

Affine 3-D Reconstruction from Two Projective Images of Independently Translating Planes

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

Abstract

Consider two views of a multi-body scene consisting of k planar bodies moving in pure translation one relative to the other. We show that the fundamental matrices, one per body, live in a 3-dimensional subspace, which when represented as a step-3 extensor is the common transversal on the collection of extensors defined by the homography matrices H_1, \dots, H_k of the moving planes. We show that as much as five bodies are necessary for recovering the common transversal from the homography matrices, from which we show how to recover the fundamental matrices and the affine calibration between the two cameras.

1 Introduction

In the context of multiple-view geometry it is a well known fact that two perspective views of a 3D scene consisting of two planar configurations of points (planar objects) completely determine the projective calibration between the two cameras. For instance, let H denote the 3×3 collineation (homography matrix) from image 1 to 2 induced by some planar object, and let F denote the (unknown) fundamental matrix, then $H^T F$ is a skew-symmetric matrix. Therefore, each planar object provides 6 linear constraints on F , thus two planar objects are sufficient to uniquely determine F . Once F is known the camera projection matrices and the projective reconstruction of scene points can be recovered (see [10] for a recent detailed overview of such material).

Given the growing body of work on *dynamic* scenes [1, 14, 12, 16, 9, 8, 4], i.e., 3D scenes which contain multiply moving points or collections of points (bodies) seen under multiple views, we wish to extend the basic paradigm described above to the case where the scene contains multiple planar objects moving relative to each other by pure translational motion while the camera is changing position.

Our goal is to utilize the new source of information induced by the movement of the planar objects relative to each other in order to recover (i) the fundamental matrices F_1, \dots, F_k , one for each planar object, from the corresponding homography matrices H_1, \dots, H_k , and (ii) the affine calibration between the two cameras.

As mentioned above, a single homography matrix is not sufficient for recovering the fundamental matrix, however, the relative motion among the multiple planes can be harnessed to appropriately introduce additional constraints from which the fundamental matrix of each planar object can be recovered and, moreover, to recover the homography at infinity H_∞ (which in turn provides the affine calibration between the two cameras).

We show that the additional constraints are embedded in the problem of finding a common *transversal* in \mathcal{P}^8 which represents the family of 3×3 matrices up to scale. Because the fundamental matrices F_1, \dots, F_k share the internal parameters of the two cameras and the relative rotation among the planes remains fixed, one can show that all such matrices live in a 3-dimensional subspace \mathcal{F} of \mathcal{P}^8 , i.e., three fundamental matrices span the entire family of fundamental matrices associated with moving objects under pure translational motion. This 3-dimensional subspace can be captured as a common transversal of other 3-dimensional subspaces \mathcal{F}_i each defined from the homography matrices H_1, \dots, H_k . In other words, each H_i , $i = 1, \dots, k$, defines the corresponding F_i up to a null space of dimension three (because the skew-symmetry of $H_i^T F_i$ provides 6 constraints on the 9 entries of F_i) denoted by \mathcal{F}_i . All the spaces \mathcal{F}_i intersect with \mathcal{F} , thus the question is *how many intersections are required in order to uniquely determine \mathcal{F} ?*

It is worthwhile noting that the issue of finding a common transversal in the context of dynamic scenes was first introduced in [1]. There the application of transversals was classic, i.e., finding the common intersecting 3D line (the trajectory of a moving point) of 4 other lines is a well known

exercise in invariant theory (see for example, [15]).

We will start with the necessary mathematical tools required for representing subspaces as a single object (extensors) and the operation of subspace addition (the “join”) required for the calculation of transversals. We will then proceed to the general case (translational motion is general) and show that 5 planar objects are necessary for recovering (linearly) \mathcal{F} . We will then discuss the recovery of affine calibration, special cases (such as collinear motion) and briefly touch upon the issue of incorporating non-linear constraints in the estimation.

2 Mathematical Preliminaries: Extensors and the Join Operation

The mathematical component of our work deals with intersecting and joining subspaces for the purposes of finding common transversals in the 8-th dimensional projective space \mathcal{P}^8 . A convenient way to do so is to treat a k -dimensional subspace as a single object (instead of as a collection of k basis vectors) which is done using Grassmann coordinates also known as an “extensor of step k ”. Generally, the algebra of extensors with the operations of intersection (“meet”) and union (“join”) are also known as double algebra or Grassmann-Cayley algebra. These were first introduced in the context of multiple-view geometry by [3, 7, 6] and also in the context of projection matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$ [17]. A concise introduction to extensors and the operations of meet and join can be found in [15, 2]. Some of the material described below, especially Claim 1, is not found in the scope of the references above thus it is recommended to read through the entire section before proceeding to the remainder of the paper.

