Two-body Segmentation from Two Perspective Views

Lior Wolf and Amnon Shashua*
School of Computer Science and Engineering,
The Hebrew University,
Jerusalem 91904, Israel
e-mail: {shashua,lwolf}@cs.huji.ac.il

Abstract

We consider a scene containing two independently and generally moving objects, viewed by two general perspective views. Using matching points arising from both objects simultaneously we derive a geometrical constraint, applicable to points from both objects, we call the segmentation matrix. We then use this constraint in order to recover the fundamental matrices associated with each object, or simply to segment the scene into the two objects. Moreover, when the two bodies move in pure translation relative to each other we can both segment the scene and recover the affine calibration (homography at infinity) of the camera geometry. Unlike algorithms suggested in the past we need only two images, we work with general projective cameras (rather than affine or orthographic) and with general body motion, and no prior information beyond point matches is required.

1 Introduction

We consider the problem of segmenting a collection of matching points across a pair of images arising from two 3D bodies moving relatively to each other in a general manner — which we refer to as the two-body motion segmentation problem. All we are given as input are matching points from which we would like to decide which pair of matching points comes from which object, recover the underlying fundamental matrices (each body is associated with one), and in case the relative motion between the bodies is pure translation to recover the affine calibration of the camera geometry. We do so using geometric constraints alone, i.e., without the need for statistical model fitting in the presence of outliers techniques like Least-Median Squares, Ransac, or M-estimation. In other words, *we look for multi-linear constraints which apply to the matching points as a whole regardless from which object they arise from* — from these constraints we can recover all aspects of the individual bodies' motion (epipoles, fundamental matrix, segmentation).

The closest work to ours is the factorization-based motion segmentation of [2] applicable to the affine (parallel projection) camera model. There it was shown that the measurement matrix of all points across a sequence of images lies in a linear subspace whose dimension is determined by the number of independent bodies and where the motion of each body lies in a separate subspace — the rearrangement of the data can be done in order to separate between the subspaces (see also [6]). In our work we assume a full projective model (but also address the affine model) and assume only two views (whereas the factorization would require at least 4 views).

Our analytical approach follow the path of a growing body of work on *dynamic* scenes [2, 1, 10, 3, 7, 11, 4, 12], i.e., 3D scenes which contain multiply moving points or collections of points (bodies) seen under multiple views. Among these, the most relevant are [12] who propose a segmentation tensor applicable to two bodies, but which move in pure translation to each other (instead of general motion), the technique requires three views (instead of two) and 4 points are assumed to be pre-segmented (whereas here we require no information beyond the matching points themselves).

2 The 2-body Multi-linear Constraints

We will model the image plane as the projective plane $\mathcal{P}^2$. Points and lines are represented as triplets of numbers (not all zero) which are defined up to scale. A point p lies on a line l if their inner-product $l^\top p$ vanishes. A conic is represented by some symmetric 3×3 matrix C. A point p is on a conic C if $p^\top C p = 0$. If the conic matrix is of rank two, the conic is a union of two lines s and r and is given by: $sr^\top + rs^\top$.

Consider two images of two independently moving bod-

*This work has been partly funded by the Israeli Science Ministry grant 1229 and by a grant by the Israeli Science Foundation (ISF). A.S. is on sabbatical at the Computer Science dept. at Stanford University and can be reached at shashua@cs.stanford.edu

ies where each image taken by a different camera at a different time. Thus, due to the relative motion (rotation and translation) between the two bodies, the image measurements per body would be subject to a distinct fundamental matrix. Therefore, if p, p' is a pair of matching points from one of the bodies, and q, q' is a pair of matching points from the other body, then they satisfy the epipolar constraints $p'^{\top}\mathcal{F}^1 p = 0$, $q'^{\top}\mathcal{F}^2 q = 0$. Note that segmentation can be achieved if we can first recover the two fundamental matrices, or recover the epipoles (provided they are distinct) — as detailed later in the paper. Our goal is therefore to describe a constraint applicable to both matching constraints and from which one can later recover the individual fundamental matrices or the epipoles.

2.1 The Segmentation Matrix $\mathcal{S}$

Consider a matching pair p, p' of points. Since the corresponding 3D object point can arise from either one of the two bodies, then the constraint

$$(p'^{\top}\mathcal{F}^1 p)(p'^{\top}\mathcal{F}^2 p) = 0,$$

must hold regardless from which body it arises from. This constraint is a bi-quadratic expression. It is quadratic in each of the measurements p and p'. Using tensorial notation it is written as: $p^i p^j p'^k p'^l \mathcal{W}_{ijkl} = 0$ where $\mathcal{W}_{ijkl} \cong \mathcal{F}^1_{ik}\mathcal{F}^2_{jl}$. The tensor contains 81 elements but cannot be recovered linearly from the measurements themselves because of symmetries. Those symmetries arise from the fact that $\mathcal{W}$ is bi-quadratic — thus each measured point appears twice and in turn we have a 4-fold symmetry:

$$p^i p^j p'^k p'^l = p^j p^i p'^k p'^l = p^i p^j p'^l p'^k = p^j p^i p'^l p'^k.$$

The symmetry implies that, for example, we cannot solve for $\mathcal{W}_{1211}$ linearly, but only for the sum $\mathcal{W}_{1211} + \mathcal{W}_{2111}$. We define instead a more compact structure we call the *segmentation matrix* whose structure is detailed below.

