TWO-VIEW SEGMENTATION OF DYNAMIC SCENES FROM THE MULTIBODY FUNDAMENTAL MATRIX

by

René Vidal, Stefano Soatto and Shankar Sastry

Memorandum No. UCB/ERL M02/2

26 February 2002

TWO-VIEW SEGMENTATION OF DYNAMIC SCENES FROM THE MULTIBODY FUNDAMENTAL MATRIX

by

René Vidal, Stefano Soatto and Shankar Sastry

Memorandum No. UCB/ERL M02/2

26 February 2002

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Two-View Segmentation of Dynamic Scenes from the Multibody Fundamental Matrix *

René Vidal† Stefano Soatto‡ Shankar Sastry†

†Department of EECS, UC Berkeley
301 Cory Hall, Berkeley CA 94710
{rvidal,sastry}@eecs.berkeley.edu

‡Department of Computer Sciences, UCLA
Boelter Hall 4531, Los Angeles CA 90095
soatto@cs.ucla.edu

February 26, 2002

Abstract

We present a geometric approach to 3D motion segmentation of multiple moving objects seen in two perspective views. Our approach is based on the *multibody epipolar constraint* and its associated *multibody fundamental matrix*, which are generalizations of the epipolar constraint and of the fundamental matrix to multiple moving objects. We show how to linearly solve for the multibody fundamental matrix and for the number of independent motions. We also show how the epipoles of each independent motion can be computed from the nullspace of the multibody fundamental matrix, while epipolar lines and fundamental matrices can be recovered using tensor factorization. Motion segmentation is then obtained from either the epipoles or the fundamental matrices. We demonstrate the proposed approach to segment a real image sequence.

1 Introduction

Motion is one of the most important cues for segmenting an image sequence into different objects. Classical approaches to 2D motion segmentation try to separate the image flow into different regions either by looking for flow discontinuities [14], while imposing some regularity conditions [2], or by fitting a mixture of probabilistic models [9, 16]. The latter is usually done using an iterative process that alternates between segmentation and motion estimates using the Expectation-Maximization (EM) algorithm [5]. Alternative approaches are based on local features that incorporate spatial and temporal motion information. Similar features are grouped together using, for example, normalized cuts [13] or the eigenvectors of a similarity matrix [17].

3D motion segmentation and estimation based on 2D imagery is a more recent problem and various special cases have been analyzed using a geometric approach: multiple points moving linearly with constant speed [8, 12] or in a conic section [1], multiple moving objects seen by an orthographic camera [3, 10], self-calibration from multiple motions [7], or two-object segmentation from two perspective views [18]. Alternative probabilistic approaches to 3D motion segmentation are based on model selection techniques [15, 10] or combine normalized cuts with a mixture of probabilistic models [6].

In this paper we generalize the work of Wolf and Shashua [18] by considering a scene with multiple independently moving objects seen in two perspective views. We introduce the *multibody epipolar constraint* as a geometric relationship between camera motion and image points that is satisfied by all image points, regardless of the body to which they belong to. The multibody epipolar constraint defines the so-called *multibody fundamental matrix*, which is a generalization of the fundamental matrix to multiple bodies. We show how such a matrix can be computed linearly from image measurements, after embedding all image points in a higher-dimensional space. The number of independent motions can also be derived from

*Research supported by ONR grant N00014-00-1-0621 and ARO grant DAAD19-99-1-0139.

the corresponding estimation matrix. The epipoles of each independent motion can be computed from the nullspace of the multibody fundamental matrix, while epipolar lines and fundamental matrices can be recovered using tensor factorization. Motion segmentation is then obtained from either the epipoles or the fundamental matrices. We demonstrate the proposed approach to segment a real image sequence.

2 The Multibody Fundamental Matrix

We consider two frames of a scene containing n_o independent and generally moving objects. The motion of each object between the two frames is described by the fundamental matrix $F^k \in \mathbb{R}^{3\times 3}$ associated to object $k = 1, \dots, n_o$. The image of a point $P^i \in \mathbb{R}^3$ in frame j is denoted by $\mathbf{x}_j^i \in \mathcal{P}^2$, $i = 1, \dots n$, $j = 1, 2$. We drop the superscript when we refer to a generic image point, thus we use $(\mathbf{x}_1, \mathbf{x}_2)$ when we refer to an image pair that could correspond to any object and $(\mathbf{x}_{1k}, \mathbf{x}_{2k})$ when we refer to an image pair associated to object $k = 1, \dots, n_o$. We will use $\mathbf{x}$ to refer to an arbitrary point in $\mathcal{P}^2$.