An extensor of step k describes a subspace of dimension k of some n -dimensional vector space V . All extensors of step k lie in the linear space $\bigwedge^k(V)$ which has the dimension $\binom{n}{k}$. The join operator (\vee) is a multilinear antisymmetric operator that takes two extensors of steps j and k and produces an extensor of step $j + k$. The joint extensor is associated with the direct sum of the linear spaces associated with the two extensors. This join extensor vanishes if the two generating extensors intersect. If e_1, e_2, \dots, e_n is a basis of V then the basis for $\bigwedge^k(V)$ is given by $\binom{n}{k}$ basis elements:

$$\{e_{j_1} \vee e_{j_2} \vee \dots \vee e_{j_k} \mid 1 \leq j_1 < \dots < j_k \leq n\}$$

Let $\bar{A} = \text{span}\{a_1, \dots, a_k\}$ be a k -dimensional subspace of V where a_1, \dots, a_k is some choice of basis. The step k extensor $A = a_1 \vee \dots \vee a_k$ also denoted by $A = a_1 a_2 \dots a_k$ is an element of the vector space $\bigwedge^k(V)$:

$$A = \sum_{1 \leq j_1 < \dots < j_k \leq n} A_{j_1, \dots, j_k} e_{j_1} \vee \dots \vee e_{j_k}$$

where the scalars A_{j_1, \dots, j_k} are $k \times k$ minors:

$$A_{j_1, \dots, j_k} = \begin{vmatrix} a_{1j_1} & a_{1j_2} & \dots & a_{1j_k} \\ a_{2j_1} & a_{2j_2} & \dots & a_{2j_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{kj_1} & a_{kj_2} & \dots & a_{kj_k} \end{vmatrix}$$

Thus the extensor A has $\binom{n}{k}$ coefficients (choices of $k \times k$ minors from the $k \times n$ matrix whose rows consist of a_1, \dots, a_k). The extensor A represents the subspace \bar{A} as we note that

$$\bar{A} = \{u \in V \mid A \vee u = 0\}$$

(all $(k + 1) \times (k + 1)$ minors vanish thus $u \in \text{span}\{a_1, \dots, a_k\}$) while on the other hand the determinant expansions are invariant to a change of basis of \bar{A} .

Let $A = a_1 \dots a_k$ and $B = b_1 \dots b_j$ be extensors of step k, j representing subspaces \bar{A}, \bar{B} and $k + j \leq n$. Then $A \vee B = a_1 \dots a_k b_1 \dots b_j$ is non-zero (at least one coefficient does not vanish) iff the set $a_1, \dots, a_k, b_1, \dots, b_j$ is linearly independent (i.e., $\bar{A} \wedge \bar{B} = \emptyset$). In this case,

$$\bar{A} + \bar{B} = \overline{A \vee B} = \text{span}\{a_1, \dots, a_k, b_1, \dots, b_j\}$$

Thus, the algebraic join of extensors corresponds to the geometric join of linear subspaces. Conversely, in case $k + j > n$ the subspaces \bar{A}, \bar{B} always have a non-vanishing intersection into a $k + j - n$ dimensional linear space. Thus, it is possible to define a “meet” operation $A \wedge B$ which would be a linear combination of extensors of step $k + j - n$. We will not make use of meet operations¹ in this paper. Further details can be found in [15, 2].

It would be useful for later to describe the coefficients of $C = A \vee B$ as a function of the coefficients A_{i_1, \dots, i_k} of the extensor A and the coefficients B_{i_1, \dots, i_j} of the extensor B . This function is bilinear and has the following form:

Claim 1 *Let $A = a_1 \dots a_k$ and $B = b_1 \dots b_j$ be extensors of step k, j , respectively, where $k + j \leq n$. Let $C = A \vee B$ be their join. Each coefficient $C_{l_1, \dots, l_{k+j}}$, $1 \leq k_1 < \dots < k_p \leq n$, can be described as follows:*

$$\sum_{\substack{\sigma \in S_{k+j} \\ \sigma_1 < \dots < \sigma_k \\ \sigma_{k+1} < \dots < \sigma_{k+j}}} \text{sgn}(\sigma) A_{l_{\sigma_1}, \dots, l_{\sigma_k}} B_{l_{\sigma_{k+1}}, \dots, l_{\sigma_{k+j}}} \quad (1)$$

where S_p is the group of permutations of p letters and $\sigma(i)$, denoted by σ_i is the permuted position of the i 'th letter.