Let $p = (x, y, w)^{\top}$ be the image coordinates and let $\hat{p} = (x^2, xy, y^2, xw, yw, w^2)^{\top}$ be the result of "lifting" p onto a 6-dimensional projective space, and let $\hat{p}'$ be the analog expression for p'. The bi-quadratic expression $p^i p^j p'^k p'^l \mathcal{W}_{ijkl} = 0$ can be transformed into a bi-linear expression in $\hat{p}$ and $\hat{p}'$ of the form $\hat{p}'^{\top}\mathcal{S}\hat{p} = 0$. The 6×6 matrix $\mathcal{S}$ is referred to as the *segmentation matrix*. Each element of $\mathcal{S}$ is a sum of one or more elements of the tensor $\mathcal{W}$. For example $\mathcal{S}_{11} = \mathcal{W}_{1111}$, $\mathcal{S}_{23} = \mathcal{W}_{1222} + \mathcal{W}_{2122}$ and $\mathcal{S}_{24} = \mathcal{W}_{1213} + \mathcal{W}_{2113} + \mathcal{W}_{1231} + \mathcal{W}_{2131}$.

Stating the relationship between $\mathcal{S}$ and $\mathcal{W}$ more formally, let I be the vector $(1, 1, 2, 1, 2, 3)^{\top}$, and J be the vector $(1, 2, 2, 3, 3, 3)^{\top}$. Then $\mathcal{S}_{ij} = c(\mathcal{W}_{I(i)J(i)I(j)J(j)} + \mathcal{W}_{J(i)I(i)I(j)J(j)} + \mathcal{W}_{I(i)J(i)J(j)I(j)} + \mathcal{W}_{J(i)I(i)J(j)I(j)})$, where c is $1/4$ if $I(i) = J(i)$ and $I(j) = J(j)$, $1/2$ if only one of these two equalities hold, and 1 if none of these equalities hold.

Since every pair of matching points p, p' provides one linear constraint $\hat{p}'^{\top}\mathcal{S}\hat{p} = 0$ on $\mathcal{S}$, we need a minimal of 35 matching pairs to achieve a unique solution. Note that $\mathcal{S}$ is determined by a maximal number of 14 parameters (7 for each fundamental matrix), and is therefore subject to 21 independent non-linear constraints. This high number of non-linear constraints make the linear computation of $\mathcal{S}$ somewhat unstable. We will later or show how to use some non-linear constraints in order to make the computation of $\mathcal{S}$ more robust to noise in the image measurements.

Before we investigate the properties of $\mathcal{S}$ and how the motion geometry (fundamental matrix and epipoles) could be extracted from it, it is worthwhile noting that the minimal number of points on a single object does not exceed the number of points required for solving (linearly) the fundamental matrix in a segmented situation. In other words, a minimal of 8 points arising from one object is required (proof omitted) and the remaining (27) points from the other object.

3 Properties of $\mathcal{S}$

The segmentation matrix $\mathcal{S}$ encapsulates the 2-body motion geometry. We show in this section that the epipolar points of the respective fundamental matrices are contained in the null space of $\mathcal{S}$ and $\mathcal{S}^{\top}$.

Claim 1 (Slices of $\mathcal{S}$) *$\mathcal{S}\hat{p}$ is a vector representing the degenerate conic defined by the epipolar lines $\mathcal{F}^1 p$ and $\mathcal{F}^2 p$.*

Proof: Let $d = \mathcal{S}\hat{p}$. The vector d can be folded into a symmetric 3×3 matrix D:

$$D = \begin{bmatrix} d_1 & d_2/2 & d_4/2 \\ d_2/2 & d_3 & d_5/2 \\ d_4/2 & d_5/2 & d_6 \end{bmatrix}.$$

Note that $p'^{\top}Dp' = d^{\top}\hat{p}'$. By construction, $\hat{p}'^{\top}\mathcal{S}\hat{p} = 0$ iff p' lies on the degenerate conic which includes the two epipolar lines $\mathcal{F}^1 p$ and $\mathcal{F}^2 p$. However, any such p' must satisfy $p'^{\top}Dp' = 0$ as well. Therefore D must to be that degenerate conic. □

Note that this property is completely symmetric between the images. $\mathcal{S}^{\top}\hat{p}'$ is a representation of the degenerate conic which is composed out of the lines $\mathcal{F}^{1\top}p'$ and $\mathcal{F}^{2\top}p'$.

Also note that for every point p in the first view we get a degenerate conic containing two epipolar lines, $\mathcal{F}^1 p$ and $\mathcal{F}^2 p$. Each epipolar line contains the epipole of its fundamental matrix. Accumulating such lines over a number of points gives us two pencils of epipolar lines. This is shown in Fig. 1.

Recovery of the epipoles in both images (four epipolar points all together) is therefore quite simple, using for instance a RANSAC procedure. We will show in the next

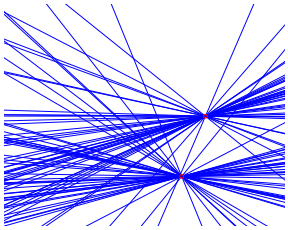


Figure 1: a plot of 40 conics in the second image to which a segmentation matrix is mapping points in the first image.

section a simple algebraic way of deriving the epipoles. It would be based on the following observation on the null space of the segmentation matrix $\mathcal{S}$.

