On Projection Matrices $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 3, \cdots, 6$, and their Applications in Computer Vision

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

Projection matrices from projective spaces $\mathcal{P}^3$ to $\mathcal{P}^2$ have long been used in multiple-view geometry to model the perspective projection created by the pin-hole camera. In this work we introduce higher-dimensional mappings $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 3, 4, 5, 6$ for the representation of various applications in which the world we view is no longer rigid. We also describe the multi-view constraints from these new projection matrices (where $k > 3$) and methods for extracting the (non-rigid) structure and motion for each application.

1 Introduction

The projective camera model, represented by the mapping between projective spaces $\mathcal{P}^3 \rightarrow \mathcal{P}^2$, has long been used to model the perspective projection of the pin-hole camera in Structure from Motion (SFM) applications in computer vision. These applications include photogrammetry, ego-motion estimation, feature alignment for visual recognition, and view-synthesis for graphics rendering. There is a large body of literature on the projective camera model in a multi-view setting with the resulting multi-linear tensors as the primitive building-blocks of 3D computer vision. A summary of the past decade of work in this area with a detailed exposition of the multi-linear maps with their associated tensors (bifocal, trifocal and quadrifocal) can be found in [8] and earlier work in [4].

The literature mentioned above is mostly relevant to a static scene, i.e., a rigid body viewed by an uncalibrated camera. Recently, however, a new body of work has appeared [1, 12, 10, 13, 7] which assumes a configuration of points in which every single point in the configuration can move independently along some arbitrary trajectory (straight line path and in some cases second-order) while the camera is undergoing general motion (in 3D projective space). For brevity, we will refer to such a scene as *dynamic* whereas the conventional rigid body configuration would be referred to as *static*. Dynamic configurations, for example, include as a particular case multi-body motion, i.e., when each body contains multiple points rigidly attached to the same coordinate system [3, 6]

In this paper we address the geometry of multiple views of dynamic scenes from the point of view of *lifting* the problem to a static scene embedded in a higher dimensional space. In other words, we investigate camera projection matrices of $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 3, 4, 5, 6$ for modeling a static body in k-dimensional projective space $\mathcal{P}^k$ projected onto the image space $\mathcal{P}^2$. These projection matrices model dynamic situations in 2D and 3D. We will consider, for example, three different applications of $\mathcal{P}^4 \longrightarrow \mathcal{P}^2$ which include (i) multiple linearly moving coplanar points under constant velocity, (ii) 3D points moving in constant velocity along a common single direction, and (iii) Two-body segmentation in 3D — the resulting tensor is referred to as the 3D *segmentation* tensor ($\mathcal{P}^3 \longrightarrow \mathcal{P}^2$ models a 2D segmentation problem). Projection matrix $\mathcal{P}^5 \longrightarrow \mathcal{P}^2$ is shown to model moving 3D points under constant velocity and coplanar trajectories (all straight line paths are on a plane). Projection matrix $\mathcal{P}^6 \longrightarrow \mathcal{P}^2$ is shown to model the general constant velocity multiple linearly moving points in 3D. The latter was derived in the past by [7] for orthographic cameras while here we take this further and address the problem in the general perspective pin-hole (projective) setting.

Following the introduction of $\mathcal{P}^k \longrightarrow \mathcal{P}^2$ and their role in dynamic SFM, we describe the construction of tensors from multi-view relations of each model and the process for recovering the camera motion parameters (the physical cameras) and the 3D structure of the scene.

2 Applications of $\mathcal{P}^k \rightarrow \mathcal{P}^2$

We will describe below a number of different applications for values of $k = 3, 4, 5, 6$. These applications include multi-body segmentation (we call "segmentation tensors") and multiple linearly moving points.

2.1 Applications for $\mathcal{P}^3 \rightarrow \mathcal{P}^2$

The family of 3×4 matrices have been extensively studied in the context of SFM. These matrices model the (uncalibrated) pin-hole camera viewing a rigid configuration of points, i.e., a *static* 2D from 3D scenario. We present an additional instantiation of $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ in the context of "2D segmentation" defined below:

Problem Definition 1 (2D segmentation) *We are given 2D general views of a planar point configuration consisting of two bodies moving relatively to each other by pure translation. Describe algebraic constraints necessary for segmenting the two bodies from image measurements.*

Clearly, 4 point matches per body (8 points in total) uniquely determine the 2D homography between the two views of the plane, thus a segmentation can be achieved by searching over all quadruples of matching points until a consistent set is found (i.e., the resulting homography agrees on a sufficiently large subset of points). This approach is general and will work even when the relative motion between the two bodies is full projective.

