ECCV Workshop on Vision and Modeling of Dynamic Scenes, Copenhagen, Denmark, May 2002

Segmentation of Dynamic Scenes from the Multibody Fundamental Matrix*

René Vidal
Dept. of EECS, UC Berkeley
Berkeley, CA 94720

Stefano Soatto
CS Department, UCLA
Los Angeles, CA 90095

Yi Ma
ECE Dept., UIUC
Urbana, IL 61801

Shankar Sastry
Dept. of EECS, UC Berkeley
Berkeley, CA 94720

Abstract

We present a geometric approach for the analysis of dynamic scenes containing multiple rigidly moving objects seen 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 moving objects. We derive a rank constraint on the image points from which one can estimate the number of independent motions and linearly solve for the multibody fundamental matrix. We prove that the epipoles of each independent motion lie exactly in the intersection of the left null space of the multibody fundamental matrix with the so-called Veronese surface. We then show that individual epipoles and epipolar lines can be uniformly and efficiently computed by using a novel polynomial factorization technique. Given the epipoles and epipolar lines, the estimation of individual fundamental matrices becomes a *linear* problem. Then, motion and feature point segmentation is automatically obtained from either the epipoles and epipolar lines or the individual fundamental matrices. We test the proposed approach by segmenting a real sequence.

Keywords: Multibody structure from motion, multibody epipolar constraint, multibody fundamental matrix, motion segmentation.

1 Introduction

Motion is one of the most important cues for segmenting an image sequence into different objects. Classical approaches to 2-D motion segmentation try to separate the image flow into different regions either by looking for flow discontinuities [Spoerri and Ullman 1987], while imposing some regularity conditions [Black and Anandan 1991], or by fitting a mixture of probabilistic models [Jepson and Black 1993]. The latter is usually done using an iterative process that alternates between segmentation and motion estimates using the Expectation-Maximization algorithm. Alternative approaches are based on local features that incorporate spatial and temporal motion information. Similar features are grouped together using, for example, normalized cuts [Shi and Malik 1998] or the eigenvectors of a similarity matrix [Weiss 1999].

3-D motion segmentation and estimation based on 2-D imagery is a more recent problem and various special cases have been analyzed using a geometric approach: 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 moving objects seen by an orthographic camera [Costeira and Kanade 1995; Kanatani 2001], self-calibration from multiple motions [Fitzgibbon and Zisserman 2000; Han and Kanade 2001], or two-object segmentation from two perspective views [Wolf and Shashua 2001]. Alternative probabilistic approaches to 3-D motion segmentation are based on model selection techniques [Torr 1998; Kanatani 2001] or combine normalized cuts with a mixture of probabilistic models [Feng and Perona 1998].

*This work is supported by grants ONR N00014-00-1-0621, ARO DAAD19-99-1-0139 and UIUC ECE Department startup fund.

This paper presents a geometric approach for the analysis of dynamic scenes containing an arbitrary number of rigidly moving objects seen in two perspective views. In section 2 we introduce the *multibody epipolar constraint* as a geometric relationship between the motion of the objects and the image points that is satisfied by all the image points, regardless of the body to which they belong. The multibody epipolar constraint defines the so-called *multibody fundamental matrix*, which is a generalization of the fundamental matrix to multiple bodies. Section 3 derives a rank constraint on the image measurements from which one can estimate the number of motions and linearly solve for the multibody fundamental matrix after embedding all the image points in a higher-dimensional space. In Section 4 we prove that the epipoles of each independent motion lie exactly in the intersection of the left null space of the multibody fundamental matrix with the so-called Veronese surface.

A complete solution and an algorithm for segmentation and estimation of multiple motions is presented in Section 5, where we show that individual epipoles and epipolar lines can be uniformly and efficiently computed using a novel polynomial factorization technique. Given the epipoles and the epipolar lines, the estimation of the individual fundamental matrices becomes a simple *linear* problem. Then, motion and feature point segmentation is automatically obtained from either the epipoles and epipolar lines or the individual fundamental matrices. We present preliminary results on the segmentation of a real image sequence in Section 6.