Claim 2 (null space of $\mathcal{S}$) *Let $\mathcal{S}$ be a segmentation matrix as above, arising from the fundamental matrices $\mathcal{F}^1$ and $\mathcal{F}^2$. Let $e_1 \cong null(\mathcal{F}^1)$, $e_2 \cong null(\mathcal{F}^2)$. Let $\hat{e}_1, \hat{e}_2$ be their extension to a vector of six numbers as above. Then $\mathcal{S}\hat{e}_1 = 0, \mathcal{S}\hat{e}_2 = 0$. Also, let $\hat{e}'_1$ and $\hat{e}'_2$ be the extensions of $e'_1 \cong null(\mathcal{F}^{1\top})$, $e'_2 \cong null(\mathcal{F}^{2\top})$, then $\mathcal{S}^\top \hat{e}'_1 = 0, \mathcal{S}^\top \hat{e}'_2 = 0$*

Proof: From the construction of the segmentation matrix $\mathcal{S}\hat{e}_1$ is composed out of sums of the elements of the tensor contraction $\hat{e}^i_1 \hat{e}^j_1 W_{ijkl}$. Where $\mathcal{W}_{ijkl} \cong \mathcal{F}^1_{ik}\mathcal{F}^2_{jl}$. Since $\mathcal{F}^1 e_1 = 0, e^i_1 e^j_1 W_{ijkl} = 0$ as well and so is $\mathcal{S}\hat{e}_1$. Similar arguments show that $\mathcal{S}\hat{e}_2 = 0$, and that $\mathcal{S}^\top \hat{e}'_1 = 0, \mathcal{S}^\top \hat{e}'_2 = 0$. □

The rank of $\mathcal{S}$ is 4 in the general case when the epipoles of the two fundamental matrices are distinct. We will show later how to recover the epipoles in this case. The matter is more subtle when the epipoles coincide as shown next:

Claim 3 (null space of $\mathcal{S}$ when the epipoles coincide) *Let $\mathcal{S}$ be a segmentation matrix as above, arising from the fundamental matrices $\mathcal{F}^1$ and $\mathcal{F}^2$. If $null(\mathcal{F}^1) \cong null(\mathcal{F}^2)$ or $null(\mathcal{F}^{1\top}) \cong null(\mathcal{F}^{2\top})$ then* $\text{rank}\mathcal{S} \leq 3$.

Proof: We will prove the claim for $null(\mathcal{F}^1) \cong null(\mathcal{F}^2)$. The details for the other case are similar. Consider what happens in the lifted coordinates ($\hat{p}$) when we apply a transformation A to a point p. The point $\hat{A}p$ (the lifted transformed point) is

$$((A_1^\top p)^2, (A_1^\top p)(A_2^\top p), (A_2^\top p)^2, (A_1^\top p)(A_3^\top p), (A_2^\top p)(A_3^\top p), (A_3^\top p)^2)^\top$$

where $A_i, i = 1..3$ are the rows of A. This can be written as a matrix $\hat{A}$ times the point $\hat{p}$. If A is invertible, so is $\hat{A}$.

If we apply the transformation A to the first image, then the two new fundamental matrices become $\mathcal{F}^1 A^{-1}$ and $\mathcal{F}^2 A^{-1}$. The new segmentation matrix is $\mathcal{S}\hat{A}^{-1}$. This is because for every p in the first image $\mathcal{S}\hat{p} = \mathcal{S}\hat{A}^{-1}(\hat{A}\hat{p})$.

Since $\mathcal{F}^1$ and $\mathcal{F}^2$ have the same null space we can find an invertible transformation A^{-1} such that the last row of $\mathcal{F}^1 A^{-1}$ and the last row of $\mathcal{F}^2 A^{-1}$ vanish. In the new coordinate system every element of W_{ijkl} such that i or k is 3 vanishes. From the construction of S from W the last three rows of $\mathcal{S}\hat{A}^{-1}$ has to vanish. Since $\hat{A}^{-1}$ is invertible, the rank of $\mathcal{S}$ is at most three. □

To conclude, the rank of S indicates wheather we are in a case where the epipoles coincide or in the general case. We next address the issue of recovering the distinct epipoles from $null(\mathcal{S})$ (in the general case).

3.1 Recovery of the Epipoles

Assume for now that the epipoles are distinct, i.e., $e_1 \cong null(\mathcal{F}^1)$, and $e_2 \cong null(\mathcal{F}^2)$ are distinct. Recall that $\hat{p} = (x^2, xy, y^2, xw, yw, w^2)^\top$ is the lifted representation of p. We will define another representation of p as a 3×3 symmetric matrix $P = pp^\top$. Note that $\hat{p}$ has exactly the same elements as P.