We show that on this kind of problem, where the relative motion between the two bodies is pure translation, we can do better. We will first use 8 unsegmented point matches after which we will need only 3 segmented point matches (i.e. search over triplets of matching points). The formulation of the problem is described next.

Let A, B be the (unknown) homography matrices from the world plane to views 1,2 respectively. Let s be a point on the first body. The image of s in the first view is $p \cong As$ and in the second view $p' \cong Bs$. The image of a point r on the second body would be $p \cong Ar$ in the first view, and

$$p' \cong Br + B \begin{bmatrix} dx \\ dy \\ 0 \end{bmatrix}$$

on the second view, where $t = (dx, dy, 0)$ is the fixed (unknown) translational motion between the two bodies.

To formulate this as a $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ problem we "lift" s and r to 3D space by defining $P_s \cong (s, 0)^\top$ for point s and $P_r \cong (r, 1)^\top$ for point r on the second body. Define the following projection matrices:

$$\begin{aligned} M_1 &\cong [A \quad 0_{3\times1}] \\ M_2 &\cong [B \quad Bt] \end{aligned}$$

Therefore, M_1, M_2 apply to both bodies in a uniform manner without the need for prior segmentation. Since we have formulated the 2D segmentation problem in the domain of $\mathcal{P}^3 \rightarrow \mathcal{P}^2$, then all the body of work on static SFM from two views (and more than 2 views) apply here. For example, a "fundamental" matrix F can be computed from 8 (unsegmented) points, i.e., $p'^\top Fp = 0$ for all matching points regardless of which body they come from. The image of F, i.e., Fp, is a line in the second view which passes through the two possible images of the point. The null vector of $F^\top$ is the point Bt. Each body is represented as a plane in $\mathcal{P}^3$, thus having 3 segmented points would allow us to fix the plane and in turn segment the scene.

2.2 Applications for $\mathcal{P}^4 \rightarrow \mathcal{P}^2$

We introduce three different instantiations of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ in the context of dynamic SFM. First application would be three views of multiple linearly moving coplanar points under constant velocity, second is constant velocity multiple linearly moving points in 3D where all trajectories are parallel to each other, and third is the 3D segmentation tensor.

Problem Definition 2 (Coplanar Dynamic Scene) *We are given views of a planar configuration of points where each point may move independently along some straight-line path with a constant velocity motion. Describe the algebraic constraints necessary for reconstruction of camera motion (homography matrices), static versus dynamic segmentation, and reconstruction of point velocities.*

The problem above is a particular case of a more general problem (same as above but without the constant velocity constraint) addressed by [12]. The algebraic constraints there were in the form of a $3 \times 3 \times 3$ tensor called "Htensor" which requires 26 triplets of point-matches for a solution. We will show next that the constant-velocity assumption reduces the requirements considerably to 13 triplets of point-matches, not to mention that Htensor becomes degenerate for constant-velocity. The key is a $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ problem formulation as follows.

Let H_j, $j = 0, 1, 2$ denote the homography from world plane to the j'th view onto the image points $p_j = (x_j, y_j, 1)^\top$. Let $(X, Y, 1)$ be the coordinates of the world point projecting onto p_j. Note that since the reconstruction is up to a 3D Affine ambiguity (because of the constant velocity assumption), then we are allowed to fix the third coordinate of the world plane to 1. Let dX, dY be the direction of the constant-velocity motion of the point $(X, Y, 1)^\top$. Let H_j^* denote the left 3×2 sub-matrix of H_j. We have the following relation:

$$p_j \cong H_j \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} + jH_j \begin{pmatrix} dX \\ dY \\ 0 \end{pmatrix} = \tilde{H}_j \begin{pmatrix} X \\ Y \\ 1 \\ dX \\ dY \end{pmatrix}$$

where $\tilde{H}_j$ is a 3×5 matrix $[H_j, jH_j^*]$. We have therefore a $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ formalism $p_j \cong \tilde{H}_j P$ where $P \in \mathcal{P}^4$. The geometry of such projections is described in more detail in section 3 and as an example, the center for projection is no longer a point but an extensor of step 2, i.e., a line.