2 The multibody epipolar constraint and the multibody fundamental matrix

Consider two images of a scene containing an *unknown* number n of independent and rigidly moving objects. The motion of each object relative to the camera between the two frames is described by the fundamental matrix $F_i \in \mathbb{R}^{3\times 3}$ associated with the motion of object $i = 1,\dots,n$. We assume that the motions of the objects are such that all the fundamental matrices are distinct and different from zero, and hence the relative translation between the two image frames is non-zero. The image of a point $q^j \in \mathbb{R}^3$ with respect to image frame I_k is denoted as $\boldsymbol{x}_k^j \in \mathbb{P}^2$, for $j = 1,\dots,N$ and $k = 1, 2$. In order to avoid degenerate cases, we will assume that the image points correspond to 3-D points in general configuration in $\mathbb{R}^3$, *i.e.* they 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 always use the homogeneous representation $\boldsymbol{x} = [x, y, z]^T \in \mathbb{R}^3$ to refer to an arbitrary image point in $\mathbb{P}^2$. We define the *multibody structure from motion problem* as follows:

Problem 1 (Multibody structure from motion problem) *Given a set of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ corresponding to an unknown number of independent and rigidly moving objects that satisfy the assumptions above, estimate the number of independent motions n, the fundamental matrices $\{F_i\}_{i=1}^n$, and the segmentation of the image pairs, i.e. the object to which each image pair belongs.*

Let $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be an arbitrary image pair corresponding to any motion. Then, there exists an fundamental matrix F_i satisfying the epipolar constraint $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0$. Thus, regardless of the object to which the image pair belongs, the following constraint must be satisfied by the number of independent motions n, the relative motions of the objects $\{F_i\}_{i=1}^n$ and the image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$

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

We call this constraint the *multibody epipolar constraint*, since it is a natural generalization of the epipolar constraint valid for $n = 1$.

The multibody epipolar constraint converts Problem 1 into that of solving for the number of independent motions n and the fundamental matrices $\{F_i\}_{i=1}^n$ from the *nonlinear* equation (1). This nonlinear constraint defines a homogeneous polynomial of degree n in either $\boldsymbol{x}_1$ or $\boldsymbol{x}_2$. For example, if we let $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T$, then equation (1) 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

$$M_n \doteq (n+1)(n+2)/2 \tag{2}$$

different monomials. Thus, we can use the Veronese map of degree n, $\nu_n : \mathbb{P}^2 \to \mathbb{P}^{M_n - 1}$, $[x, y, z]^T \mapsto [x_1^n, x_1^{n-1}y_1, x_1^{n-1}z_1, \ldots, z_1^n]^T$, to write the multibody epipolar constraint (1) in bilinear form as stated by the following Lemma.

Lemma 1 (The bilinear multibody epipolar constraint)
The multibody epipolar constraint (1) can be written as

$$\boxed{\nu_n(\boldsymbol{x}_2)^T F \nu_n(\boldsymbol{x}_1) = 0,} \tag{3}$$

where $F \in \mathbb{R}^{M_n \times M_n}$ is a matrix representation of the symmetric tensor product of all the fundamental matrices $\{F_i\}_{i=1}^n$.

Proof: See [Vidal et al. 2002] for the proof. ∎

We call the matrix F the *multibody fundamental matrix* since it is a natural generalization of the fundamental matrix to the case of multiple moving objects. Since equation (3) clearly resembles the bilinear form of the epipolar constraint for a single rigid body motion, we will refer to both equations (1) and (3) as the *multibody epipolar constraint*.

3 Estimation of the number of motions and of the multibody fundamental matrix

Notice that, by definition, the multibody fundamental matrix F depends explicitly on the number of independent motions n. Therefore, even though the multibody epipolar constraint (3) is *linear* in F, we cannot use it to estimate F without knowing n in advance. We now derive a rank constraint on the image points from which one can estimate n, hence F. First, we rewrite the multibody epipolar constraint (3) as $(\nu_n(\boldsymbol{x}_2) \otimes \nu_n(\boldsymbol{x}_1))^T \boldsymbol{f} = 0$, where $\boldsymbol{f} \in \mathbb{R}^{M_n^2}$ is the stack of the columns of F and $\otimes$ represents the Kronecker product. Therefore, 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

$$A_n \boldsymbol{f} = 0, \tag{4}$$

where the j^{th} row of $A_n \in \mathbb{R}^{N \times M_n^2}$ equals $(\nu_n(\boldsymbol{x}_2^j) \otimes \nu_n(\boldsymbol{x}_1^j))^T$, for $j = 1, \ldots, N$. In order to determine $\boldsymbol{f}$ uniquely (up to a scale factor) from (4), we must have that

$$\text{rank}(A_n) = M_n^2 - 1. \tag{5}$$

This rank contraint on A_n provides an effective criterion to determine the number of independent motions n from the given image pairs, as stated by the following Theorem.