Let $\hat{n}_1, \hat{n}_2$ be the two vectors in the null space of $\mathcal{S}$ and let N_1, N_2 be the representation of $\hat{n}_1\hat{n}_2$ as 3×3 symmetric matrices. We wish to find $\hat{e}_1 = \hat{n}_1 + \lambda_1\hat{n}_2$ and $\hat{e}_2 = \hat{n}_1 + \lambda_2\hat{n}_2$, such that there exist points e_1, e_2 which map to $\hat{e}_1, \hat{e}_2$. Let $E_1 = N_1 + \lambda_1 N_2 = e_1 e_1^\top$ and $E_2 = N_1 + \lambda_2 N_2 = e_2 e_2^\top$.

Note that E_1 and E_2 are rank one matrices, thus any minor of these matrices must vanish. In other words, let M_1, M_2 be any 2×2 sub matrices of N_1, N_2, respectively, then the determinant of $(M_1 + \lambda_i M_2)$ vanishes and therefore λ_i are generalized eigen-values. This gives us 9 second order polynomials.

Note that all of these polynomial share the same roots, and so they are all equal up to scale. In order to increase our accuracy we first compute the norm of each such polynomial (viewed as a vector) and discard those whose norm is below some threshold, then normalize the remaining polynomials. We then solve the polynomial which is the mean of those normalized polynomials.

In case the epipoles in at least one of the images coincide, the solution given above for the recovery of the epipoles does not hold. The null space of the segmentation matrix is not spanned only by valid points. However in this case the epipoles can be recovered for example by first recovering the fundamental matrices (see section 4).

3.2 Segmentation using Epipolar Points

If the two epipoles in both images are distinct we can use them to perform segmentation. We do not need to recover the underlying fundamental matrices, we only need these epipoles and the segmentation matrix.

Assume that we have recovered the segmentation matrix $\mathcal{S}$, and extracted the epipoles in the second image from

it. We are given a point p in the first image and a matching point p' in the second image. The segmentation matrix maps p to the degenerate conic represented by $\mathcal{S}\hat{p}$. This conic is composed out of the two epipolar lines $\mathcal{F}^1 p$, $\mathcal{F}^2 p$. One of these epipolar lines passes through the first epipole in the second image and the other through the second epipole in the second image.

We then check to see on which one of the two epipolar lines the point in the second image p' lies. If it lies on an epipolar line passing through the first epipole, it belongs to the first object. If it lies on an epipolar line passing through the second epipole, it belongs to the second object. If the point p' lies on the intersection of these two epipolar lines, then the two views are not sufficient in order to segment this point using epipolar constraints.

4 Recovery of the Fundamental Matrices using $\mathcal{S}$

The segmentation method based on the epipoles applies only when the epipoles are distinct. Alternatively we can recover the fundamental matrices directly from $\mathcal{S}$ as described below (the method works well even if the epipoles coincide).

Consider the vector $\mathcal{S}[1,0,0,0,0,0]^\top$ representing the degenerate conic in the second image associated with the point $p = [1,0,0]^\top$ in the first image. According to Claim 1 the underlying lines of that conic are $\mathcal{F}^1 p$, $\mathcal{F}^2 p$. These lines are the first columns of the two fundamental matrices $\mathcal{F}^1$ and $\mathcal{F}^2$. In a similar fashion we can recover the second and third columns as well. The problem is that we do not know to which fundamental matrix each column belongs to (the labeling problem), and the relative scales of these columns are unknown as well.

We overcome the labeling problem by considering all eight possibilities. Each possibility is a pair: a fundamental matrix for the first object, and a fundamental matrix for the second object.

To compute the scales we look at the degenerate conic in the first image $\mathcal{S}^\top[1,0,0,0,0,0]^\top$ associated with the point $p' = [1,0,0]^\top$ in the second image. This conic is composed out of the first row of $\mathcal{F}^1$ and the first row of $\mathcal{F}^2$. We now assign a label to each of the lines which make this conic. One row would be the first row of the first fundamental matrix, and the other would be the first row of the second fundamental matrix. Using these rows we can set the scales between the columns of all the eight pairs of possible fundamental matrices.

There might be a problem in cases where there are vanishing elements in these rows. However in this case we can use other rows as well. Alternatively, we can also pick points not in the standard basis. For example, picking the point $p = [1,1,0]^\top$ allows us to set the scales between the first two columns of the fundamental matrices. [1].

To check which one of the eight possibilities is the correct one, we can for example construct a segmentation matrix from each possibility, and measure the similarity to the estimated segmentation matrix $\mathcal{S}$. Alternatively, one can apply the segmentation matrix to a random collection of points and measure the similarity of the epipolar lines to the epipolar lines of the estimated $\mathcal{S}$.

Once the fundamental matrices have been recovered, the segmentation can be achieved by examining the distance of the matched point in the second image from the epipolar line we obtain from the fundamental matrices of each of the objects.

5 Reduced Configurations

So far we have addressed the general motion segmentation problem where the views are full projective and the relative motion between the two bodies is general as well. There are reduced situations of interest such as (i) the affine camera model, and (ii) motion between the bodies is pure translation (we can recover the homography at infinity and thus achieve affine calibration).