Let $s_j = (1, 0, -x_j)$ and $r_j = (0, 1, -y_j)$. Let l_2 be any line such that $l_2^\top p_2 = 0$. Then, $0 = s_j^\top p_j = s_j^\top \tilde{H}_j P$, $0 = l_2^\top \tilde{H}_2 P$. Therefore, two points and a line provide a constraint as follows:

$$det \left(\left[\begin{array}{c} s_0^\top \tilde{H}_0 \\ r_0^\top \tilde{H}_0 \\ s_1^\top \tilde{H}_1 \\ r_1^\top \tilde{H}_1 \\ l_2^\top \tilde{H}_2 \end{array} \right] \right) = 0$$

The determinant expansion provides a multilinear constraint with a $3 \times 3 \times 3$ tensor described next. It will be useful to switch notation: let p, p', p'' replace p_0, p_1, p_2 respectively, and likewise let s, s', s'' and r, r', r'' replace s_j, r_j, $j = 0, 1, 2$, respectively. The multilinear constraint is expressed as follows:

$$p^i p'^j s''_k \mathcal{A}_{ij}^k = 0,$$

where the index notations follow the covariant-contravariant tensorial convention, i.e., $p^i s_i$ stands for the scalar product $p^\top s$ and superscripts represent points and subscripts represent lines. The entries of the tensor $\mathcal{A}_{ij}^k$ is a multilinear function of the entries of $\tilde{H}_j$. The constraint itself is a point-point-line constraint, thus a triplet p, p', p'' provides two linear constraints $p^i p'^j s''_k \mathcal{A}_{ij}^k = 0$ and $p^i p'^j r''_k \mathcal{A}_{ij}^k = 0$ on the entries of $\mathcal{A}_{ij}^k$. Therefore, 13 matching triplets are sufficient for a solution (compared to 26 triplets for the Htensor of [12]). Further details on the properties of $\mathcal{A}_{ij}^k$, how to extract the homographies up to an Affine transformation, segment static from non-static points, and how to reconstruct structure and motion are found in section 3.

Problem Definition 3 (3D Dynamic Scene, Collinear Motion) *We are given (general) views of a 3D configuration of points where each point may move independently along some straight-line path with a constant velocity motion. All the line trajectories are along the same direction (parallel to each other). Describe the algebraic constraints necessary for reconstruction of camera motion (3×4 projection matrices), static versus dynamic segmentation, and reconstruction of point velocities.*

Let $P_i = (X_i, Y_i, Z_i, 1)^\top$, $i = 1, ..., n$, be a configuration of points in 3D (Affine space) moving along a fixed direction $dP = (dX, dY, dZ, 0)^\top$ such that at time $j = 0, ..., m$ the position of each point is $P_i + j\lambda_i dP$. Let M_j denote the j'th 3×4 camera matrix, and let p_{ij} denote the projection of P_i on view j:

$$p_{ij} \cong M_j(P_i + j\lambda_i dP) = [\, M_j \quad jM_j dP \,] \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \\ \lambda_i \end{pmatrix},$$

which is again a $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ problem formulation. Further details can be found in the section 3.

Problem Definition 4 (3D Segmentation) *We are given three general views of a 3D point configuration consisting of two bodies moving relatively to each other by pure translation. Describe algebraic constraints necessary for segmenting the two bodies from image measurements.*

Clearly, one can approach this problem using trifocal tensors. The motion of each body is captured by a trifocal tensor which requires 7 points (or 6 points for a non-linear solution up to a 3-fold ambiguity). Thus, a segmentation can be achieved by searching over all 6-tuples (or 7-tuples) of matching points until a consistent set is found. This approach is general and applies even when the relative motion between the two bodies is full projective.

Just like in the 2D Segmentation problem, since the relative motion between the two bodies is pure translation, we can do better. In fact we need to search over all quadruples of points instead of 6-tuples. The key is the $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ problem formulation which allows us to describe a multilinear constraint common to both bodies — as described next.