Theorem 1 (Number of independent motions)
Let $A_i \in \mathbb{R}^{N \times M_i^2}$ be the matrix in (4), but computed with the Veronese map ν_i of degree $1 \le i \le n$. Under the assumptions of Problem 1, if $N \ge M_n^2 - 1$ and at least 8 points correspond to each motion, then

$$\text{rank}(A_i) \begin{cases} > M_i^2 - 1, & \text{if } i < n, \\ = M_i^2 - 1, & \text{if } i = n, \\ < M_i^2 - 1, & \text{if } i > n. \end{cases} \tag{6}$$

Therefore, the number of independent motions n is given by

$$\boxed{n \doteq \min\{i : \text{rank}(A_i) = M_i^2 - 1\}.} \tag{7}$$

Proof: See [Vidal et al. 2002] for the proof. ∎

Therefore, we can use Theorem 1 to estimate the number of independent motions incrementally from equation (7). Given n, we can linearly solve for the multibody fundamental matrix F from (4). Notice that the minimum number of image pairs needed is $N \ge M_n^2 - 1$, which grows in the order of $O(n^4)$ for large n.

4 Multibody epipolar geometry

Recall that the (left) epipole $\boldsymbol{e}_i$ is defined as the (left) kernel of F_i, that is $\boldsymbol{e}_i^T F_i \doteq 0$. Hence we have that

$$\left(\boldsymbol{e}_i^T F_1 \boldsymbol{x}\right)\left(\boldsymbol{e}_i^T F_2 \boldsymbol{x}\right) \cdots \left(\boldsymbol{e}_i^T F_n \boldsymbol{x}\right) = \nu_n(\boldsymbol{e}_i)^T F \nu_n(\boldsymbol{x}) = 0, \tag{8}$$

for any $\boldsymbol{e}_i$, $i = 1, \ldots, n$. Therefore, because the M_n monomials in $\nu_n(\boldsymbol{x})$ are linearly independent, we have that $\nu_n(\boldsymbol{e}_i)^T F = 0$, *i.e.* the *embedded epipoles* $\{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n$ lie on the left null space of F, $\text{Null}(F)$. Thus, the rank of F is bounded depending on the number of *distinct* (pairwise linearly independent) epipoles as stated by Lemmas 2 and 3 below.

Lemma 2 (Null space of F when the epipoles are distinct)
If the epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$ are distinct, then the embedded epipoles $\{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n$ are linearly independent. Therefore the rank of F is bounded by

$$\boxed{\text{rank}(F) \le (M_n - n).} \tag{9}$$

Proof: See [Vidal et al. 2002] for the proof. ∎

Notice that, even though embedded epipoles lie in the left null space of F, we cannot estimate them directly from a basis of that null space, because linear combinations of $\{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n$ are not embedded epipoles in general. Furthermore, we do not know if the n-dimensional subspace spanned by $\{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n$ equals $\text{Null}(F)$, *i.e.* we do not know if $\text{rank}(F) = M_n - n$ (simulations confirm that this is true when all the epipoles are distinct). An additional complication, is that some of the epipoles may be repeated. In this case, one would expect that the dimension of $\text{Null}(F)$ decreases. However, this is *not* the case. The null space of F is actually enlarged by higher-order derivatives of the Veronese map as stated by the following Lemma.

Lemma 3 (Null space of F when one epipole is repeated)
Let $\boldsymbol{e}_1$ be repeated k times, i.e. $\boldsymbol{e}_1 = \cdots = \boldsymbol{e}_k$, and let the other $n - k$ epipoles be distinct. Then the M_{k-1} vectors of partial derivatives of $\nu_n(\boldsymbol{x})$ of order $k - 1$ evaluated at $\boldsymbol{e}_1$ lie in the left null space of F. Furthermore, $\nu_n(\boldsymbol{e}_1)$ is a linear combination of these M_{k-1} vectors. Therefore,

$$\boxed{\text{rank}(F) \le M_n - M_{k-1} - (n - k).} \tag{10}$$

Proof: See [Vidal et al. 2002] for the proof. ∎

Therefore, in order to uniquely recover the embedded epipoles $\nu_n(\boldsymbol{e}_i)$ from the left null space of F we will need to exploit the algebraic structure of the Veronese map. The following theorem guarantees the uniqueness of the recovery of the embedded epipoles from the intersection of Null(F) with the so-called Veronese surface.

Theorem 2 (Veronese null space of F) *The intersection of the left null space of F Null(F) with the Veronese surface $\nu_n(\mathbb{P}^2)$ is exactly*

$$\boxed{\text{Null}(F) \cap \nu_n(\mathbb{P}^2) = \{\nu_n(\boldsymbol{e}_i)\}_{i=1}^n.} \tag{11}$$