¹This may appear at first as counter-intuitive since this paper is about finding a common transversal which requires intersection of subspaces. However, the meet operation is not defined for $k + j \leq n$ therefore we will use the fact that $A \vee B$ vanishes in the process of recovering the common transversal.

Proof: We need to show that the function above is equal to the determinant expansion

$$\begin{vmatrix} a_{1l_1} & a_{1l_2} & \dots & a_{1l_{k+j}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{kl_1} & a_{kl_2} & \dots & a_{kl_{k+j}} \\ b_{1l_1} & b_{1l_2} & \dots & b_{1l_{k+j}} \\ \vdots & \vdots & \ddots & \vdots \\ b_{jl_1} & b_{jl_2} & \dots & b_{jl_{k+j}} \end{vmatrix}$$

Which in turn is equal to

$$\sum_{\sigma \in S_{k+j}} \text{sgn}(\sigma) a_{1\sigma_1} \cdots a_{k\sigma_k} b_{1\sigma_{k+1}} \cdots b_{j\sigma_{k+j}} \quad (2)$$

Note that each term of eqn. 1 is a superposition of $j!k!$ basic terms (from $A_{l_{\sigma_1}, \dots, l_{\sigma_k}} B_{l_{\sigma_{k+1}}, \dots, l_{\sigma_p}}$) matched to those of eqn. 2 multiplied by the appropriate sign. The number of terms in eqn. 1 is $\binom{j+k}{k}$ which brings a total of $\binom{j+k}{k} j!k! = (k+j)!$ basic terms. Thus, since each basic term of eqn. 1 is distinct and matched to one of the terms of eqn. 2 and the total number of terms match, we have an equality. \square

3 The Space of Fundamental Matrices of Translating Bodies

Consider a collection of k planar point-configurations undergoing pure translation viewed by two projective images, and assume that the corresponding homography matrices from view 1 to 2 induced by the planar configurations have been recovered and denote them by H_1, \dots, H_k . Let $H_\infty = K'RK^{-1}$ be the (unknown) homography induced by the plane at infinity, where K, K' are the matrices representing the internal parameters of the two cameras and R is an orthonormal matrix representing the relative rotational component between the cameras positions. Since in the stated problem domain K, K', R remain fixed as only the translational component is changing among planar objects, the k fundamental matrices, one per object, have the form $F_i \cong [t_i]_\times H_\infty$, where $[u]_\times$ is the skew-symmetric matrix of vector products, i.e., $[u]_\times v = u \times v$. Note that if all the planar objects were static, then $t_1 = \dots = t_k = t$ which is the epipole in view 2 (projection of camera center 1 onto view 2).

In the general case where all objects are in motion and the motion vectors t_i span a 3D space, one can easily show that the fundamental matrices live in a 3-dimensional subspace of the 9-dimensional space of 3×3 matrices. This is shown next.

Let \bar{F} be the vector representing a fundamental matrix F by scanning the matrix column by column. We wish to show that $\text{rank}[\bar{F}_1, \dots, \bar{F}_k] \leq 3$. Let h_1, h_2, h_3 be the three

columns of H_∞ , then $\bar{F}_i = -([h_1]_\times t_i, [h_2]_\times t_i, [h_3]_\times t_i)^\top$. Thus,

$$[\bar{F}_1, \dots, \bar{F}_k]_{9 \times k} = - \begin{bmatrix} [h_1]_\times \\ [h_2]_\times \\ [h_3]_\times \end{bmatrix}_{9 \times 3} [t_1, \dots, t_k]_{3 \times k}$$

which makes the point that all such fundamental matrices live in a 3-dimensional linear subspace of R^9 . Denote the step-3 extensor representing that linear subspace as $\mathcal{F} = f_1 f_2 f_3$ where f_1, f_2, f_3 is some basis of this subspace. Next we will show how to find the extensor \mathcal{F} as a common transversal with other extensors induced by the known homography matrices H_1, \dots, H_k .

