldsymbol{e}_n^T\boldsymbol{\ell}) = 0$. Given a set of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$

Algorithm 1 *(Two-view multibody structure from motion algorithm).*

Given a collection of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ corresponding to N points undergoing n different rigid-body motions relative to a moving perspective camera, recover the number of independent motions n, the fundamental matrix F_i associated with motion $i = 1, \ldots, n$, and the object to which each image pair belongs as follows:

1. **Number of motions.** Apply the Veronese map ν_i of degree $i = 1, 2, \ldots, n$ to the image points $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ to form the embedded data matrix $\boldsymbol{V}_i$ in (15). Compute the number of independent motions n from a rank constraint on $\boldsymbol{V}_i$ as described in Remark 3.
2. **Multibody fundamental matrix.** Compute the multibody fundamental matrix $\mathcal{F}$ as the least-squares solution to the linear system $\boldsymbol{V}_n \boldsymbol{f} = 0$ in (15), where $\boldsymbol{V}_n$ is computed using the Veronese map ν_n of degree n.
3. **Epipolar transfer.** Compute the epipolar lines $\{\boldsymbol{\ell}^j\}_{j=1}^N$ in the second view associated with each image pair $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ from the partial derivative of the multibody epipolar constraint with respect to $\boldsymbol{x}_2$ evaluated at each image pair, as described in Theorem 4.
4. **Multibody epipole.** Use the epipolar lines $\{\boldsymbol{\ell}^j\}_{j=1}^N$ to estimate the multibody epipole $\tilde{\boldsymbol{e}}$ as the coefficients of the polynomial $p_n(\boldsymbol{\ell})$ in (26) by solving the linear system $P_n\tilde{\boldsymbol{e}} = 0$ in (27), where P_n is the matrix of embedded epipolar lines.
5. **Individual epipoles.** Compute the individual epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$ from the multibody epipole $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_n}$ by evaluating the derivatives of $p_n(\boldsymbol{\ell}) = \tilde{\boldsymbol{e}}^T\nu_n(\boldsymbol{\ell})$ at the n lines $\{l_i = \mathcal{L}_1 + t_i\mathcal{L}_2\}_{i=1}^n$, where $\mathcal{L}_1$ and $\mathcal{L}_2$ are chosen at random and $\{t_i\}_{i=1}^n$ are the roots of the univariate polynomial $p_n(\mathcal{L}_1 + t\mathcal{L}_2)$.
6. **Individual fundamental matrices.** Assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to motion $i = \arg\min_{\ell=1,\ldots,n}(\boldsymbol{e}_\ell^T\boldsymbol{\ell}^j)^2$. Then obtain the individual fundamental matrices $\{F_i\}_{i=1}^n$ by applying the eight-point algorithm to each group, as described in Section 3.4.
7. **Feature clustering from fundamental matrices.** Assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to motion i if

$$i = \arg\min_{\ell=1,\ldots,n} \frac{\left(\boldsymbol{x}_2^{jT} F_\ell \boldsymbol{x}_1^j\right)^2}{\left\|[e_3]_\times F_\ell^T \boldsymbol{x}_2^j\right\|^2 + \left\|[e_3]_\times F_\ell \boldsymbol{x}_1^j\right\|^2}.$$

corresponding to N points undergoing n different linear motions $\{\boldsymbol{e}_i \in \mathbb{P}^2\}_{i=1}^n$, one can use the set of epipolar lines $\{\boldsymbol{\ell}^j = [\boldsymbol{x}_2^j]_\times \boldsymbol{x}_1^j\}_{j=1}^N$ to estimate the epipoles $\boldsymbol{e}_i$ using Steps 4 and 5 of Algorithm 1. Notice that the epipoles are recovered directly using polynomial differentiation *without* estimating the multibody fundamental matrix $\mathcal{F}$ first. Furthermore, given the epipoles, the fundamental matrices are trivially obtained as $F_i = [\boldsymbol{e}_i]_\times$. The clustering of the image points is then obtained from Step 7 of Algorithm 1. We conclude that if the motions are linear, we only need $N = M_n - 1$ image pairs versus $N = M_n^2 - 1$ needed in the general case. So when n is large, the total number of image pairs needed grows as $O(n^2)$ for the linear motion case versus $O(n^4)$ for the general case. Therefore, the number of feature points that need to be tracked on each object grows linearly in the number of independent motions. For instance, when $n = 10$, one only needs to track 7 points on each object, which is a mild requirement given that the case $n = 10$ occurs rather rarely in most applications.