5.1 Affine Cameras

In the affine camera model (parallel projection) the upper left 2×2 matrix of each of the fundamental matrices vanishes. In this case many of the elements in the segmentation matrix vanish as well:

Claim 4 (Segmentation matrix in Affine case) *Let $\mathcal{S}$ be a segmentation matrix arising from* Affine *fundamental matrices $\mathcal{F}^1$ and $\mathcal{F}^2$, then there exist 21 vanishing elements in $\mathcal{S}$.*

Proof: Consider the multi-linear tensor $\mathcal{W}_{ijkl} \cong \mathcal{F}^1_{ik}\mathcal{F}^2_{jl}$. Any element of $\mathcal{W}_{ijkl}$ such that i and k are either 1 or 2 is zero. So is every element with indexes j and l which are either 1 or 2. Let I be the vector $(1,1,2,1,2,3)^\top$, and J be the vector $(1,2,2,3,3,3)^\top$. Recall that $\mathcal{S}_{ij} = c(W_{I(i)J(i)I(j)J(j)} + W_{J(i)I(i)I(j)J(j)} + W_{I(i)J(i)J(j)I(j)} + W_{J(i)I(i)J(j)I(j)})$, where c is two to the power of minus the number of equalities that hold out of the following two: $I(i) = J(i)$, $I(j) = J(j)$. $\mathcal{S}_{ij}$ vanishes if at least three out of the four indexes $I(i), J(i), I(j), J(j)$ are either 1 or 2. This happens exactly 21 times, where (i,j) if one of the following pairs: $(1,1), (1,2), (1,3), (1,4), (1,5), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (3,4), (3,5), (4,1), (5,1), (4,2), (5,2), (4,3), (5,3)$. □

[1] since the fundamental matrices are of rank 2, we cannot set all the scales with only one p using this method

Therefore there are additional 21 linear constraints on the segmentation matrix and one needs only 14 points in order to solve it. Note that in [2] less points were required (the minimal number is not reported, but it has to be at least 8), but here we use only two views.

5.2 Pure Translation

In [12] three views and 13 points were used to solve the case where motion between the bodies is pure translation (however 4 pre-segmented points were required). If we add the assumption of pure relative translation and solve for the segmentation matrix, we obtain a useful property: *the mapping between the point in the first view and the center of the degenerate conic that the segmentation matrix maps to is the homography at infinity* H_∞. This is because $(H_\infty p)^\top \mathcal{F}^1 p = (H_\infty p)^\top \mathcal{F}^2 p = 0$ for all p. In other words, the point $H_\infty p$ is on both epipolar lines, thus it has to be on their intersection.

Using the property that the mapping between p and the center of the degenerate conic (in the second view) is governed by H_∞, we can recover H_∞ and obtain a 3D affine reconstruction for each one of the moving bodies.

The recovery of the homography at infinity from the point to conic centers mapping requires some care — we must avoid conics of rank 1. Consider points on the first image which the homography at infinity maps to points between the two epipoles in the second image. The matching conic in the second image is composed out of the same line repeated twice, and the center of the conic is ill defined. This phenomenon is demonstrated in the experiment described in Fig. 2. To overcome this we can compute the homography at infinity using only points which are mapped to conics composed out of two well separated lines (rank 2 conics).

6 Non-linear Constraints

There are several sources for the existence of non-linearities between the elements of the segmentation matrix. The first source of non-linear constraints comes from the fact that the 35 elements up to scale of the segmentation matrix are sums of bi-linear products of two 3×3 matrices defined up to scale. This gives us $35 - 16 = 19$ second order constraints.

Second source of non-linearity arises from the fact that each fundamental matrix $\mathcal{F}^1, \mathcal{F}^2$ is a rank 2 matrix. This gives two additional constraints, which are cubic in the elements of the fundamental matrices, and can be of higher order on the elements of the segmentation matrix. For example, the 5'th order constraints arising from the property that every 5×5 minor of the segmentation matrix vanishes (the segmentation matrix has at most rank 4) has its source in the cubic constraints of the fundamental matrices. A third source of non-linearity comes from the fact the the two cameras share the same internal parameters. Then, for example, if we know the cameras calibration, then each fundamental matrix is determined only by three elements for rotation and two for translation up to scale.

In this paper we chose to ignore most of the non-linear constraints and instead obtain a simple and manageable algorithm. Ultimately it is a matter of a trade-off between the possibility of achieving higher accuracies in the presence of noise and the resulting added complexity of doing so. We did however incorporate some non-linear constraints in an iterative process aimed at making the estimation of the segmentation matrix more stable. This is described in the next subsection.

6.1 Improving the Estimation of $\mathcal{S}$ using an Iterative Process

Assume for now that we know the epipoles in the first image e_1, e_2. We know that each one of the rows of $\mathcal{S}$ is perpendicular to $\hat{e}_1, \hat{e}_2$. Assuming e_1 and e_2 are distinct, thus each row of $\mathcal{S}$ is spanned by four known basis vectors.

Let U be a 6×4 matrix with columns which span the rows of $\mathcal{S}$. Let C be a 36×24 block diagonal matrix which has U as each one of its six blocks. The 36 elements of $\mathcal{S}$ can be computed as the matrix C multiplied by a vector of 24 entries x.

Let E be our estimation matrix for the matrix $\mathcal{S}$ (the dimension of E is the number of points times 36). The algebraic error of each computed segmentation matrix $\mathcal{S}$ is simply the norm of E times the 36 elements of S.