Recall that $H_i^\top F_i$ is a skew-symmetric matrix, i.e., for every point p in view 1, $H_i p$ is a point along the epipolar line $F_i p$ — therefore, $(H_i p)^\top F_i p = 0$. We have therefore 6 linear constraints on \bar{F}_i . Let f_1^i, f_2^i, f_3^i be a basis of the null space of the linear system and let $\mathcal{F}_i = f_1^i f_2^i f_3^i$ be the corresponding step-3 extensor. We also know that the join $\mathcal{F} \vee \mathcal{F}_i$ must vanish (because \bar{F}_i is contained in both subspaces). Therefore, the step-3 extensor \mathcal{F} is a common transversal on all the step-3 extensors \mathcal{F}_i .

Recall from Claim 1 that the join $\mathcal{F} \vee \mathcal{F}_i$ can be expressed as a bilinear function of the coefficients of the extensors \mathcal{F} and \mathcal{F}_i . Since the coefficients of \mathcal{F}_i are known, we end up with a linear system of $\binom{9}{3} = 84$ homogeneous equations on \mathcal{F} . Among the 84 equations only 20 are linearly independent. To see why this is so, consider the general question: given $A = a_1 \cdots a_k$ a step- k extensor representing the subspace \bar{A} of the n -dimensional vector space V , how many step- j extensors B , representing subspace \bar{B} , satisfy $\bar{A} \wedge \bar{B} = \emptyset$? Consider a change of coordinates of V such that $A = e_1 \cdots e_k$. The set of all basis extensors of step j is

$$\{e_{i_1} \vee \cdots \vee e_{i_j} | 1 \leq i_1 < \cdots < i_j \leq n\}.$$

The basis extensors which do not intersect A must satisfy $i_1, \dots, i_j \notin \{1, \dots, k\}$. Thus we have $\binom{n-k}{j}$ basis extensors B which satisfy $\bar{A} \wedge \bar{B} = \emptyset$. In particular, $n = 9, k = j = 3$ results in $\binom{6}{3} = 20$ (out of 84). In other words, each given extensor \mathcal{F}_i provides at most 20 linearly independent constraints for the common transversal \mathcal{F} .

The next issue is whether the combined set of 40 linear equations arising from two extensors \mathcal{F}_i and \mathcal{F}_j on the unknown extensor \mathcal{F} is linearly independent? One can show that the second extensor provides only 19 linearly independent equations on top of the 20 equations provided by the first extensor. To see why this is so, consider again the general question: Given A, B extensors of step k, j which satisfy $\bar{A} \wedge \bar{B} = \emptyset$, how many extensors E of step q satisfy both $\bar{A} \wedge \bar{E} = \emptyset$ and $\bar{B} \wedge \bar{E} = \emptyset$? Similar to before, we select a change of coordinates of V such that $A = e_1 \cdots e_k$ and $B = e_{k+1}, \dots, e_{k+j}$, thus we have $\binom{n-(k+j)}{q}$ basis extensors E . In particular, $n = 9, k = j = q = 3$ results

in $\binom{3}{3} = 1$. In other words, the additional set of 20 equations provided by the second extensor has one equation in common with the previous set of 20 equations from the first extensor. Likewise, the third extensor will provide only 18 independent equations because it will have one equation in common with the first extensor and another equation with the second extensor, and so forth. Thus 5 intersections are needed ($20+19+18+17+16=90$) for a complete system for recovering \mathcal{F} . To summarize this discussion we have the following result:

Claim 2 *All fundamental matrices associated with translating moving bodies viewed from two fixed views live in a 3-dimensional linear subspace of R^9 represented by the step-3 extensor \mathcal{F} . Given that each body i is a planar object with a known homography matrix H_i , then the contribution of each H_i is captured by a step-3 extensor \mathcal{F}_i which satisfies $\mathcal{F} \vee \mathcal{F}_i = 0$, i.e., \mathcal{F} is a common transversal on all \mathcal{F}_i . The vanishing join equation $\mathcal{F} \vee \mathcal{F}_i = 0$, $i = 0, 1, 2, \dots, k$, contributes $20 - i$ linearly independent constraints on \mathcal{F} , thus 5 planar bodies are sufficient to uniquely define a solution for the 84 coefficients of \mathcal{F} .*

3.1 Recovering the Fundamental Matrices F_i from \mathcal{F}

We have shown that 5 moving planes are sufficient for uniquely (and linearly) recovering the step-3 extensor \mathcal{F} which all fundamental matrices (of all moving bodies) live in. We have recovered \mathcal{F}_i (from the known homography matrix H_i) and we have now \mathcal{F} . We wish to recover next the fundamental matrices F_i associated with the moving planes.