(b) **Constant motions.** In many vision and control applications, the motion of the objects in the scene changes slowly relative to the image sampling rate. Thus, if the sampling rate is even, we may assume that for a number of image frames, say m, the motion of each object between consecutive pairs of images is the same. Hence *all* the feature points corresponding to the $m - 1$ image pairs in between can be used to estimate the *same* multibody fundamental matrix. In essence, this corresponds to segmenting the trajectories of the image points rather than the points themselves. For example, when $m = 5$ and $n = 4$, we only need to track $(M_4^2 - 1)/4 = 225/4 \approx 57$ image points between each of the 4 consecutive pairs of images instead of 255. That is about $57/4 \approx 15$ features on each object on each image frame, which is easier to achieve in practice. In general if $m = O(n)$, $O(n^2)$ feature points per object need to be tracked in each image. For example, when $m = n + 1 = 6$, one needs to track about 18 points on each object, which is not so demanding given the nature of the problem.

(c) **Internal structure of $\mathcal{F}$.** The only step of Algorithm 1 that requires $O(n^4)$ image pairs is the estimation of the multibody fundamental matrix $\mathcal{F}$. Step 2 requires a lot of data points, because $\mathcal{F}$ is estimated linearly without taking into account the rich internal algebraic structure of $\mathcal{F}$ (e.g., $\text{rank}(\mathcal{F}) \leq M_n - n$). Future research is needed to reduce the minimum number of image pairs by considering constraints among entries of $\mathcal{F}$, in the same spirit that the 8-point algorithm for $n = 1$ can be reduced to 7 points if the algebraic property $\det(F) = 0$ is used.

5. **Noise sensitivity.** Algorithm 1 gives a purely algebraic solution to the multibody structure from motion problem. Future research is needed to address the sensitivity of the algorithm to noise in the image measurements. In particular, one should pay attention to Step 2, which is sensitive to noise, because it does not exploit the algebraic structure of the multibody fundamental matrix $\mathcal{F}$.
6. **Optimality.** Notice that linearly solving for the multibody fundamental matrix through the Veronese embedding is sub-optimal from a statistical point of view. We refer the interested reader to Vidal and Sastry (2003) for the derivation of the optimal function for motion estimation and segmentation in the case where the correspondences are corrupted with i.i.d. zero-mean Gaussian noise.

At the end of our theoretical development, Table 1 summarizes our results with a comparison of the geometric entities associated with two views of 1 rigid-body motion and two views of n rigid-body motions.

4. Experimental Results

In this section, we evaluate the performance of our motion segmentation algorithm with respect to the amount of noise in the image measurements. We also present experimental results on the segmentation of an indoor sequence.

We first test the algorithm on synthetic data. We randomly pick $n = 2$ collections of $N = 100n$ feature points and apply a different (randomly chosen) rigid body motion $(R_i, T_i) \in SE(3)$, with $R_i \in SO(3)$ the rotation and $T_i \in \mathbb{R}^3$ the translation. We add zero-mean Gaussian noise with standard deviation (std) from 0 to 1 pixels to the images $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. The image size is 1000 pixels. We run 1000 trials for each noise level. For each trial, the classification error is computed as the percentage of misclassified points, and the error between the true motions $\{(R_i, T_i)\}_{i=1}^n$ and their estimates

Table 1. Comparison between the geometry for two views of 1 rigid-body motion and the geometry of n rigid-body motions.