Given a generic image pair $(\mathbf{x}_1, \mathbf{x}_2)$, there exists a k such that $\mathbf{x}_2^T F^k \mathbf{x}_1 = 0$. Therefore, the following constraint is satisfied between the motion of the objects and the image pair, regardless of the object to which the image pair belongs to:

$$L(\mathbf{x}_1, \mathbf{x}_2) = \prod_{k=1}^{n_o} \mathbf{x}_2^T F^k \mathbf{x}_1 = 0. \tag{1}$$

We call (1) the *multibody epipolar constraint*, since it is a generalization of the epipolar constraint [11] valid for $n_o = 1$. The case $n_o = 2$ can be found in [18].

Although we have written the multibody epipolar constraint as a function of $\mathbf{x}_1$ and $\mathbf{x}_2$, both $\mathbf{x}_1$ and $\mathbf{x}_2$ are actually repeated n_o times as follows:

$$L(\underbrace{\mathbf{x}_1, \dots, \mathbf{x}_1}_{n_o}, \underbrace{\mathbf{x}_2, \dots, \mathbf{x}_2}_{n_o}). \tag{2}$$

If the segmentation of the points were known, *i.e.*, if we knew that pair $(\mathbf{x}_{1k}, \mathbf{x}_{2k})$ belonged to object k, then we could write (2) as the multilinear expression $L(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_o}, \mathbf{x}_{21}, \dots, \mathbf{x}_{2n_o})$ to which we could associate the tensor product $F^1 \otimes F^2 \otimes \cdots \otimes F^{n_o}$. Since the segmentation of the points is unknown, we obtain the symmetric expression (2) instead, from which we cannot estimate the full tensor, but only its symmetric part:

$$\mathcal{F} = \sum_{\sigma \in S^{n_o}} F^{\sigma(1)} \otimes F^{\sigma(2)} \otimes \cdots \otimes F^{\sigma(n_o)}, \tag{3}$$

where S^{n_o} is the set of permutation of n_o letters. We call $\mathcal{F}$ the *fundamental tensor*, since it is just the symmetric tensor product of all the fundamental matrices.

It turns out that one can convert the tensor $\mathcal{F}$ into a matrix F that can be estimated linearly from image measurements. To see this, we realize that the multibody epipolar constraint is a homogeneous polynomial of degree n_o in the entries of $\mathbf{x}_1$ and $\mathbf{x}_2$. For example, if we let $\mathbf{x}_1 = (x, y, z)^T$, then (1) viewed as a function of $\mathbf{x}_1$ can be written as a linear combination of the following monomials $\{x^{n_o}, x^{n_o-1}y, x^{n_o-1}z, \dots, z^{n^o}\}$. We conclude that (1) can be seen as bi-linear expression after embedding the image points into a higher-dimensional space.

More explicitly, let $N_{n_o} = (n_o+1)(n_o+2)/2$ and let $\ell_{n_o} : \mathcal{P}^2 \to \mathcal{P}^{N_{n_o}-1}$ be the n_o^{th}-order lifting:

$$\ell_{n_o}(\mathbf{x}_1) = \begin{bmatrix} x^{n_o} & x^{n_o-1}y & x^{n_o-1}z & \cdots & z^{n^o} \end{bmatrix}^T. \tag{4}$$

The multibody epipolar constraint (1) can be re-written as:

$$\ell_{n_o}(\mathbf{x}_2)^T F \ell_{n_o}(\mathbf{x}_1) = 0 \tag{5}$$

where $F \in \mathbb{R}^{N_{n_o} \times N_{n_o}}$ is a matrix representation of the fundamental tensor $\mathcal{F}$. We call F the *multibody fundamental matrix* since it is a natural generalization of the fundamental matrix to the case of multiple moving objects.

3 Estimation of the Multibody Fundamental Matrix

In this section, we present a linear algorithm for estimating the multibody fundamental matrix F and the number of independent motions n_o from a set of n image pairs. Even though it is out of the scope of the paper to provide a robust algorithm for estimating F in the presence of noise, in Theorem 1 we provide a statistically optimal function from which F can be estimated using non-linear optimization techniques.

3.1 Linear estimation of F

Since (5) is linear on the entries of F, the multibody fundamental matrix can be computed up to scale as the solution of the linear system:

$$\chi_{n_o}(\mathbf{x}_1,\mathbf{x}_2)f = \begin{bmatrix} \ell_{n_o}(\mathbf{x}_2^1)^T * \ell_{n_o}(\mathbf{x}_1^1)^T \\ \vdots \\ \ell_{n_o}(\mathbf{x}_2^n)^T * \ell_{n_o}(\mathbf{x}_1^n)^T \end{bmatrix} f = 0, \tag{6}$$

where $A * B$ is the Kronecker product of A and B and $f \in \mathbb{R}^{N_{n_o}^2}$ is the stack of the columns of F. It is clear that the minimum number of image points needed to linearly recover F from (6) is

$$n \geq N_{n_o}^2 - 1 = \left[\frac{(n_o+1)(n_o+2)}{2}\right]^2 - 1. \tag{7}$$

Notice that when there is only one object, we have $N_1 = 3$ and the associated fundamental matrix can be computed from 8 points.