In order to find an $\mathcal{S}$ whose null space is spanned by the epipoles e_1, e_2, and which minimizes the algebraic error, we minimize ECx subject to $\| Cx \| = 1$. An algorithm to solve such a minimization problem is given in [5] p.566.

We now have a way to perform a mapping from the two epipoles in the first image to a segmentation matrix which minimizes algebraic error. Finding the best epipoles in terms of the algebraic error is a non-linear minimization problem, which can be solved using the Levenberg-Marquardt algorithm. We represent each epipole as three numbers in order to allow the epipoles to be at infinity, and our non-linear minimization problem has a manageable size of 6 parameters.

This algorithm was inspired by similar algorithms given in [5] for the computation of the fundamental matrix and of the trifocal tensor. Unlike those algorithms, the segmentation matrix we get at each stage of the minimization procedure is not guaranteed to satisfy *all* the non-linear constraints. This is basically because both the fundamental matrix and the trilinear tensor are linear in homography matrices, while the segmentation matrix is bi-linear in the underlying remaining parameters.

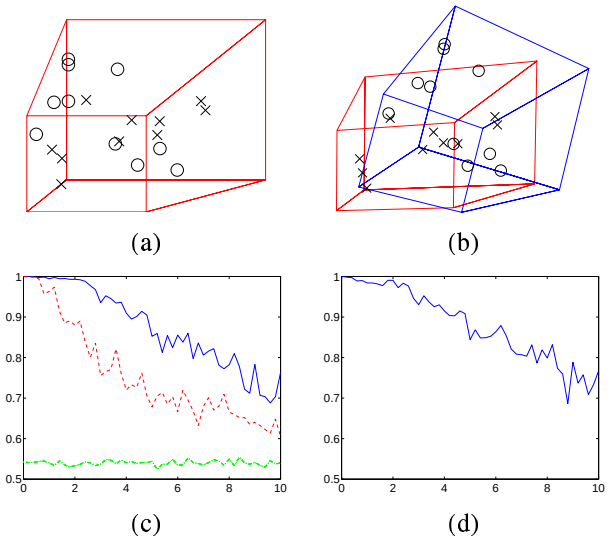


Figure 2: (a) The image of the box in 3D from which points were uniformly sampled. (b) Motion of the two objects and the camera is illustrated by the image of the moved boxes. (c) Grades of segmentations using epipole-based segmentation matrix algorithm (top line), fundamental matrices extracted from the segmentation matrix (middle line), and of totally random segmentation (bottom line). Grades are shown against standard deviation of the noise in a 1800×1800 image (d) Grades of the segmentation picked by the heuristic against standard deviation of the noise.

7 Experiments

We begin with a synthetic experiment designed to test the sensitivity of the method in the presence of (normally distributed) noise. For each of the 20 experimental trials, 50 points (with added noise with varying variance) were sampled from each object (all points were sampled uniformly from a box, see Fig. 2a,b). The performance was measured as a function of the number of miss-classification by the segmentation program — a performance rating of 0.5 was considered as pure chance. The noise level varied from 0 to a standard deviation of 10 pixels (where the image size was 1800×1800).

Fig. 2c shows the performance of segmentation using the segmentation matrix based on epipoles (top line), of a segmentation based on using fundamental matrices which were extracted from the segmentation matrix (middle line) and of a totally random segmentation. Each reported grade is a mean value of the 20 trials. Note that in the majority of the runs the performance of segmentation based on epipoles outperformed that based on the fundamental matrices, although in some runs the fundamental matrix based segmentation was better.

We used a simple heuristic in order to pick automatically the best segmentation out of the two (based on epipole, based on fundamental matrix) at each trial. Each segmentation gives us two groups. We picked the larger group from each one of these pairs. Then we computed a fundamental matrix from these larger groups. We computed the median of the distances from the epipolar lines to all the 100 points. The heuristic picked the segmentation with the lower median. The grades of the segmentation result picked by this heuristic is shown in 2d. The heuristic picked the best segmentation in about 75% of the cases, and the mean difference of the grade of segmentation picked by the heuristic and the best grade was less then 0.02.

7.1 Real Imagery Experiments

In Fig. 3 we evaluate the performance of the estimation of the homography at infinity in the case of pure translation. We have chosen an outdoor setting in which one body is the background scene and the second body is a moving vehicle along a straight section of the roadway (the motion between the two bodies is therefore pure translation). Fig. 3(a,b) displays two images of the moving car taken by a moving camera. Points were tracked using openCV's [8] pyramid LK tracker. The segmentation results using our algorithm are shown in Fig. 3(c,d) — which illustrates a good segmentation – there were no miss-classifications.

We then estimated the homography at infinity from the point to conic center matching. Fig. 3(e) shows the warp of the first image with H_∞ using all point to conic matches. Once can see that some of the matches were with conics of rank=1 which had a detrimental effect on the result. Fig. 3(f) shows the warp when only rank-2 conics were used — the warped image looks undistorted.

Once the first image is warped with H_∞, the rotational component of camera motion is cancelled. One way to measure the quality of H_∞ is to create in-between frames and observe the perspective distortion — a lack of perspective distortion suggests a good fit. A sequence of synthetic images are shown in the last row of Fig. 3 illustrating a smooth and undistorted animation.