Comparison of	2 views of 1 body	2 views of n bodies
An image pair	$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^3$	$\nu_n(\boldsymbol{x}_1), \nu_n(\boldsymbol{x}_2) \in \mathbb{R}^{M_n}$
Epipolar constraint	$\boldsymbol{x}_2^T F \boldsymbol{x}_1 = 0$	$\nu_n(\boldsymbol{x}_2)^T \mathcal{F} \nu_n(\boldsymbol{x}_1) = 0$
Fundamental matrix	$F \in \mathbb{R}^{3\times 3}$	$\mathcal{F} \in \mathbb{R}^{M_n \times M_n}$
Linear estimation from N image pairs	$\begin{bmatrix} \boldsymbol{x}_2^1 \otimes \boldsymbol{x}_1^1 \\ \boldsymbol{x}_2^2 \otimes \boldsymbol{x}_1^2 \\ \vdots \\ \boldsymbol{x}_2^N \otimes \boldsymbol{x}_1^N \end{bmatrix} \boldsymbol{f} = 0$	$\begin{bmatrix} \nu_n(\boldsymbol{x}_2^1) \otimes \nu_n(\boldsymbol{x}_1^1) \\ \nu_n(\boldsymbol{x}_2^2) \otimes \nu_n(\boldsymbol{x}_1^2) \\ \vdots \\ \nu_n(\boldsymbol{x}_2^N) \otimes \nu_n(\boldsymbol{x}_1^N) \end{bmatrix} \boldsymbol{f} = 0$
Epipole	$\boldsymbol{e}^T F = 0$	$\nu_n(\boldsymbol{e})^T \mathcal{F} = 0$
Epipolar lines	$\boldsymbol{\ell} = F\boldsymbol{x}_1 \in \mathbb{R}^3$	$\tilde{\boldsymbol{\ell}} = \mathcal{F}\nu_n(\boldsymbol{x}_1) \in \mathbb{R}^{M_n}$
Epipolar line & point	$\boldsymbol{x}_2^T \boldsymbol{\ell} = 0$	$\nu_n(\boldsymbol{x}_2)^T \tilde{\boldsymbol{\ell}} = 0$
Epipolar line & epipole	$\boldsymbol{e}^T \boldsymbol{\ell} = 0$	$\tilde{\boldsymbol{e}}^T \nu_n(\boldsymbol{\ell}) = 0$

$\{(\hat{R}_i, \hat{T}_i)\}_{i=1}^n$ are computed as

$$\text{Rotation error} = \frac{1}{n}\sum_{i=1}^{n} \left| \text{acos}\left(\frac{\text{trace}(R_i \hat{R}_i^T) - 1}{2} \right) \right| \text{ degrees}$$

$$\text{Translation error} = \frac{1}{n}\sum_{i=1}^{n} \left| \text{acos}\left(\frac{T_i^T \hat{T}_i}{\|T_i\| \|\hat{T}_i\|} \right) \right| \text{ degrees}$$

Figure 5 plots the mean classification error, the rotation error and the translation error (degrees) as a function of noise. In all trials the number of motions was correctly estimated from Eq. (19) as $n = 2$.[8] The mean classification error is less than 7% using an assignment based on epipoles and epipolar lines, and can be reduced to about 3.25% using an assignment based on the Sampson error. The rotation error is less than 0.38° and the translation error is less than 0.83°.

We also tested the proposed approach by segmenting a real sequence in which a moving camera observes a can moving in front of a static background consisting of a T-shirt and a book. We manually extracted a total of $N = 170$ correspondences: 70 for the can and 100 for the background. For comparison purposes, we estimated the ground truth motion (R_i, T_i) by applying the eight-point algorithm to manually segmented correspondences. Figure 6 shows the first frame of the sequence as well as the relative displacement of the correspondences between the two frames. We applied Algorithm 1 to estimate the number of motions as $n = 2$.[9] We obtained a misclassification error of 5.88% when the clustering is obtained using epipolar lines and epipoles only. We used this segmentation to obtain the motion parameters for each group. The error in rotation was 0.07° for the background and 4.12° for the can.