3.2 Number of independent motions

From the previous analysis one may think that in order to determine the multibody fundamental matrix, one needs to know the number of motions beforehand. It turns out that for 3D points in general configuration and for general and independent motions, one can determine n_o from the rank of the matrices χ_k as shown by the following lemma:

Lemma 1 *Assume that the collection of image points corresponds to 3D points in general configuration and that the motions of the objects are general and independent. Let $\chi_k(\mathbf{x}_1,\mathbf{x}_2)$ be the matrix defined in (6), but computed using the k^{th}-order lifting of $\mathbf{x}_1$ and $\mathbf{x}_2$. Then:*

$$rank(\chi_k) = \begin{cases} N_k^2 & \text{if } k < n_o \\ N_k^2 - 1 & \text{if } k = n_o \\ r < N_k^2 - 1 & \text{if } k > n_o. \end{cases} \tag{8}$$

Therefore the number of independently moving objects is given by:

$$n_o = \min\{k : rank(\chi_k) = N_k^2 - 1\}. \tag{9}$$

3.2.1 Optimal estimation of F

In the presence of noise, the solution of (6) may be a biased estimator of the true multibody fundamental matrix. As in the single body case, it is possible to obtain an unbiased estimator by minimizing a non-linear objective function. The following theorem gives the expression for the optimal function for estimating F.

Theorem 1 *Let $(\mathbf{x}_1,\mathbf{x}_2) = (\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2) + (n_1,n_2)$ be an image pair corrupted with i.i.d zero mean Gaussian noise. Let $D\ell_{n_o}(\mathbf{x}) \in \mathbb{R}^{N_o \times 3}$ be the Jacobian of the lifting ℓ_{n_o} and let $e_3 = (0,0,1)^T$. The optimal function for estimating F is given by:*

$$J(F,\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2) = \sum_{i=1}^{n} \frac{\left(\ell_{n_o}(\tilde{\mathbf{x}}_1)^T F^T D\ell_{n_o}(\tilde{\mathbf{x}}_2)\mathbf{x}_2 + \ell_{n_o}(\tilde{\mathbf{x}}_2)^T F D\ell_{n_o}(\tilde{\mathbf{x}}_1)\mathbf{x}_1\right)^2}{\|[e_3]_\times D\ell_{n_o}^T(\tilde{\mathbf{x}}_1)F^T\ell_{n_o}(\tilde{\mathbf{x}}_2)\|^2 + \|[e_3]_\times D\ell_{n_o}^T(\tilde{\mathbf{x}}_2)F\ell_{n_o}(\tilde{\mathbf{x}}_1)\|^2}. \tag{10}$$

An approximated function obtained after neglecting higher-order terms is:

$$J(F)=\sum_{i=1}^{n}\frac{(\ell_{n_o}(\mathbf{x}_2)^T F\ell_{n_o}(\mathbf{x}_1))^2}{\|[e_3]_\times D\ell_{n_o}^T(\mathbf{x}_1)F^T\ell_{n_o}(\mathbf{x}_2)\|^2+\|[e_3]_\times D\ell_{n_o}^T(\mathbf{x}_2)F\ell_{n_o}(\mathbf{x}_1)\|^2}. \tag{11}$$

For $n_o=1$ we have $\ell_1(\mathbf{x})=\mathbf{x}$ and $D\ell_1(\mathbf{x})=I_3$, thus we obtain the standard error function:

$$J(F)=\sum_{i=1}^{n}\frac{(\mathbf{x}_2^T F\mathbf{x}_1)^2}{\|[e_3]_\times F^T\mathbf{x}_2\|^2+\|[e_3]_\times F\mathbf{x}_1\|^2}. \tag{12}$$

4 Motion Segmentation from the Multibody Fundamental Matrix

Given the multibody fundamental matrix, we are now interested in recovering the segmentation of the image points and the motion parameters. We show how this can be done from the epipoles of each fundamental matrix and the epipolar lines of each image point.

4.1 Estimation of the epipolar lines

Given an image in the first frame $\mathbf{x}_1$, let $\mathbf{l}_2^k=F^k\mathbf{x}_1, k=1,\dots n_o$. Then there exists a k such that $\mathbf{l}_2^k$ corresponds to the epipolar line passing through $\mathbf{x}_2$, *i.e.*, there exists a k such that $\mathbf{x}_2^T\mathbf{l}_2^k=0$. Since

$$L(\mathbf{x},\mathbf{x}_1)=\prod_{k=1}^{n_o}\mathbf{x}^T F^k\mathbf{x}_1=\prod_{k=1}^{n_o}\mathbf{x}^T\mathbf{l}_2^k=\ell_{n_o}(\mathbf{x})^T F\ell_{n_o}(\mathbf{x}_1), \tag{13}$$

we conclude that $F\ell_{n_o}(\mathbf{x}_1)$ represents the coordinates of the symmetric tensor

$$\sum_{\sigma\in S_{n_o}}\mathbf{l}_2^{\sigma(1)}\otimes\mathbf{l}_2^{\sigma(2)}\otimes\cdots\otimes\mathbf{l}_2^{\sigma(n_o)}. \tag{14}$$