Let $f_1, f_2, f_3 \in R^9$ be some basis of the 3-dimensional subspace represented by \mathcal{F} , i.e., $\mathcal{F} = f_1 f_2 f_3$. In order to find such a basis, consider again an application of Claim 1 as follows. Recall that a point $P \in \text{span}\{f_1, f_2, f_3\}$ iff $\mathcal{F} \vee P = 0$. Let $C = \mathcal{F} \vee P$,

$$C = \sum_{1 \leq k_1 < \dots < k_4 \leq 9} C_{k_1, \dots, k_4} e_{k_1} \vee \dots \vee e_{k_4},$$

where from Claim 1 the coefficients of C can be described as a linear combination of the (known) coefficients of \mathcal{F} and the (unknown) entries of P :

$$C_{k_1, \dots, k_4} = \sum_{\sigma \in S_4, \sigma_1 < \sigma_2 < \sigma_3} \text{sgn}(\sigma) \mathcal{F}_{k_{\sigma_1}, k_{\sigma_2}, k_{\sigma_3}} P_{k_{\sigma_4}}$$

This provides us with $\binom{9}{4} = 126$ linear constraints on the 9 entries of P . The measurement matrix M of this system must be of rank 6 because \mathcal{F} is a step-3 extensor, hence the three eigenvectors associated with the vanishing eigenvalues of $M^T M$ provide us with a basis f_1, f_2, f_3 .

We are given the basis f_1, f_2, f_3 and from each H_i we have the null vectors f_1^i, f_2^i, f_3^i . The vector \bar{F}_i , representing

the fundamental matrix F_i , is in the span of both sets of vectors, i.e., there exist coefficients α_j^i, β_j^i (up to scale for each i) which satisfy:

$$\alpha_1^i f_1^i + \alpha_2^i f_2^i + \alpha_3^i f_3^i = \beta_1^i f_1 + \beta_2^i f_2 + \beta_3^i f_3 \quad (3)$$

which provides a system of linear equations for those coefficients (per i). The existence and uniqueness of the solution is guaranteed since we know that F_i is unique and is in the span of both sets. Once those coefficients have been recovered, then $\bar{F}_i = \beta_1^i f_1 + \beta_2^i f_2 + \beta_3^i f_3$.

3.2 Recovering H_∞

Given that the family of fundamental matrices F_i associated with bodies moving in pure translation is of the form $F_i = [t_i]_\times H_\infty$, one can easily recover H_∞ and in turn obtain an affine calibration of the camera geometry.

Note that the family $H_i^\pi = \lambda H_\infty + t_i n^\top$ satisfies the constraint that $F_i^\top H_i^\pi$ is a skew-symmetric matrix for all choices of the scalar λ and the vector n . Therefore, given that H_∞ is of full rank (which is a valid assumption for perspective cameras since $H_\infty = K' R K^{-1}$) the family of homography matrices which satisfy this constraint for all i are $H_i^\pi \cong H_\infty$.

The constraint that $F_i^\top H_\infty$ is skew-symmetric provides 6 linear equations on H_∞ , per index i , however only 5 of which are linearly independent. To see why this is so, recall that [13] have first noted that the family of homography matrices H_π over all choices of planes π (including $\pi = \infty$) live in a 4-dimensional linear subspace of R^9 . Therefore, the collection of matrices H which satisfy $F_i = [t_i]_\times H$ live in a 4-dimensional space. Since each of the linear constraints $F_i^\top H + H^\top F_i = 0$ is an element of the 5-dimensional null space over the set of all matrices H , there could be at most 5 linearly independent constraints from $F_i^\top H + H^\top F_i = 0$.

Thus, to conclude, two fundamental matrices (of two bodies) are sufficient to uniquely constrain H_∞ .

4 Collinear Motion and Miscellaneous Items

So far we considered the general case in which the relative translations among the various bodies spans a full rank space. Consider the case where all the motions are along the same direction, i.e., $F_i = [\alpha_i v' + \beta_i t]_\times H_\infty$ where t is the fixed direction, β_i is the magnitude of translation of the i 'th body, v' is the epipolar point (projection of the first camera center onto the image plane of the second camera) and α_i accounts for the global scale factor. Therefore, the family of such fundamental matrices live in a 2-dimensional subspace, i.e., \mathcal{F} is a step-2 extensor (instead of step-3).