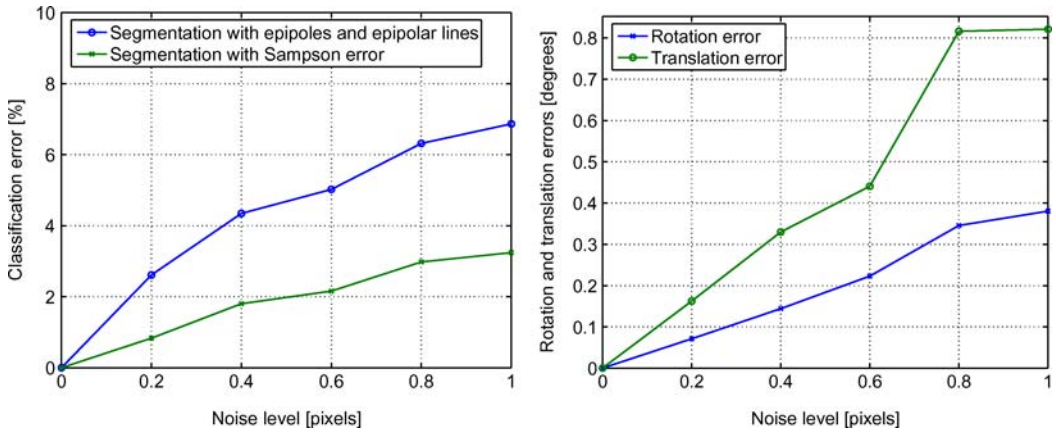

Figure 5. Percentage of correct classification and error in the estimation of rotation and translation (in degrees).

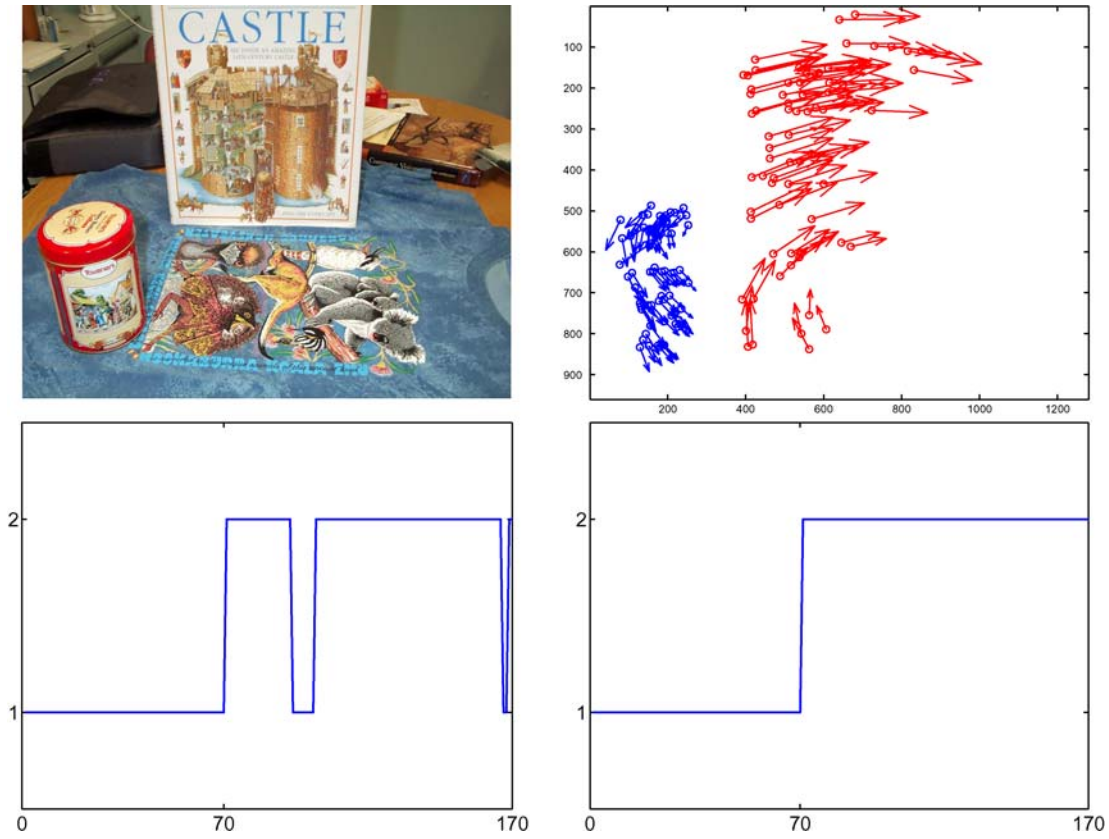

Figure 6. Top: first frame of a sequence with two rigid-body motions—the can and the background—and the 2-D displacements of the 140 correspondences from the first view ('$\circ$') to the second ('$\rightarrow$'). Bottom: segmentation of the 170 correspondences using epipoles and epipolar lines (left) and using Sampson distance (right).