Similarly, $F^T\ell_{n_o}(\mathbf{x}_2)$ represents the coordinates of the symmetric tensor

$$\sum_{\sigma\in S_{n_o}}\mathbf{l}_1^{\sigma(1)}\otimes\mathbf{l}_1^{\sigma(2)}\otimes\cdots\otimes\mathbf{l}_1^{\sigma(n_o)} \tag{15}$$

where $\mathbf{l}_1^k=F^{kT}\mathbf{x}_2$ and there exists a k such that $\mathbf{x}_1^T\mathbf{l}_1^k=0$.

We conclude that finding the epipolar line $\mathbf{l}_2$ associated to $\mathbf{x}_2$ is equivalent to factorizing the tensor $F\ell_{n_o}(\mathbf{x}_1)$ into its n_o components $\mathbf{l}_2^1,\dots,\mathbf{l}_2^{n_o}$ and then checking which one of them satisfies $\mathbf{x}_2^T\mathbf{l}_2^k=0$. Similarly, one can find the epipolar line $\mathbf{l}_1$.

In order to solve for the epipolar lines, we need to be able to factorize a symmetric tensor into its n_o components. For $n_o=2$ this can be done using the eigenvalue decomposition as follows:

1. Convert $\alpha=F\ell_{n_o}(\mathbf{x}_1)\in\mathbb{R}^6$ to the rank-2 symmetric matrix [18]:

$$A=\begin{bmatrix}\alpha_1 & \alpha_2/2 & \alpha_3/2\\ \alpha_2/2 & \alpha_4 & \alpha_5/2\\ \alpha_3/2 & \alpha_5/2 & \alpha_6\end{bmatrix}\in\mathbb{R}^{3\times3}.$$

2. Find the eigenvalue decomposition of $A=U\Lambda U^T$, $\Lambda=diag(\lambda_1,\lambda_2,0)$. Then:

$$\begin{bmatrix}\mathbf{l}_2^1 & \mathbf{l}_2^2\end{bmatrix}=U\begin{bmatrix}\sqrt{|\lambda_1|} & sgn(\lambda_1)\sqrt{|\lambda_1|}\\ \sqrt{|\lambda_2|} & sgn(\lambda_2)\sqrt{|\lambda_2|}\end{bmatrix}\frac{1}{\sqrt{2}}.$$

The key for solving the problem for $n_o = 2$ is the existence of the SVD, $A = U\Sigma V^T$, for matrices (second-order tensors) that converts A into a diagonal matrix. Even though there are multilinear versions of the singular value decomposition for $n_o > 2$ [4], it is not possible to ensure that Σ will be diagonal in general.

We are interested in a special case in which the tensor to be factorized is symmetric. Therefore, we will exploit the structure of the problem to propose an iterative scheme that solves the factorization problem. The main idea is to reduce the problem to the case $n_o = 2$, which we know how to solve, by fixing the other factors. We give the details of the case $n_o = 3$ below and briefly outline the case $n_o > 3$.

Let $\alpha \in \mathbb{R}^{10}$ be the coordinates of the symmetric part of the third-order tensor $\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$, where $\mathbf{x} = (x_1, x_2, x_3)^T$, $\mathbf{y} = (y_1, y_2, y_3)^T$ and $\mathbf{z} = (z_1, z_2, z_3)^T$. We have:

$$\alpha = \begin{bmatrix} z_1 & 0 & 0 & 0 & 0 & 0 \\ z_2 & z_1 & 0 & 0 & 0 & 0 \\ z_3 & 0 & z_1 & 0 & 0 & 0 \\ 0 & z_2 & 0 & z_1 & 0 & 0 \\ 0 & z_3 & z_2 & 0 & z_1 & 0 \\ 0 & 0 & z_3 & 0 & 0 & z_1 \\ 0 & 0 & 0 & z_2 & 0 & 0 \\ 0 & 0 & 0 & z_3 & z_2 & 0 \\ 0 & 0 & 0 & 0 & z_3 & z_2 \\ 0 & 0 & 0 & 0 & 0 & z_3 \end{bmatrix} \begin{bmatrix} x_1y_1 \\ x_1y_2 + x_2y_1 \\ x_1y_3 + x_3y_1 \\ x_2y_2 \\ x_2y_3 + x_3y_2 \\ x_3y_3 \end{bmatrix} = G(\mathbf{z})\,(\mathbf{x}\otimes\mathbf{y} + \mathbf{y}\otimes\mathbf{x}) \tag{16}$$