Fig 4(a,b) shows two images taken from a camera moving inside the cabin of a vehicle while the vehicle is in motion. The bodies are the static cabin features and the background scene. The relative motion between the two bodies is general as the car was in the middle of making a turn. Fig 4(c,d) illustrate the segmentation results — there were 3 miss-classifications when extracting the background and no miss-classifications when extracting in-cabin features.

The last experiment shown is an example of the algorithm's behavior on low relief objects which occupy a narrow region in the image. Fig. 5(a,b) are two images of a cat moving its head with overlayed tracked points. The cat's head captures only a small part of the image, and not many points were tracked on the head. We can expect the compu-

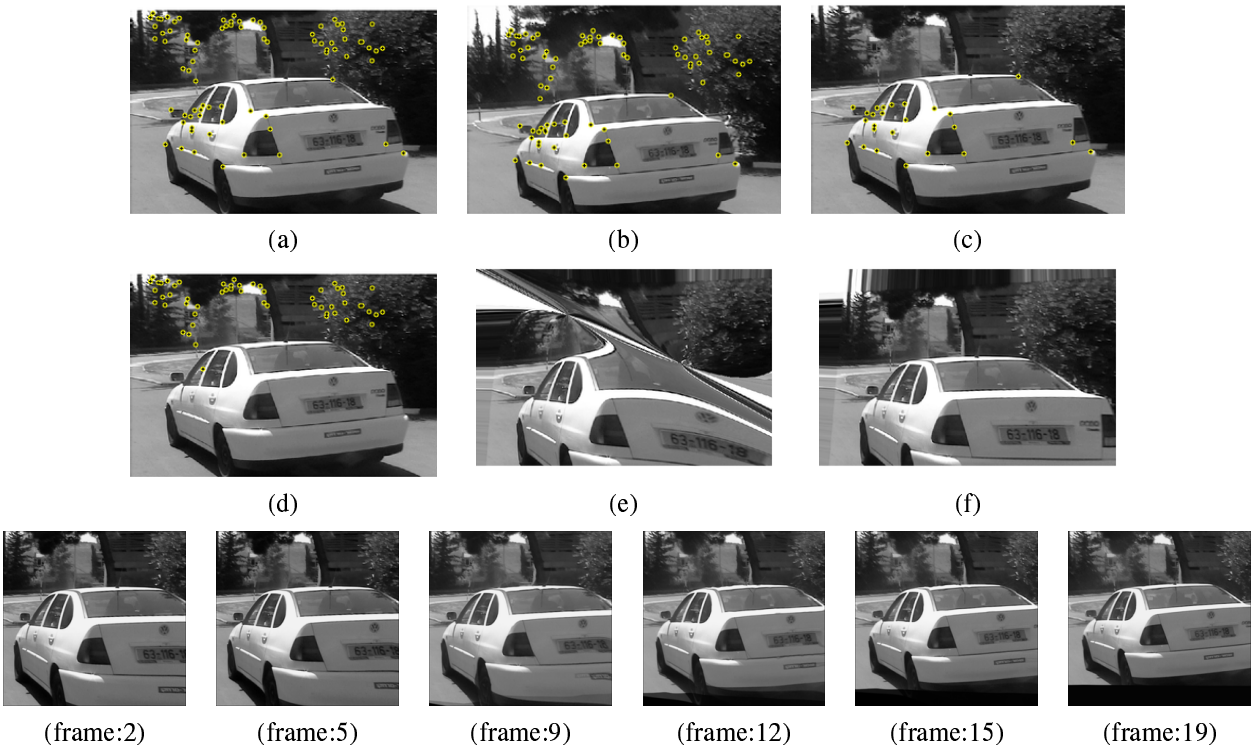


Figure 3: (a), (b) Two images of a moving car with overlaid tracked points. (c) Points on first object (car) segmented using fundamental matrices extracted from the segmentation matrix. (d) Points on the second object (background). Since we are in a pure translation case we can compute the homography at infinity using the point to center of conic mapping. (e) shows the results warping the first image by this mapping using all matches, including those corresponding with rank-1 conics. Note that on the line connecting the epipoles this mapping is degenerate. (the epipoles are seen as locus of bent lines). (f) shows the first image warped by the homography at infinity computed using only rank-2 conics. Last row shows the result of interpolation made after canceling the effects of rotation using the homography at infinity.