Proof: See [Vidal et al. 2002] for the proof. ∎

5 Multibody structure from motion

5.1 Estimation of the epipolar lines

Given a point $\boldsymbol{x}_1$ in the first image frame, the epipolar lines associated with it are defined as $\boldsymbol{l}_i \doteq F_i \boldsymbol{x}_1 \in \mathbb{R}^3$, $i = 1, \ldots, n$. Since

$$\nu_n(\boldsymbol{x}_2)^T F \nu_n(\boldsymbol{x}_1) = \prod_{i=1}^n \left(\boldsymbol{x}_2^T F_i \boldsymbol{x}_1\right) = \prod_{i=1}^n (\boldsymbol{x}_2^T \boldsymbol{l}_i) = \nu_n(\boldsymbol{x}_2)^T \tilde{\boldsymbol{l}},$$

we conclude that the *multibody epipolar line* $\tilde{\boldsymbol{l}} \doteq F \nu_n(\boldsymbol{x}_1) \in \mathbb{R}^{M_n}$ represents the coefficients of the homogeneous polynomial in $\boldsymbol{x}$

$$\boxed{g(\boldsymbol{x}) \doteq (\boldsymbol{x}^T \boldsymbol{l}_1)(\boldsymbol{x}^T \boldsymbol{l}_2) \cdots (\boldsymbol{x}^T \boldsymbol{l}_n) = \nu_n(\boldsymbol{x})^T \tilde{\boldsymbol{l}}.} \tag{12}$$

Therefore, in order to recover the epipolar lines $\{\boldsymbol{l}_i\}_{i=1}^n$ associated with $\boldsymbol{x}_1$ from the multibody epipolar line $\tilde{\boldsymbol{l}} = F \nu_n(\boldsymbol{x}_1)$, we need to factorize the homogeneous polynomial of degree n, $g(\boldsymbol{x})$, into the n homogeneous polynomials of degree one $\{\boldsymbol{x}^T \boldsymbol{l}_i\}_{i=1}^n$. We showed in [Vidal et al. 2002] that this polynomial factorization problem has a unique solution (up to scale of 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. 2002] for further details.

We can interpret the factorization of the multibody epipolar line $\tilde{\boldsymbol{l}} = F \nu_n(\boldsymbol{x}_1)$ as a generalization of the conventional "epipolar transfer" to multiple motions. In essence, the multibody fundamental matrix 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

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

5.2 Estimation of the individual epipoles

Given a set of epipolar lines, we now describe how to compute the epipoles associated with each one of the n motions. 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 $\boldsymbol{e}_i \in \mathbb{R}^3$ lying in the (left) null space of F_i, that is $\boldsymbol{e}_i$ satisfies $\boldsymbol{e}_i^T F_i = 0$. Now let $\boldsymbol{l} \in \mathbb{R}^3$ be an arbitrary epipolar line associated with some image point in the first frame. Then there exists an i such that $\boldsymbol{e}_i^T \boldsymbol{l} = 0$. Therefore, every epipolar line $\boldsymbol{l}$ has to satisfy the following polynomial constraint

$$\boxed{h(\boldsymbol{l}) \doteq (\boldsymbol{e}_1^T \boldsymbol{l})(\boldsymbol{e}_2^T \boldsymbol{l}) \cdots (\boldsymbol{e}_n^T \boldsymbol{l}) = \tilde{\boldsymbol{e}}^T \nu_n(\boldsymbol{l}) = 0,} \tag{13}$$

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.

Now, given a collection $\{\boldsymbol{l}^j\}_{j=1}^m$ of $m \geq M_n - 1$ epipolar lines (which can be computed from the multibody epipolar transfer process described before), the multibody epipole $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_n}$ satisfies the following system of linear equations

$$B_n \tilde{\boldsymbol{e}} = 0, \tag{14}$$

where the j^{th} row of $B_n \in \mathbb{R}^{N \times M_n}$ is given by $\nu_n(\boldsymbol{l}^j)^T$, for $j = 1, \ldots, m$. In order for equation (14) to have a unique solution (up to a scale factor), we will need to replace n by the number of distinct epipoles n_e, which can be computed from the following Lemma.

Lemma 4 (Number of distinct epipoles) *Let $\{\boldsymbol{l}^j\}_{j=1}^m$ be a collection of $m \geq M_n - 1$ epipolar lines, with at least 2 lines corresponding to each motion. Then*

$$\text{rank}(B_i) \begin{cases} > M_i - 1, & \text{if } i < n_e, \\ = M_i - 1, & \text{if } i = n_e, \\ < M_i - 1, & \text{if } i > n_e. \end{cases} \tag{15}$$

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

$$\boxed{n_e \doteq \min\{i : \text{rank}(B_i) = M_i - 1\}.} \tag{16}$$

Proof: See [Vidal et al. 2002] for the proof. ∎

Once the number of distinct epipoles, n_e, has been computed, the vector $\tilde{\boldsymbol{e}} \in M_{n_e}$ can be obtained from the linear system $B_{n_e} \tilde{\boldsymbol{e}} = 0$. Once $\tilde{\boldsymbol{e}}$ has been computed, the individual epipoles $\{\boldsymbol{e}_i\}_{i=1}^{n_e}$ can be computed by applying polynomial factorization to $h(\boldsymbol{l}) = \tilde{\boldsymbol{e}}^T \nu_n(\boldsymbol{l})$.