$$= \begin{bmatrix} x_1y_1 & 0 & 0 \\ (x_1y_2 + x_2y_1) & x_1y_1 & 0 \\ (x_1y_3 + x_3y_1) & 0 & x_1y_1 \\ x_2y_2 & (x_1y_2 + x_2y_1) & 0 \\ (x_2y_3 + x_3y_2) & (x_1y_3 + x_3y_1) & (x_1y_2 + x_2y_1) \\ x_3y_3 & 0 & (x_1y_3 + x_3y_1) \\ 0 & x_2y_2 & 0 \\ 0 & (x_2y_3 + x_3y_2) & x_2y_2 \\ 0 & x_3y_3 & (x_2y_3 + x_3y_2) \\ 0 & 0 & x_3y_3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = H(\mathbf{x},\mathbf{y})\mathbf{z}. \tag{17}$$

These equations are homogeneous in $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, so in order to solve the problem we can further impose the constraint $\|\mathbf{x}\| = \|\mathbf{y}\| = \|\mathbf{z}\| = 1$. Given the form of the equations, this can be done regardless of the norm of α. We solve the equations as follows: Given $\mathbf{z}$ we solve linearly for $\mathbf{x}\otimes\mathbf{y} = \left(G^T(\mathbf{z})G(\mathbf{z})\right)^{-1}G^T(\mathbf{z})\alpha$ and given $\mathbf{x}\otimes\mathbf{y}+\mathbf{y}\otimes\mathbf{x}$ we solve linearly for $\mathbf{z} = \left(H^T(\mathbf{x},\mathbf{y})H(\mathbf{x},\mathbf{y})\right)^{-1}H^T(\mathbf{x},\mathbf{y})\alpha$. We have observed experimentally that this iterative algorithm converges to the correct solution with random initialization for $\mathbf{z}$. In most of the cases the number of iterations required for convergence is between 5 and 30, although there are cases for which convergence is slow. After the algorithm has converged, $\mathbf{x}$ and $\mathbf{y}$ are obtained from the factorization of $\mathbf{x}\otimes\mathbf{y}+\mathbf{y}\otimes\mathbf{x}$ using the algorithm for $n_o = 2$.

For $n_o > 3$ we do as follows. If n_o is even, then we factorize the tensor in $n_o/2$ second-order tensors. If n_o is odd, then we factorize it in $(n_o - 1)/2$ second-order tensors and 1 first-order tensor.

4.2 Estimation of the epipoles

The right and left epipoles T_1^k and T_2^k associated with motion k are the right and left nullspace of F^k, respectively. In order to solve for T_1^k and T_2^k we will look for a set of linear constraints on the lifted epipoles $\ell_{n_o}(T_1^k)$ and $\ell_{n_o}(T_2^k)$. We will basically show that the lifted epipoles lie on the intersection of the nullspace of the epipolar lines and the nullspace of the multibody fundamental matrix. For $n_o = 2$ this will be enough to uniquely recover the epipoles. For $n_o \geq 3$ additional polynomial constraints will be needed. For example, the case $n_o = 3$ can be solved from a second-order polynomial in one variable.

Let $(\mathbf{x}_1, \mathbf{x}_2)$ be an image pair and let $\mathbf{l}_1 = F^{kT}\mathbf{x}_2$ and $\mathbf{l}_2 = F^k\mathbf{x}_1$ (for some unknown k) be the associated epipolar lines which can be computed using the algorithm in Section 4.1. Since $\mathbf{l}_1^T T_1^k = 0$ and $\mathbf{l}_2 T_2^k = 0$, we obtain:

$$(\mathbf{l}_1^T T_1^k)^{n_o} = \widetilde{\ell}_{n_o}(\mathbf{l}_1)^T \ell_{n_o}(T_1^k) = 0 \quad \text{and} \quad (\mathbf{l}_2^T T_2^k)^{n_o} = \widetilde{\ell}_{n_o}(\mathbf{l}_2)^T \ell_{n_o}(T_2^k) = 0, \tag{18}$$

where $\widetilde{\ell}_{n_o}$ is the same as ℓ_{n_o}, except for some coefficients resulting from raising a trinomial to the n_o^{th} power. For example, for $\mathbf{x} = (x, y, z)^T$, we have

$$\begin{aligned}
\widetilde{\ell}_2(\mathbf{x}) &= (x^2, 2xy, 2xz, y^2, 2yz, z^2)^T, \\
\widetilde{\ell}_3(\mathbf{x}) &= (x^3, 3x^2y, 3x^2z, 3xy^2, 6xyz, 3xz^2, y^3, 3y^2z, 3yz^2, z^3)^T.
\end{aligned}$$