Let $P \in \mathcal{P}^3$ be a point in 3D. If P is on the first body, then a set of camera matrices M_j^1, $j = 0, 1, 2$, provide the image points $p_j \cong M_j^1 P$. Likewise, if P is on the second body then $p_j \cong M_j^2 P$. Because the relative motion between the two bodies consists of pure translation the homography A_∞^j due to the plane at infinity is the same for the j'th camera matrix of both bodies:

$$M_j^1 \cong [\, A_\infty^j v_j^1 \,] \quad M_j^2 \cong [\, A_\infty^j v_j^2 \,] .$$

We "lift" P onto $\mathcal{P}^4$ by defining $\tilde{P}$ as follows. If P belongs to the first body, then $\tilde{P} \cong (\, P_1 \quad P_2 \quad P_3 \quad P_4 \quad 0 \,)^T$. If P belongs to the second body, then $\tilde{P} \cong (\, P_1 \quad P_2 \quad P_3 \quad 0 \quad P_4 \,)^T$. The $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ projection matrix would then be:

$$M_j \cong [\, A_\infty^j v_j^1 v_j^2 \,] .$$

The resulting $3 \times 3 \times 3$ tensor would be derived exactly as above and would require 13 (unsegmented) points for a linear solution. Each body is represented by an extensor of step 4 in $\mathcal{P}^4$, thus 4 (segmented) point matches are required to solve for the extensor. Therefore, once the tensor is found, 4 segmented points are required to provide a segmentation of the entire point configuration.

2.3 Applications for $\mathcal{P}^5 \rightarrow \mathcal{P}^2$

There are a number of instantiations of $\mathcal{P}^5 \rightarrow \mathcal{P}^2$. The first is the projection from 3D lines represented by Plücker coordinates to 2D lines [5]: $l \cong \tilde{M}L$ where the three rows of $\tilde{M}$ are the result of the "meet" [2] operation of pairs of rows of the original 3×4 camera projection matrix, i.e., each row of $\tilde{M}$ represents the line of intersection of the two planes represented by the corresponding rows of M.

The resulting multi-view tensors in the straight-forward sense represent the "trajectory triangulation" introduced in [1] which models the application of a moving point P along a straight line L such that in the j'th view we observe the projection of p_j of P. Thus, $p_j^\top \tilde{M}L = 0$ for all views of P. In the situation of trajectory triangulation, in each view we have an image P_i of a point which lies on the line in 3D. So $p_i^T M_i L \cong p_i^T l_i = 0$. The determinant of the 6×6 matrix whose rows are $p_j^\top \tilde{M}$ must vanish. The resulting tensor is 3^6 and thus would require 728 matching points across 6 views in order to obtain a linear solution. Naturally, this situation is unwieldy application-wise.

A more tractable tensor (in terms of size) would arise from adding two more assumptions (i) the motion of the point is with constant velocity, and (ii) all the line trajectories are coplanar. We have the following problem definition:

Problem Definition 5 (3D Dynamic Scene, Coplanar Motion) *We are given (general) views of a 3D configuration of points where each point may move independently along some straight-line path with a constant velocity motion. All the line trajectories are coplanar. Describe the algebraic constraints of this situation.*

Following the derivation of Problem 3, the j'th projection matrix $\tilde{M}_j$ has the form $[M_j, jM_jdP_1, jM_jdP_2]$ where M_j is the corresponding 3×4 camera matrix and dP_1, dP_2 span the 2D plane of trajectories. The points in $\mathcal{P}^5$ have the form $P_i = (X_i, Y_i, Z_i, 1, \lambda_i, \mu_i)^\top$, thus $p_{ij} \cong \tilde{M}_j P_j$. The resulting tensorial relation follows from 3 views, as follows. For a triplet of matching points p, p', p'' denote the lines $s = (1, 0, -x)$ and $r = (0, 1, -y)$ coincident with p and likewise the lines s', r' and the lines s'', r''. Thus the two rows $s^\top \tilde{M}$, and $r^\top \tilde{M}$ per camera (and likewise with $\tilde{M}'$ and $\tilde{M}''$) form a 6×6 matrix with a vanishing determinant. The determinant expansion provides a multilinear constraint of p, p', p'' with a $3 \times 3 \times 3$ tensor $p^i p'^j p''^k \mathcal{E}_{ijk} = 0$. Therefore 26 matching triplets across 3 views are sufficient for a solution (compared to 728 points across 6 views).