5.3 Estimation of the fundamental matrices

Given the epipolar lines and the epipoles, we now show how to recover each one of the individual fundamental matrices $\{F_i\}_{i=1}^n$. Notice that the only case in which both left and right epipoles are repeated is when the rotation axes of two (or more) motions are equal to each other and parallel to the common translation direction. Therefore, except for this degenerate case, we can assume that the epipoles (either left or right) are distinct, *i.e.* $n_e = n$, without loss of generality.

Let $F_i = [\boldsymbol{f}_i^1\ \boldsymbol{f}_i^2\ \boldsymbol{f}_i^3] \in \mathbb{R}^{3\times 3}$ be the fundamental matrix associated with motion i, with columns $\boldsymbol{f}_i^1, \boldsymbol{f}_i^2, \boldsymbol{f}_i^3 \in \mathbb{R}^3$. We know from Section 5.2 that, given $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T \in \mathbb{R}^3$, the vector $F \nu_n(\boldsymbol{x}_1) \in \mathbb{R}^{M_n}$ represents the coefficients of the following homogeneous polynomial in $\boldsymbol{x}$

$$g(\boldsymbol{x}) = \left(\boldsymbol{x}^T(\boldsymbol{f}_1^1 x_1 + \boldsymbol{f}_1^2 y_1 + \boldsymbol{f}_1^3 z_1)\right) \cdots \left(\boldsymbol{x}^T(\boldsymbol{f}_n^1 x_1 + \boldsymbol{f}_n^2 y_1 + \boldsymbol{f}_n^3 z_1)\right).$$

Thus, given the multibody fundamental matrix F, one can compute any linear combination of the columns of the fundamental matrix F_i up to a scale factor, *i.e.* we can get vectors $\boldsymbol{l}_i \in \mathbb{R}^3$ satisfying

$$\lambda_i \boldsymbol{l}_i \doteq (\boldsymbol{f}_i^1 x_1 + \boldsymbol{f}_i^2 y_1 + \boldsymbol{f}_i^3 z_1), \quad \lambda_i \in \mathbb{R},\ i = 1, \ldots, n.$$

These vectors are nothing but the epipolar lines associated with the multibody epipolar line $F \nu_n(\boldsymbol{x}_1)$, which can be computed using polynomial factorization as described in Section 5.2. Notice that, in particular, we can obtain the three columns of F_i up to a scale factor by choosing $\boldsymbol{x}_1 = [1, 0, 0]^T$, $\boldsymbol{x}_1 = [0, 1, 0]^T$ and $\boldsymbol{x}_1 = [0, 0, 1]^T$, respectively. However:

1. We do not know the fundamental matrix to which the recovered epipolar lines belong.
2. The recovered epipolar lines, hence the columns of each F_i, can be obtained up to a scale factor only. Hence, we do not know the relative scales between the columns of each F_i.

The first problem is easily solvable: if a recovered epipolar line $\boldsymbol{l} \in \mathbb{R}^3$ corresponds to a linear combination of columns of the fundamental matrix F_i, then it must be perpendicular to the previously computed epipolar line $\boldsymbol{e}_i$, *i.e.* we must have $\boldsymbol{e}_i^T \boldsymbol{l} = 0$. As for the second problem, for each i let $\boldsymbol{l}_i^j$ be the epipolar line associated with $\boldsymbol{x}_1^j$ that is perpendicular to $\boldsymbol{e}_i$, for $j = 1, \ldots, m$. Since the $\boldsymbol{x}_1^j$'s can be chosen arbitrarily, we choose the first three to be $\boldsymbol{x}_1^1 = [1,0,0]^T$, $\boldsymbol{x}_1^2 = [0,1,0]^T$ and $\boldsymbol{x}_1^3 = [0,0,1]^T$ to form a simple basis. Then for every $\boldsymbol{x}_1^j = [x_1^j, y_1^j, z_1^j]^T, j \geq 1$, there exist unknown scales $\lambda_i^j \in \mathbb{R}$ such that

$$\begin{aligned} \lambda_i^j \boldsymbol{l}_i^j &= \boldsymbol{f}_i^1 x_1^j \qquad + \boldsymbol{f}_i^2 y_1^j \qquad + \boldsymbol{f}_i^3 z_1^j \qquad j \geq 4, \\ &= (\lambda_i^1 \boldsymbol{l}_i^1) x_1^j + (\lambda_i^2 \boldsymbol{l}_i^2) y_1^j + (\lambda_i^3 \boldsymbol{l}_i^3) z_1^j, \quad j \geq 4. \end{aligned}$$