Equation (18) gives one linear constraint on the lifted epipoles. Additional constraints can be derived from Lemma 2, which generalizes the case $n_o = 2$ [18]:

Lemma 2 (Relationship between the nullspace of F and that of F^k) *Let F be the multibody fundamental matrix associated to the fundamental matrices $F^1, \ldots, F^{n_o}$. Let T_1^k and T_2^k be the left and right epipoles associated to F^k, i.e., $F^k T_1^k = 0$ and $T_2^{kT} F^k = 0$. Then $F\,\ell_{n_o}(T_1^k) = 0$ and $\ell_{n_o}(T_2^k)^T F = 0$ for all $k = 1, \ldots n_o$.*

Proof: For any $\mathbf{x} \in \mathcal{P}^2$ we have $\ell_{n_o}(\mathbf{x})^T F\,\ell_{n_o}(T_1^k) = L(\mathbf{x}, T_1^k) = \prod_{k'=1}^{n_o} \mathbf{x}^T F^{k'} T_1^{k'}$. Since $F^k T_1^k = 0$ and the span of $\ell_{n_o}(\mathbb{R}^3)$ equals $\mathbb{R}^{N_{n_o}}$, we have $F\,\ell_{n_o}(T_1^k) = 0$. Similarly $\ell_{n_o}(T_2^k)^T F = 0$. ∎

In general we have $\text{rank}(F) = N_{n_o} - n_o$ when the lifted epipoles are linearly independent. Let $V = \begin{bmatrix} v_1 & \cdots & v_{n_o} \end{bmatrix} \in \mathbb{R}^{N_{n_o} \times n_o}$ be a basis for the null space of F. Then for any epipole T_1 we have $\ell_{n_o}(T_1) = V\lambda$, for some vector of coefficients $\lambda \in \mathbb{R}^{n_o}$. We would like to find λ such that $V\lambda \in \ell_{n_o}(\mathbb{R}^3)$. Since $\ell_{n_o}(\mathbb{R}^3)$ is a three-dimensional manifold in $\mathbb{R}^{N_{n_o}}$, there are $N_{n_o} - 3$ independent constraints on λ. Given the form of ℓ_{n_o}, such constraints are actually homogeneous polynomials of degree n_o on the $n_o - 1$ independent entries of λ[1].

For $n_o = 2$ the nullspace of $\begin{bmatrix} F \\ \widetilde{\ell}_{n_o}(\mathbf{l}_1)^T \end{bmatrix} \in \mathbb{R}^{7\times 6}$ is one-dimensional. Therefore, we obtain a unique solution for $\ell(T_1^1)$, hence T_1^1. Notice that the previous computation was done for an arbitrary image pair $(\mathbf{x}_1, \mathbf{x}_2)$. Thus one can compute the epipole T_1 associated to each image pair. Image pairs with the same epipole belong to the same moving object, hence segmentation of the image points is trivially obtained (See Section 5 for details).

For $n_o = 3$, let $W = [w_1, w_2] \in \mathbb{R}^{10\times 2}$ be a basis for the nullspace of $\begin{bmatrix} F \\ \widetilde{\ell}_{n_o}(\mathbf{l}_1)^T \end{bmatrix} \in \mathbb{R}^{11\times 10}$. We have $\ell_{n_o}(T_1) = W \begin{bmatrix} \lambda \\ 1 \end{bmatrix}$ for some $\lambda \in \mathbb{R}$ satisfies the following set of second-order polynomials:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \end{bmatrix} = \begin{bmatrix} x^3 \\ x^2y \\ x^2z \\ xy^2 \\ xyz \\ xz^2 \\ y^3 \\ y^2z \\ yz^2 \\ z^3 \end{bmatrix} \Longrightarrow \begin{cases}
\alpha_2\alpha_4 = \alpha_1\alpha_7 \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_2e_4^T - e_1e_7^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0 \\
\alpha_3\alpha_6 = \alpha_1\alpha_{10} \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_3e_6^T - e_1e_{10}^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0 \\
\alpha_8\alpha_9 = \alpha_7\alpha_{10} \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_8e_9^T - e_7e_{10}^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0 \\
\alpha_2\alpha_3 = \alpha_1\alpha_5 \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_2e_3^T - e_1e_5^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0 \\
\alpha_4\alpha_8 = \alpha_5\alpha_7 \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_4e_8^T - e_5e_7^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0 \\
\alpha_6\alpha_9 = \alpha_5\alpha_{10} \Leftrightarrow \begin{bmatrix} \lambda & 1 \end{bmatrix} W^T(e_6e_9^T - e_5e_{10}^T)W \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = 0
\end{cases}. \tag{19}$$

where $e_i, i = 1, \ldots, 10$ is the standard basis for $\mathbb{R}^{10}$. There will be one common root to these second-order polynomials, from which one obtains the epipole T_1^1. As before, one can find the epipole associated to each image pair and segment image pairs depending on their associated epipoles (See Section 5 for details).