The error in translation was 0.21° for the background and 4.51° for the can. Given the motion parameters for each group, we re-clustered the features using the Sampson error (39). The misclassification error reduced to 0%.

5. Discussions, Conclusions and Future Work

We have presented a novel geometric approach for the analysis of dynamic scenes containing multiple rigidly moving objects seen in two perspective views. Instead of iterating between feature clustering and single body motion estimation, our approach eliminates the clustering problem and solves directly for the motion parameters. This is achieved by exploiting the algebraic and geometric properties of the so-called *multibody epipolar constraint* and its associated *multibody fundamental matrix*, which are natural generalizations of the epipolar constraint and of the fundamental matrix to multiple moving objects. Overall, the proposed algorithm provides a principled solution to the problem and paves the way to a more systematic study of its many variations, such as special motions, special structures, multiple frames, etc.

Issues such as the effect of noise, outliers, incorrect correspondences and missing data have not been systematically studied. The present algorithm can tolerate a moderate amount of noise, provided that the multibody fundamental matrix is well estimated. However, further research is needed to improve the estimation of the number of motions and of the multibody fundamental matrix, e.g., by incorporating its internal algebraic structure in the current approach. We refer the reader to Vidal and Sastry (2003) for more details on the estimation of the multibody fundamental matrix in the presence of zero-mean Gaussian noise. Our discussion has also suggested that the use of multiple images may reduce the amount of feature points needed from each image (pair), thus improving the performance of the

algorithm. We have shown that this is indeed the case in Hartley and Vidal (2004), where we considered the motion segmentation problem in the case of three perspective views. The case of multiple affine views with missing correspondences can be found in Vidal and Hartley (2004).

Notes

1. The relationships between epipoles and epipolar lines will be studied in the next section, where we will show how both of them can be computed from the derivatives of the multibody epipolar constraint.
2. This is simply because the M_n monomials in $\nu_n(\boldsymbol{x})$ are linearly independent.
3. This assumption is not always satisfied, e.g., for $n = 3$ motions with epipoles along the X, Y and Z axes. However, as long as the $\boldsymbol{e}_i$'s are different up to scale, one can always find a non-singular linear transformation $\boldsymbol{e}_i \mapsto T\boldsymbol{e}_i$ on $\mathbb{R}^3$ that makes the assumption true. Furthermore, this linear transformation induces a linear transformation on the lifted space $\mathbb{R}^{M_n}$ that preserves the rank of the matrix U.
4. For $n = 1$, $D\nu_n(\boldsymbol{x})$ is clearly full rank. For $n \geq 2$, notice that if $\boldsymbol{y} = [y_1, y_2, y_3]^T$ is such that $D\nu_n(\boldsymbol{x})\boldsymbol{y} = 0$, then $nx_j^{n-1}y_j = 0$ for $j = 1, \ldots, 3$. Thus $\boldsymbol{y} = 0$ if all entries of $\boldsymbol{x}$ are nonzero. One can show that $\boldsymbol{y} = 0$ even when one or two entries of $\boldsymbol{x}$ are zero by properly choosing additional equations from $D\nu_n(\boldsymbol{x})\boldsymbol{y} = 0$.
5. This is the so-called (real) Veronese surface in Algebraic Geometry (Harris, 1992).
6. Notice that this is not a strong assumption. If two individual fundamental matrices share the same (left) epipoles, one can consider the right epipoles (in the first image frame) instead, because it is extremely rare that two motions give rise to the same left and right epipoles. In fact, this happens only when the rotation axes of the two motions are equal to each other and parallel to the translation direction.
7. This happens only when the rotation axes of the two motions are equal to each other and parallel to the translation direction.
8. We use $\mu = 5 \times 10^{-3}$ in equation (19) for computing the number of motions.
9. We use $\mu = 5 \times 10^{-3}$ in equation (19) for computing the number of motions.