Cross-multiplying both sides by $\boldsymbol{l}_i^j$, we obtain

$$0 = \boldsymbol{l}_i^j \times \left((\lambda_i^1 \boldsymbol{l}_i^1) x_1^j + (\lambda_i^2 \boldsymbol{l}_i^2) y_1^j + (\lambda_i^3 \boldsymbol{l}_i^3) z_1^j \right), \quad j \geq 4 \tag{17}$$

where $\lambda_i^1, \lambda_i^2, \lambda_i^3$ are the only unknowns. Therefore, the fundamental matrices are given by

$$\boxed{F_i = [\boldsymbol{f}_i^1 \;\; \boldsymbol{f}_i^2 \;\; \boldsymbol{f}_i^3] = [\lambda_i^1 \boldsymbol{l}_i^1 \;\; \lambda_i^2 \boldsymbol{l}_i^2 \;\; \lambda_i^3 \boldsymbol{l}_i^3],} \tag{18}$$

where λ_i^1, λ_i^2 and λ_i^3 can be obtained as the solution to the linear system

$$\begin{bmatrix} x_1^4(\boldsymbol{l}_i^4 \times \boldsymbol{l}_i^1) & y_1^4(\boldsymbol{l}_i^4 \times \boldsymbol{l}_i^2) & z_1^4(\boldsymbol{l}_i^4 \times \boldsymbol{l}_i^3) \\ x_1^5(\boldsymbol{l}_i^5 \times \boldsymbol{l}_i^1) & y_1^5(\boldsymbol{l}_i^5 \times \boldsymbol{l}_i^2) & z_1^5(\boldsymbol{l}_i^5 \times \boldsymbol{l}_i^3) \\ \vdots & & \\ x_1^m(\boldsymbol{l}_i^m \times \boldsymbol{l}_i^1) & y_1^m(\boldsymbol{l}_i^m \times \boldsymbol{l}_i^2) & z_1^m(\boldsymbol{l}_i^m \times \boldsymbol{l}_i^3) \end{bmatrix} \begin{bmatrix} \lambda_i^1 \\ \lambda_i^2 \\ \lambda_i^3 \end{bmatrix} = 0. \tag{19}$$

We have given a constructive proof for the following statement:

Theorem 3 (Factorization of the multibody fundamental matrix)

Let $F \in \mathbb{R}^{M_n \times M_n}$ be the multibody fundamental matrix associated with fundamental matrices $\{F_i \in \mathbb{R}^{3\times 3}\}_{i=1}^n$. If the n epipoles are distinct, then the matrices $\{F_i\}_{i=1}^n$ can be uniquely determined up to a scale factor for each F_i.

5.4 3-D motion segmentation

3-D motion segmentation of the image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ can be easily done from either the epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$ and epipolar lines $\{\boldsymbol{l}^j\}_{j=1}^m$, or from the fundamental matrices $\{F_i\}_{i=1}^n$, as follows.

1. *Motion segmentation from the epipoles and epipolar lines*: Given $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, factorize $\tilde{\boldsymbol{l}} = F\nu_n(\boldsymbol{x}_1)$ into n epipolar lines. One of these lines, say $\boldsymbol{l}$, passes through $\boldsymbol{x}_2$, *i.e.* $\boldsymbol{l}^T \boldsymbol{x}_2 = 0$. The pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is assigned to the i^{th} motion if $\boldsymbol{l}^T \boldsymbol{e}_i = 0$.

2. *Motion segmentation from the fundamental matrices*: The image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is assigned to the i^{th} motion if $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0$.

Figure 1 illustrates how a particular image pair, say $(\boldsymbol{x}_1, \boldsymbol{x}_2)$, which belongs to the i^{th} motion, $i = 1, \ldots, n$, is successfully segmented.

5.5 Summary of multibody geometry

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

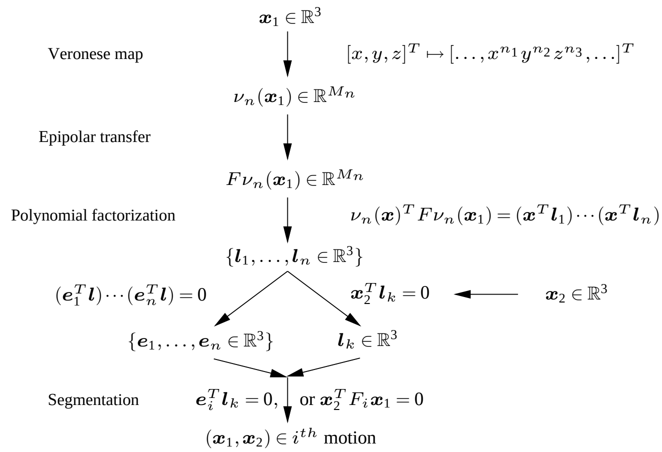


Figure 1: Transformation diagram associated with the segmentation of an image pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ in the presence of n motions.