[1] One entry of λ can always be eliminated because $\ell_{n_o}(\mathbb{R}^3)$ is a homogeneous space.

4.3 Estimation of the Fundamental Matrices

Let $e_i, i = 1...3$ be the standard basis for $\mathbb{R}^3$ and let f_i^k be the i^{th} column of the k^{th} fundamental matrix, for $i = 1..3$ and $k = 1...n_o$. For all $\mathbf{x} \in \mathcal{P}^2$ we have:

$$L(\mathbf{x}, e_i) = \prod_{k=1}^{n_o} \mathbf{x}^T F^k e_i = \prod_{k=1}^{n_o} \mathbf{x}^T f_i^k = \ell_{n_o}(\mathbf{x})^T F \ell_{n_o}(e_i). \tag{20}$$

Therefore $F\ell_{n_o}(e_i)$ represents the coordinates of the symmetric tensor

$$\sum_{\sigma \in S_{n_o}} f_i^{\sigma(1)} \otimes f_i^{\sigma(2)} \otimes \cdots \otimes f_i^{\sigma(n_o)}. \tag{21}$$

We conclude that in order to find the columns of each of the fundamental matrices, we need to factorize this tensor into its n_o components. This can be done with the techniques of Section 4.1. Since the factorization problem is symmetric in each of the factors of the tensor, we do not know which column corresponds to which fundamental matrix. Also, the factorization process outputs these columns with unit norm, hence the relative scale between two columns of each fundamental matrix is also lost. To solve this problem, we use the algorithm in [18] which can be summarized as follows

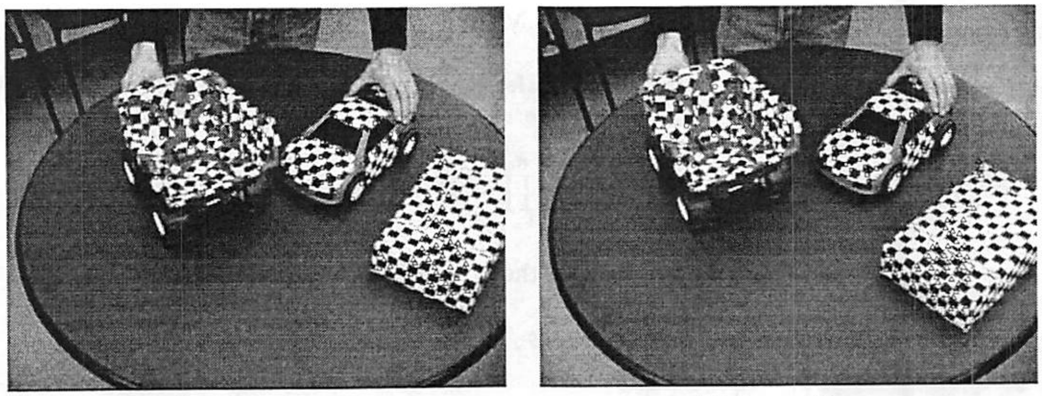
- Find all possible fundamental matrices from all column combinations.
- Find one row for each fundamental matrix from the factorization of $F^T\ell_{n_o}(e_i)$ and use them to find the unknown scales between the columns.
- Build one multibody fundamental matrix from each possibility and compare it to the correct one to obtain the correct fundamental matrices for each motion.

5 Experimental Results

Figure 1 shows two frames of a motion sequence containing two cars and a box with the tracked features superimposed. We track a total of $n = 173$ point features: 44 for the first car, 48 for the second and 81 for the box. The segmentation of the image points was obtained as follows:

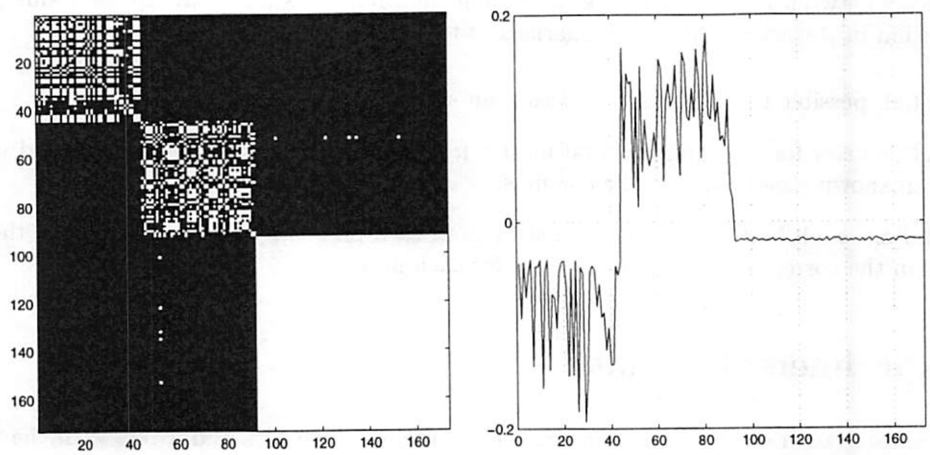
1. Estimate n_o from (9) and F from (6) using all the image points. We obtain $n_o = 3$.
2. For each image point $\mathbf{x}_1^i$, $i = 1, \ldots, n$:
   (a) Factorize the tensor $F\ell_{n_o}(\mathbf{x}_1^i)$ into $n_o = 3$ candidate epipolar lines.
   (b) Estimate the unique epipolar line $\mathbf{l}_2^i$ as the factor of the tensor that minimizes $(\mathbf{l}_2^{iT}\mathbf{x}_2^i)^2$.
   (c) Estimate the epipole T^i as $\ell_{n_o}^{-1}(W^i[\lambda^i\ 1]^T)$, where W^{iT} is a basis for the left nullspace of $[F\ \tilde{\ell}_{n_o}(\mathbf{l}_2^i)]$ and $\lambda^i \in \mathbb{R}$ is the solution of the polynomials in (19).
3. Form the segmentation matrix $\mathcal{S} = \mathcal{T}^T\mathcal{T} \in \mathbb{R}^{n\times n}$, where $\mathcal{T} = [T^1\, T^2 \cdots T^n] \in \mathbb{R}^{3\times n}$. We should have $|\mathcal{S}_{ij}| = 1$ if points i and j belong to the same object. Alternatively, one can threshold the second eigenvector of $\mathcal{S}$ to obtain the segmentation, as shown in [17].

Figure 2(a) plots the segmentation matrix $\mathcal{S}$ and Figure 2(b) plots its second eigenvector. Notice the correct block diagonal structure of the $\mathcal{S}$, except for some outliers. We obtained the correct segmentation by thresholding the second eigenvector.



(a) First frame (b) Second frame

Figure 1: A motion sequence with two cars and a box. Tracked features are marked with a 'o' for the first car, a '□' for the second car and an '△' for the box.



(a) Segmentation matrix $\mathcal{S}$ (b) Second eigenvector of $\mathcal{S}$

Figure 2: Motion segmentation results. The correct segmentation is obtained from the second eigenvector of $\mathcal{S}$. The first 44 points correspond to the first car, the next 48 to the second, and the last 81 to the box.

6 Conclusions

We have introduced the *multibody fundamental matrix* and showed that it is a geometric entity which is at the core of 3D motion segmentation. We showed how to linearly solve for the multibody fundamental matrix from the multibody epipolar constraint. We also showed how to solve for epipolar lines, epipoles, fundamental matrices and motion segmentation using a purely geometric approach. Experimental results on a real image sequence showed the applicability of the proposed algorithms to multibody motion segmentation from two perspective views.

References

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

[2] M. Black and P. Anandan. Robust dynamic motion estimation over time. In *International Conference on Computer Vision and Pattern Recognition*, pages 296–302, 1991.

[3] J. Costeira and T. Kanade. Multi-body factorization methods for motion analysis. In *IEEE International Conference on Computer Vision*, pages 1071–1076, 1995.

[4] L. de Lathauwer, B. de Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis*, 21(4):1253–1278, 2000.

[5] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(B):1–38, 1977.

[6] X. Feng and P. Perona. Scene segmentation from 3D motion. In *International Conference on Computer Vision and Pattern Recognition*, pages 225–231, 1998.

[7] A. Fitzgibbon and A. Zisserman. Multibody structure and motion: 3D reconstruction of independently moving objects. In *European Conference on Computer Vision*, 2000.

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

[9] A. Jepson and M. Black. Mixture models for optical flow computation. In *International Conference on Computer Vision and Pattern Recognition*, pages 760–761, 1993.

[10] K. Kanatani. Motion segmentation by subspace separation and model selection. In *IEEE International Conference on Computer Vision*, volume 2, pages 586–591, 2001.

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

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

[13] J. Shi and J. Malik. Motion segmentation and tracking using normalized cuts. In *IEEE International Conference on Computer Vision*, pages 1154–1160, 1998.

[14] A. Spoerri and S. Ullman. The early detection of motion boundaries. In *IEEE International Conference on Computer Vision*, pages 209–218, 1987.

[15] P. H. S. Torr. Geometric motion segmentation and model selection. *Phil. Trans. Royal Society of London A*, 356(1740):1321–1340, 1998.

[16] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *International Conference on Computer Vision and Pattern Recognition*, pages 520–526, 1997.

[17] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *IEEE International Conference on Computer Vision*, pages 975–982, 1999.

[18] L. Wolf and A. Shashua. Two-body segmentation from two perspective views. In *International Conference on Computer Vision and Pattern Recognition*, 2001.