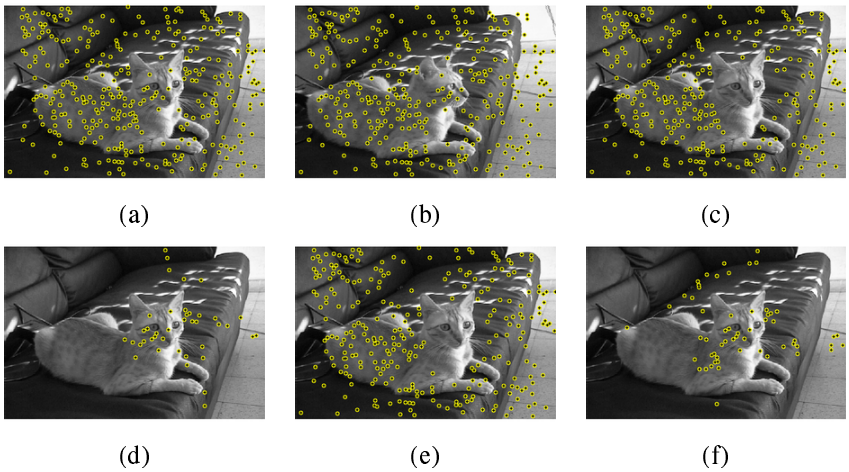
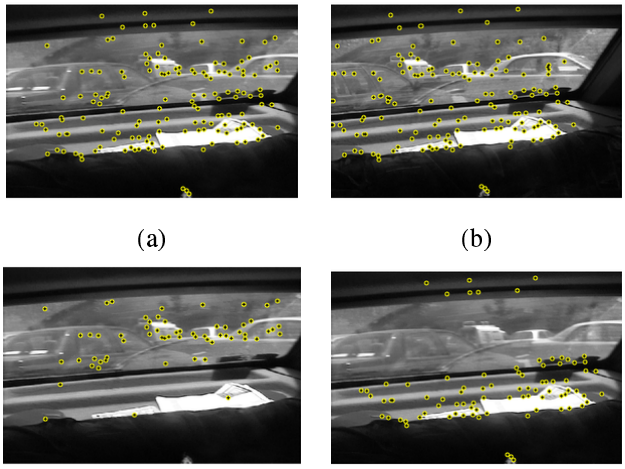


Figure 5: Example of the algorithm's behavior for low relief objects. (a), (b) Two images of a cat moving it's head with overlaid tracked points. The images were segmented by hand, then a fundamental matrix was computed for each object. These fundamental matrices were used for segmentation and the resulting objects are seen in images (c,d). (e,f) The result of segmentation using fundamental matrices computed from the segmentation matrix. Note the comparable performance.



(a) (b)

(c) (d)

Figure 4: Example of algorithm's performance for general motion. (a), (b) Two images of a back window of a car with overlaid tracked points. The car is rotating into a parking space, and so the relative motion of the two objects (inside of car, outside the car) is a combination of translation and rotation. (c,d) The result of segmentation using epipoles computed from the segmentation matrix.

tation of the fundamental matrix computed from points on the head to be ill-posed.

The scene was first segmented by hand. The two fundamental matrices were computed, and segmentation was performed using these fundamental matrices. Fig. 5(c,d) shows the resulting segmentation. Note that the group of background points was segmented without error, while the group of points on the cat's head is contaminated by outliers and ill-tracked points (such as points on the cat's front leg, which were affected by the cat's whiskers).

Applying segmentation based on fundamental matrices extracted from the segmentation matrix gives comparable performance as seen in fig. 5(e,f).

8 Summary

We have presented a method for 2-body motion segmentation using two perspective views. The method is based on the introduction of the segmentation matrix $\mathcal{S}$ which forms a constraint which applies to matching points arising from both bodies. The slices of $\mathcal{S}$ and its null spaces encapsulate all the necessary information for recovering the underlying fundamental matrices and epipoles of each of the moving bodies. Furthermore, when the bodies are known to move relative to each other in pure translation, the segmentation matrix contains the information for recovering the homography at infinity and thus obtaining an affine calibration of the camera geometry.

Acknowledgments

We would like to thank T. Kanade and J. P. Costeira for sharing with us their experimental data.

References

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

[2] J. P. Costeira and T. Kanade, A multibody factorization method for independently moving objects In *International Journal on Computer Vision*, 29-3 (1998), 159-179.

[3] A.W. Fitzgibbon and A. Zisserman. Multibody Structure and Motion: 3-D Reconstruction of Independently Moving Object. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Dublin, Ireland, June 2000.

[4] M. Han and T. Kanade. Reconstruction of a Scene with Multiple Linearly Moving Objects. In *Proc. of Computer Vision and Pattern Recognition*, June, 2000.

[5] R.I. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[6] K. Kanatani. Motion Segmentation by Subspace Separation and Model Selection. In *International Conference on Computer Vision (ICCV)*, Vancouver, Canada, July, 2001.

[7] R.A. Manning and C.R. Dyer. Interpolating view and scene motion by dynamic view morphing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 388–394, Fort Collins, Co., June 1999.

[8] Open source computer vision library *http://www.intel.com/research/mrl/research/cvlib/*

[9] P. H. S. Torr. Geometric motion segmentation and mosel selection. In *Phil. Trans. Roy. Soc.*, A-356 (1998), 1321-1340.

[10] A. Shashua and L. Wolf. Homography tensors: On algebraic entities that represent three views of static or moving planar points. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Dublin, Ireland, June 2000.

[11] Y. Wexler and A.Shashua. On the synthesis of dynamic scenes from reference views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, South Carolina, June 2000.

[12] L. Wolf and A. Shashua. On Projection Matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, ..., 6$, and their Applications in Computer Vision. In *International Conference on Computer Vision (ICCV)*, Vancouver, Canada, July, 2001.

[13] L. Wolf, A. Shashua and Y. Wexler. Join Tensors: on 3D-to-3D alignment of Dynamic Sets. In *Proc. of the Int. Conf. on Pattern Recog. (ICPR)*, Sep. 2000, Barcelona, Spain