Our calculations are now different. We are looking for a common step-2 transversal on step-3 extensors. The join

$\mathcal{F} \vee \mathcal{F}_i$ has $\binom{9}{5} = 126$ vanishing coefficients (linear constraints on the $\binom{9}{2} = 36$ coefficients comprising \mathcal{F}). Among the 126 constraints only $\binom{9-3}{2} = 15$ are linearly independent. Each additional extensor $\mathcal{F}_i, i = 1, 2, \dots, k$ adds only $15 - 3i$ independent equations. Thus, three moving planes are sufficient ($15 + 12 + 9 = 36$) for a unique solution for \mathcal{F} . We summarize all this in the following claim:

Claim 3 *All fundamental matrices associated with translating moving bodies along a fixed direction viewed from two fixed views live in a 2-dimensional linear subspace of R^9 represented by the step-2 extensor \mathcal{F} . Given that each body i is a planar object with a known homography matrix H_i , then the contribution of each H_i is captured by a step-3 extensor \mathcal{F}_i which satisfies $\mathcal{F} \vee \mathcal{F}_i = 0$, i.e., \mathcal{F} is a common transversal on all \mathcal{F}_i . The vanishing join equation $\mathcal{F} \vee \mathcal{F}_i = 0, i = 0, 1, 2, \dots, k$, contributes $15 - 3i$ linearly independent constraints on \mathcal{F} , thus 3 planar bodies are sufficient to uniquely define a solution for the 36 coefficients of \mathcal{F} .*

4.1 Parallel Planes

Going back to the general case of translations filling up a full rank space, the calculation also changes when planes are parallel to each other. Consider homography matrices H_1, H_2 induced by two parallel planes. The step-3 extensor \mathcal{F}_1 contributes 20 independent equations for \mathcal{F} . However, $\mathcal{F}_1 \vee \mathcal{F}_2$ is a step-5 extensor (instead of 6), and thus \mathcal{F}_2 would contribute only 16 new constraints (instead of 19). To see why this is so, consider H_1, H_2 two homography matrices induced by two distinct but parallel planes, i.e., $H_1 \cong H_\infty + t_1 n^t$ and $H_2 \cong H_\infty + t_2 n^t$. One can verify (by substitution) that $F = [t]_\times H_1 \cong [t]_\times H_2$, where $t = t_1 - t_2$, which means that $F^\top H_1$ and $F^\top H_2$ are skew-symmetric. Thus, the solution space for \mathcal{F}_1 (given H_1) and the solution space for \mathcal{F}_2 (given H_2) intersect at F , which in turn means that $\mathcal{F}_1 \vee \mathcal{F}_2$ is a step-5 extensor. Therefore, among the 20 constraints contributed from \mathcal{F}_2 , $\binom{9-5}{3} = 4$ of them are in common with the previous set of 20 made by \mathcal{F}_1 . Thus, we will need much more than 5 planes in order to obtain a sufficient number of constraints to uniquely solve for \mathcal{F} . Nevertheless, there is an alternative way to handle this situation (which requires only 4 translating planes) but due to space limitations we will not introduce here.

4.2 Non-linear Constraints

The final issue we address here is the non-linear constraints we so far ignored. There are two kinds of non-linearities. The first kind is associated with the fact that not every vector of 84 coefficients is an admissible step-3 extensor. Let A be a step-3 extensor, using Claim 1 on

$A \vee A = 0$, the $\binom{9}{6} = 84$ coefficients of the join are bilinear products of the coefficients of A with itself — thus comprising of 84 second order constraints on A . One can make use of these second order constraints to reduce the number of intersections required for a solution for the common transversal (up to a finite-fold ambiguity) from 5 to some smaller number. For example, 4 intersections provide $20 + 19 + 18 + 17 = 74$ linear constraints on \mathcal{F} leaving 9 parameters to be determined (linear combination of the null space of the estimation matrix) using the second order constraints on \mathcal{F} (provided there is a sufficient number of algebraically independent constraints among them).

The second kind of non-linearity arises from the fact that each fundamental matrix F_i is a rank 2 matrix. This type of non-linearity appears in two places. First is when we create the step-3 extensors \mathcal{F}_i from the null space of the 6 constraints $H_i^\top F_i + F_i^\top H_i = 0$ on F_i . Given that the rank of F_i is 2, we have an additional cubic constraint (arising from the vanishing determinant of F_i). This implies that the possible F_i do not live in a 3-dimensional subspace but in an algebraic variety — a fact that complicates considerably the process of finding a common transversal. Second is when we recover F_i from eqn. 3 we have an additional cubic constraint which can be used to further constrain the system of equations using a Levenberg-Marquardt type of iterative algorithm.