Finally, we can make the following analogy between $\mathcal{P}^5 \rightarrow \mathcal{P}^2$ and planar dynamic scenes with general motion (no constant velocity assumption). The case of planar dynamic motion across three views was introduced in [12], where the constraint is based on the fact that if p, p', p'' are projections of a moving point P along some line on a fixed world plane, then $Hp, H'p', p''$ are collinear, where H, H' are homography matrices aligning images 1,2 onto image 3 (H, H' are uniquely defined as a function of the position of the three cameras and the position of the world plane on which the points P reside). We make the following claim: in the context of $\mathcal{P}^5 \rightarrow \mathcal{P}^2$, there exist two such homography matrices H, H' from images 1,2 onto image 3, such that the projections of points $P \in \mathcal{P}^5$ onto the three image planes produces a set of 3 collinear points.

Claim 1 (Dynamic Coplanar, General Motion) *Given three views p, p', p'' of a point configuration in $P \in \mathcal{P}^5$, there exist homographies H and H' such $Hp, H'p', p''$ are collinear.*

Proof: The key observation is that without loss of generality we can choose a projective coordinate system (in $\mathcal{P}^5$) such that the first two projection matrices are of the form $[\, A_{3\times3} \quad 0_{3\times3} \,]$, and $[\, 0_{3\times3} \quad B_{3\times3} \,]$. The third projection matrix will have some general form $[\, C_{3\times3} \quad D_{3\times3} \,]$. Let $H = CA^{-1}$ and $H' = DB^{-1}$ and let $P = (p_1, \cdots, p_6)$. Then, $Hp \cong C(p_1, p_2, p_3)^\top$ and $H'p' \cong D(p_4, p_5, p_6)^\top$, whereas $p'' \cong C(p_1, p_2, p_3)^\top + D(p_4, p_5, p_6)^\top$. $\square$

2.4 Applications for $\mathcal{P}^6 \rightarrow \mathcal{P}^2$

In this section we consider the most general constant velocity tensor - the tensor of constant velocity in 3D, where direction of motion is not restricted and the cameras are general 3×4 projective cameras.

Problem Definition 6 (3D Dynamic Scene) *We are given (general) views of a 3D configuration of points. Each point may move independently along some straight-line path with a constant velocity motion. Describe the algebraic constraints necessary for reconstruction of the points in 3D and their velocities.*

Let $P_i = (X_i, Y_i, Z_i, 1)^\top$, $i = 1, ..., n$, be a configuration of points in 3D (Affine space) moving along a direction $dP_i = (dX_i, dY_i, dZ_i, 0)^\top$ such that at time $j = 0, 1, 2, 3$ the position of each point is $P_i + jdP_i$. Let M_j denote the j'th 3×4 camera matrix, and M_j^* denote the left 3×3 sub-matrix of M_j. The projection p_{ij} of P_i on view j is described by $p_{ij} \cong \tilde{M}_j \tilde{P}_i$ where $\tilde{M}_j = [\, M_j \quad M_j^* \,]$ and $\tilde{P}_i = (X_i, Y_i, Z_i, 1, dX_i, dY_i, dZ_i)^\top$.

The resulting tensorial relation follows from 4 views, as follows. denote by $s_j = (1, 0, -x_j)^\top$ and $r_j = (0, 1, -y_j)^\top$ be lines coincident with the projections $p_j \cong (x_j, y_j, 1)^\top$ of a point $\tilde{P}$. We construct a 7×7 matrix with a vanishing determinant such that it's first 6 rows are $s_j^\top \tilde{M}_j$ and $r_j^\top \tilde{M}_j$, $j = 0, 1, 2$, and for the 7'th row $l'''^\top \tilde{M}_3$ where l''' is any line coincident with the projection p_3. The determinant expansion is a multilinear relations between the image points p_0, p_1, p_2, denoted now by p, p', p'' and the line l''' with a 3^4 tensor $\mathcal{B}^q_{ijk}$, i.e., $p^i p'^j p''^k l'''_q \mathcal{B}^q_{ijk} = 0$. Since we can take any line l''' coincident with the 4'th image points each quadruple of matching points provides 2 linear constraints on the tensor, hence 40 matching points across 4 views are sufficient to uniquely (up to scale) determine the tensor. The process for extracting the camera matrices M_j up to a 3D affinity is described in section 3.