References

Avidan, S. and Shashua, A. 2000.Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):348–357.

Costeira, J. and Kanade, T. 1995. Multibody factorization methods for motion analysis, In *International Conference on Computer Vision*, pp. 1071–1076.

Fan, X. and Vidal, R. 2005. The space of multibody fundamental matrices: Rank geometry and projection. In *ICCV Workshop on Dynamical Vision*.

Feng, X. and Perona, P. 1998. Scene segmentation from 3D motion, In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 225–231.

Han, M. and Kanade, T. 2000. Reconstruction of a scene with multiple linearly moving objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 542–549.

Harris, J. 1992. *Algebraic Geometry: A First Course*, Springer-Verlag.

Hartley R. and Zisserman A. 2000. *Multiple View Geometry in Computer Vision*, Cambridge.

Hartley, R. and Vidal, R. 2004. The multibody trifocal tensor: Motion segmentation from 3 perspective views, In *IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pp. 769–775.

Kanatani, K. 2001. Motion segmentation by subspace separation and model selection, In *International Conference on Computer Vision*, vol. 2, pp. 586–591.

Kanatani, K. 2002. Evaluation and selection of models for motion segmentation, In *Asian Conference on Computer Vision*, pp. 7–12.

Lang, S. 1993. *Algebra*, Addison-Wesley Publishing Company, 3rd edition,

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

Ma, Y., Košecká, J. and Sastry, S. 2001. Optimization criteria and geometric algorithms for motion and structure estimation, *International Journal of Computer Vision*, 44(3):219–249.

Ma, Y. Soatto, S., Kosecka, J., and Sastry, S. 2003. *An Invitation to 3D Vision: From Images to Geometric Models*, Springer Verlag.

Shashua, A. and Levin, A. 2001. Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In *International Conference on Computer Vision*, vol. 2 :592–599.

Sturm, P. 2002. Structure and motion for dynamic scenes—the case of points moving in planes. In *European Conference on Computer Vision*, pp. 867–882.

Torr, P., Szeliski, R., and Anandan, P. 2001. An integrated Bayesian approach to layer extraction from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):297–303.

Torr, P. H. S. 1998. Geometric motion segmentation and model selection, *Philosophical Transactions Royal Society of London*, 356(1740):1321–1340.

Vidal, R. and Hartley, R. 2004. Motion segmentation with missing data by PowerFactorization and Generalized PCA. In *IEEE Conference on Computer Vision and Pattern Recognition*, vol. II, pp. 310–316.

Vidal, R. and Ma, Y. 2004. A unified algebraic approach to 2-D and 3-D motion segmentation, In *European Conference on Computer Vision*, pp. 1–15.

Vidal, R., Ma, Y., and Piazzi. J. 2004. A new GPCA algorithm for clustering subspaces by fitting, differentiating and dividing polynomials. In *IEEE Conference on Computer Vision and Pattern Recognition*, (1):510–517.

Vidal, R., Ma, Y., and Sastry, S. 2005. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1–15.

Vidal, R., Ma, Y., and Sastry, S. 2003. Generalized Principal Component Analysis (GPCA). In *IEEE Conference on Computer Vision and Pattern Recognition*, (1):621–628.

Vidal, R. and Sastry, S. 2003. Optimal segmentation of dynamic scenes from two perspective views. In *IEEE Conference*

on Computer Vision and Pattern Recognition, (2):281–286.

Vidal, R., Soatto, S., Ma, Y., and Sastry, S. 2002. Segmentation of dynamic scenes from the multibody fundamental matrix. In *ECCV Workshop on Visual Modeling of Dynamic Scenes*.

Vidal, R., Soatto, S., and Sastry, S. 2002. Two-view segmentation of dynamic scenes from the multibody fundamental matrix. *Technical Report*, UCB/ERL M02/02, UC Berkeley.

Wolf, L. and Shashua, A. 2001. Affine 3-D reconstruction from two projective images of independently translating planes. In *International Conference on Computer Vision*, pp. 238–244.

Wolf, L. and Shashua, A. 2001. Two-body segmentation from two perspective views. In *Conference on Computer Vision and Pattern Recognition*, pp. 263–270.