In this paper we chose to ignore the non-linear constraints and instead obtain a simple and manageable algorithm. Ultimately it is a matter of a tradeoff between the possibility of achieving higher accuracies in the presence of noise and the resulting added complexity of doing so. The experiments shown next indicate that one can obtain reasonable estimation accuracy using the linear constraints alone.

5 Experiments

In the experiments below we tested the reconstruction of the fundamental matrices and the affine calibration under general translation and translations along a fixed direction.

In the first experiment, shown in Fig. 1, four 3D objects (with planar parts) are in translational (general) motion and the fifth object is taken from the static background (the table). Fig. 1a,b displays the two views of the dynamic configuration. The homography matrices H_1, \dots, H_5 are recovered using the matching points displayed in Fig. 1c. The fundamental matrices F_1, \dots, F_5 were recovered using the algorithm presented in this paper. Fig. 1d displays marked points p on one of the objects and Fig. 1e shows the epipolar lines Fp where F is the corresponding fundamental matrix. Note that the epipolar lines pass through the matching points at a sub-pixel accuracy.

The affine calibration was constructed from the recovered H_∞ and its accuracy was estimated as follows. A

line drawn on one of the planar objects (a book) was reconstructed in 3D using the recovered calibration data - denote that line as L . Then matching points along parallel lines were marked in both images and reconstructed in 3D. For each reconstructed point a 3D line parallel to L was created and back-projected to the image. If the calibration is projective the back-projected lines should not necessarily be parallel in the image, but when the calibration is affine those lines should be parallel (on the image of the planar object). Fig. 1f displays the back-projected lines which are indeed parallel over the extent of the book (the planar object).

In the second experiment, shown in Fig. 2, the multi-body configuration moves along a fixed direction. In this particular scene the bodies consist of the person, the chair, and the static floor. Fig. 2a,b displays two views of this multi-body configuration. The homography matrices H_1, H_2, H_3 were recovered from point matches — on the person taken from the chest (approximately planar), on the chair taken from box, and from points on the floor. The three fundamental matrices were recovered, and that recovered from the floor was tested on points from the static part of the scene as shown in Fig. 2c,d. Note that the points being tested are taken from regions which are far away from the floor and thus are more susceptible to estimation error of the fundamental matrix — yet the epipolar lines pass through the matching points at sub-pixel accuracy.

6 Summary

We have shown that a configuration of multiple planes moving relatively to each other in pure translation conveys additional information (beyond the homography matrices) which can be used to recover the fundamental matrices, one per object, and in turn recover the affine calibration between the two cameras. The technique for doing so was based on the observation that all fundamental matrices live in a 3-dimensional subspace, which when represented as an extensor is the common transversal on the extensors defined by the homography matrices. We have shown that generally 5 intersections are needed for a linear solution, and when the multi-body motion is along a fixed direction then 3 intersections suffice. The affine calibration readily follows once the fundamental matrices are recovered because the homography induced by the plane at infinity is the only homography matrix shared by all the fundamental matrices, thus can be extracted linearly from at least two bodies.

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] M. Barnabei, A. Brini, and G.C. Rota. On the exterior calculus of invariant theory. *J. of Alg.*, 96:120–160, 1985.

[3] S. Carlsson The Double Algebra: An effective Tool for Computing Invariants in Computer Vision In *Applications of Invariance in Computer Vision*, Joseph L.Mundy, Andrew Zisserman, David Forsyth (Eds.), Springer-Verlag Berlin Heidelberg 1994.

[4] J. Costeira and T. Kanade. A Multibody Factorization Method for Independent Moving Objects. *1998 International Journal on Computer Vision*, Kluwer, Vol. 29, No. 3, September, 1998.

[5] O.D. Faugeras. Stratification of three-dimensional vision: projective, affine and metric representations. *Journal of the Optical Society of America*, 12(3):465–484, 1995.

[6] O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, June 1995.

[7] O.D. Faugeras and T. Papadopoulos. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors *INRIA Rapport de recherche no.3225 - july 1997*

[8] 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.

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

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

[11] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings IJCAI*, pages 674–679, Vancouver, Canada, 1981.

[12] 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.

[13] A. Shashua and S. Avidan. The rank4 constraint in multiple view geometry. In *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, April 1996.

[14] 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.