5.6 Multibody structure from motion algorithm

We are now ready to present a complete algorithm for multibody motion estimation and segmentation from two perspective views.

Algorithm 1 (Multibody structure from motion algorithm).

Given a collection of image pairs $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$ of points undergoing n different motions, recover the number of motions n, the fundamental matrices $\{F_i\}_{i=1}^n$ and the segmentation as follows:

1. **Number of motions.** Compute the number of independent motions n from the rank constraint in (7), using the Veronese map of degree $i = 1, 2, \ldots, n$ applied to the image points $\{(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)\}_{j=1}^N$.

2. **Multibody fundamental matrix.** Compute the multibody fundamental matrix F as the solution of the linear system $A_n \boldsymbol{f} = 0$ in (4), using the Veronese map of degree n.

3. **Epipolar transfer.** Pick $N \geq M_n - 1$ vectors $\{\boldsymbol{x}_1^j \in \mathbb{R}^3\}_{j=1}^N$, with $\boldsymbol{x}_1^1 = [1,0,0]^T$, $\boldsymbol{x}_1^2 = [0,1,0]^T$ and $\boldsymbol{x}_1^3 = [0,0,1]^T$, and compute their corresponding epipolar lines $\{\boldsymbol{l}_k^j\}_{k=1,\ldots,n}^{j=1,\ldots,N}$ by applying polynomial factorization to $\{F\nu_n(\boldsymbol{x}_1^j) \in \mathbb{R}^{M_n}\}_{j=1}^N$.

4. **Multibody epipole.** Use the epipolar lines $\{\boldsymbol{l}_k^j\}_{k=1,\ldots,n}^{j=1,\ldots,N}$ to estimate the multibody epipole $\tilde{\boldsymbol{e}}$ as coefficients of the polynomial $h(\boldsymbol{l})$ in (13) by solving the system $B_n \tilde{\boldsymbol{e}} = 0$ in (14).

5. **Individual epipoles.** Compute the individual epipoles $\{\boldsymbol{e}_i\}_{i=1}^n$ by applying polynomial factorization to the multibody epipole $\tilde{\boldsymbol{e}} \in \mathbb{R}^{M_n}$.

6. **Individual fundamental matrices.** For each j, choose $k(i)$ such that $\boldsymbol{e}_i^T \boldsymbol{l}_{k(i)}^j = 0$, *i.e.* assign each epipolar line to its motion. Then use (18) and (19) to obtain each fundamental matrix F_i from the epipolar lines assigned to motion i.

7. **Features segmentation by motion.** Assign image pair $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$ to motion i if $\boldsymbol{e}_i^T \boldsymbol{l}_{k(i)}^j = 0$ or if $\boldsymbol{x}_2^T F_i \boldsymbol{x}_1 = 0$.

One of the main drawbacks of Algorithm 1 is that it needs a lot of image pairs in order to compute the multibody fundamental matrix. We discuss in [Vidal et al. 2002] a few variations to Algorithm 1 that significantly reduce the data requirements.

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	$\boldsymbol{x}_2^T F \boldsymbol{x}_1 = 0$	$\nu_n(\boldsymbol{x}_2)^T F \nu_n(\boldsymbol{x}_1) = 0$
Fundamental matrix	$F \in \mathbb{R}^{3\times 3}$	$F \in \mathbb{R}^{M_n \times M_n}$
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	$\boldsymbol{e}^T F = 0$	$\nu_n(\boldsymbol{e})^T F = 0$
Epipolar lines	$\boldsymbol{l} = F \boldsymbol{x}_1 \in \mathbb{R}^3$	$\tilde{\boldsymbol{l}} = F \nu_n(\boldsymbol{x}_1) \in \mathbb{R}^{M_n}$
Epipolar line & point	$\boldsymbol{x}_2^T \boldsymbol{l} = 0$	$\nu_n(\boldsymbol{x}_2)^T \tilde{\boldsymbol{l}} = 0$
Epipolar line & epipole	$\boldsymbol{e}^T \boldsymbol{l} = 0$	$\tilde{\boldsymbol{e}}^T \nu_n(\boldsymbol{l}) = 0$

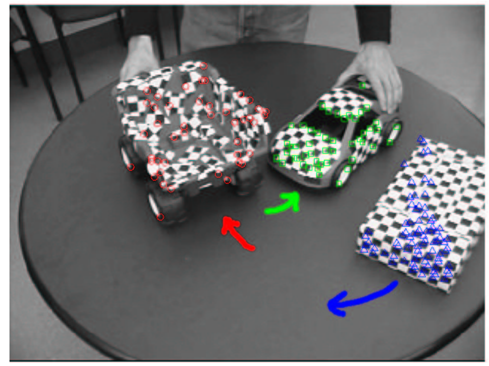
Table 1: Comparison between the geometry for two views of 1 rigid body motion and that for n rigid body motions.