[15] B. Sturmfels. *Algorithms in Invariant Theory* Springer-Verlag Wien New York, 1993.

[16] 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.

[17] Lior Wolf and A. Shashua. On Projection Matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, \dots, 6$, and their Applications in Computer Vision. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.

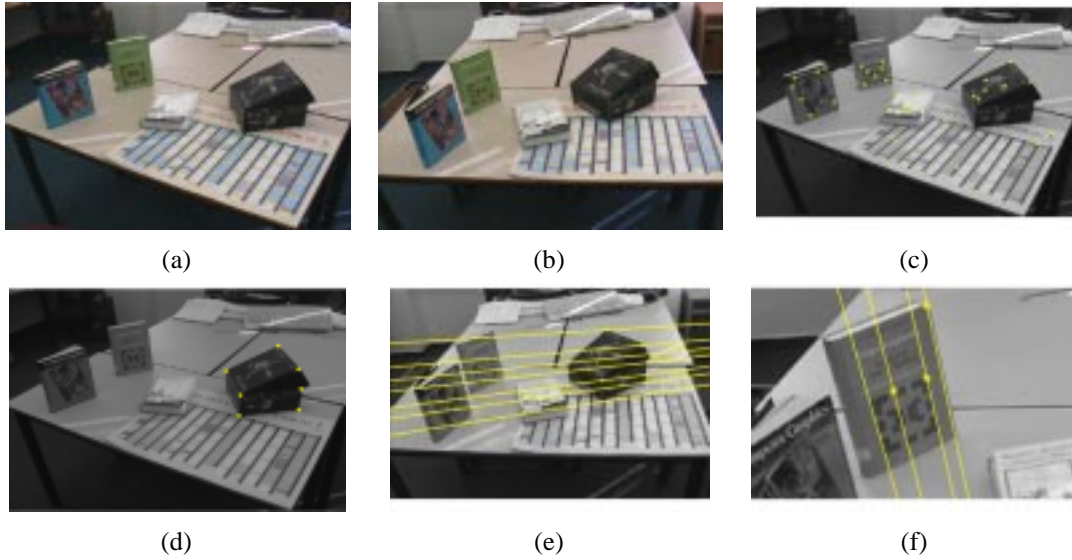


Figure 1. (a),(b) two views of a multi-body scene under general translation. (c) Points used to find the five homography matrices H_1, \dots, H_5 . (d),(e) Some corners pointed on one of the objects, and the corresponding epipolar lines on the other object. (f) Back-projection of parallel lines for testing accuracy of affine calibration — see text for details.

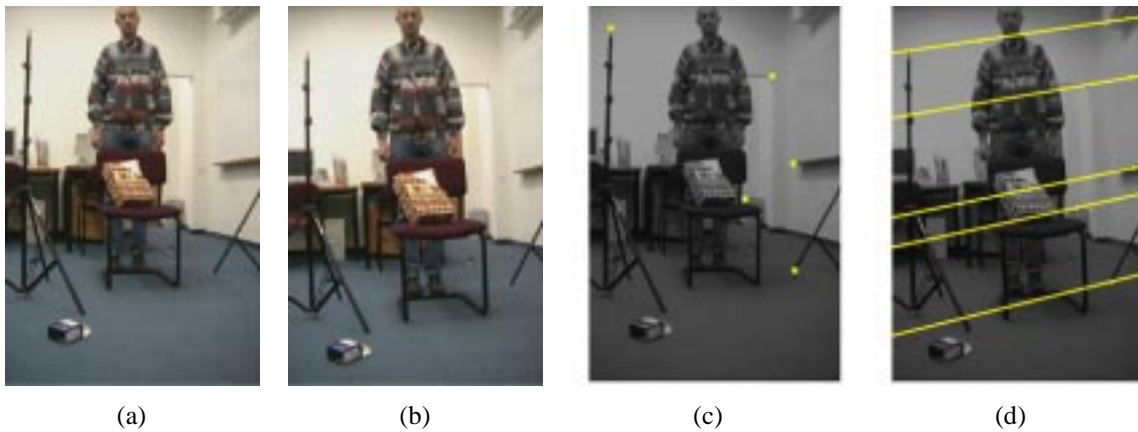


Figure 2. (a),(b) two views of a multi-body scene where the objects move along a fixed direction. Homography matrices are estimated one from the person's chest, one from the object on the chair, and one from the floor. (c),(d) Some "static" points and their corresponding epipolar lines.