6 Segmentation results

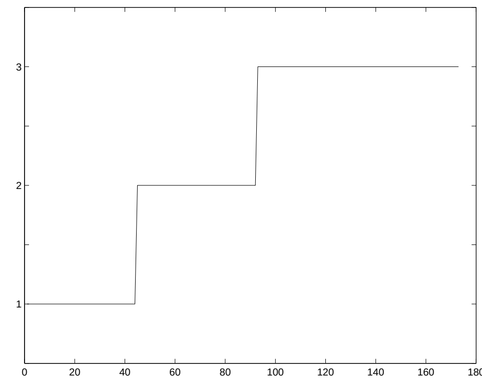
We tested the proposed approach by segmenting a real image sequence with $n = 3$ moving objects: a truck, a car and a box. Figure 2(a) shows the first frame of the sequence with the tracked features superimposed. We tracked a total of $N = 173$ point features: 44 for the truck, 48 for the car and 81 for the box. Figure 2(b) plots the segmentation of the image points from the obtained fundamental matrices. The segmentation has no mismatches.

7 Conclusions

We have proposed a geometric approach for the analysis of dynamic scenes based on the *multibody epipolar constraint* and its associated *multibody fundamental matrix*. The theory developed in this paper generalizes the well-known epipolar geometry to multiple moving objects and shows that it is possible to estimate multiple motions without prior segmentation using a purely geometric approach.



(a) First image frame



(b) Segmentation results

Figure 2: Segmentation of three independent motions. (a) First frame with tracked features superimposed: "○" for the truck, "□" for the car and "△" for the box. (b) Segmentation results: each image pair is assigned to the motion i that minimizes $(\boldsymbol{x}_2 F_i \boldsymbol{x}_1)^2$. The first 44 points correspond to the truck, the next 48 to the car, and the last 81 to the box. The correct segmentation is obtained.

References

AVIDAN, S., AND SHASHUA, A. 2000. Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. *IEEE Trans. on Pattern Analysis and Machine Intelligence 22*, 4, 348–357.

BLACK, M., AND ANANDAN, P. 1991. Robust dynamic motion estimation over time. In *International Conference on Computer Vision and Pattern Recognition*, 296–302.

COSTEIRA, J., AND KANADE, T. 1995. Multi-body factorization methods for motion analysis. In *IEEE International Conference on Computer Vision*, 1071–1076.

FENG, X., AND PERONA, P. 1998. Scene segmentation from 3D motion. In *International Conference on Computer Vision and Pattern Recognition*, 225–1.831.

FITZGIBBON, A., AND ZISSERMAN, A. 2000. Multibody structure and motion: 3D reconstruction of independently moving objects. In *European Conference on Computer Vision*.

HAN, M., AND KANADE, T. 2000. Reconstruction of a scene with multiple linearly moving objects. In *International Conference on Computer Vision and Pattern Recognition*, vol. 2, 542–549.

HAN, M., AND KANADE, T. 2001. Multiple motion scene reconstruction from uncalibrated views. In *IEEE International Conference on Computer Vision*, vol. 1, 163–170.

JEPSON, A., AND BLACK, M. 1993. Mixture models for optical flow computation. In *International Conference on Computer Vision and Pattern Recognition*, 760–761.

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

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

SHI, J., AND MALIK, J. 1998. Motion segmentation and tracking using normalized cuts. In *IEEE International Conference on Computer Vision*, 1154–1160.

SPOERRI, A., AND ULLMAN, S. 1987. The early detection of motion boundaries. In *IEEE International Conference on Computer Vision*, 209–1.818.

TORR, P. H. S. 1998. Geometric motion segmentation and model selection. *Phil. Trans. Royal Society of London A 356*, 1740, 1321–1340.

VIDAL, R., MA, Y., SOATTO, S., AND SASTRY, S. 2002. Two-view multibody structure from motion. Tech. Rep. UCB/ERL M02/11, UC Berkeley.

WEISS, Y. 1999. Segmentation using eigenvectors: a unifying view. In *IEEE International Conference on Computer Vision*, 975–982.

WOLF, L., AND SHASHUA, A. 2001. Two-body segmentation from two perspective views. In *International Conference on Computer Vision and Pattern Recognition*.