2.5 Summary of Applications

So far, we have discussed multi-view constraints of scenes containing multiple linearly moving points. The constraints were derived by "lifting" the non-rigid 3D phenomena into a rigid configuration in a higher dimensional space of $\mathcal{P}^k$. We have presented 6 applications for various values of k ranging from 3 to 6. To summarize, the table below lists the various applications of $\mathcal{P}^k \rightarrow \mathcal{P}^2$ which were presented in the preceding sections.

$\mathcal{P}^k$	Tensor Name	Size	ref.
$\mathcal{P}^3$	2D segmentation tensor	3^2	2.1
$\mathcal{P}^4$	2D constant velocity tensor	3^3	2.2
$\mathcal{P}^4$	3D segmentation tensor	3^3	2.2
$\mathcal{P}^4$	3D constant collinear velocity	3^3	2.2
$\mathcal{P}^5$	3D constant coplanar velocity	3^3	2.3
$\mathcal{P}^6$	3D constant velocity tensor	3^4	2.4

The resulting tensors for each $\mathcal{P}^k \rightarrow \mathcal{P}^2$ were reasonable in terms of size (thus practical) where the largest tensor of size 3^4 requiring 40 matching quadruples across 4 views was for the general, constant velocity, 3D dynamic motion.

3 The Geometry of $\mathcal{P}^k \rightarrow \mathcal{P}^2$

We will derive the basic elements for describing and recovering the projective matrices of $\mathcal{P}^k \rightarrow \mathcal{P}^2$. These elements are analogous to the role homography matrices and epipoles play in the $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ setting) in $\mathcal{P}^k \rightarrow \mathcal{P}^2$ geometry. We will start with some general concepts that are common to all the constructions of $\mathcal{P}^k \rightarrow \mathcal{P}^2$ and then proceed to the detailed derivation of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ and $\mathcal{P}^6 \rightarrow \mathcal{P}^2$.

We use the term *extensor* (cf. [2]) to describe the linear space spanned by a collection of points. A point will be extensor of step 1, a line is an extensor of step 2, a plane is an extensor of step 3, and a hyper-plane is an extensor of step k in $\mathcal{P}^k$. In $\mathcal{P}^n$, the union (join) of extensors of step k_1 and step k_2, where $k_1 + k_2 \leq n + 1$ is an extensor of step $k_1 + k_2$. The intersection (meet) of extensors of step k_1 and k_2 is an extensor of step $k_1 + k_2 - (n + 1)$. Given these definitions, the following statements immediately follow:

- The *center of projection* (COP) of a $\mathcal{P}^k \rightarrow \mathcal{P}^2$ projection is an extensor of step $k - 2$. Recall that the center of projection is the null space of the $3 \times (k+1)$ projection matrix, i.e., the center of projection of $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ is a *point*, of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ is a *line* and of $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ is an extensor of step 4.

- The *line of sight* (image ray) joins the COP and a point (on the image plane). Thus, for $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ the line of sight is a line, for $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ the line of sight is plane (extensor of step 2+1), and for $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ it is an extensor of step 5.

- The intersection of two lines of sight (a "triangulation" as it is known in $\mathcal{P}^3 \rightarrow \mathcal{P}^2$) is the meet of two lines of sights. Thus, in $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ the intersection is either a point or is not defined (2+2-4=0), i.e., when the two lines are skew. In $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ the intersection always exists and is also a point (3+3-5), and in $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ the intersection is a plane (5+5-7). Note that simply from these counting arguments it is clear that in $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ two views of matching points provide constraints on the geometry of camera positions, yet two views in $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ do not provide any constraints (because image rays always intersect), thus one needs at least 3 views of matching points in order to obtain a constraint, and in $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ one would need at least 4 views for a constraint (two rays intersect at a plane, a plane and a ray intersect at a point $(3 + 5 - 7)$, thus three image rays always intersect).

- The "epipole" in $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ is defined as the intersection between the line joining two COPs and an image plane (thus, for a pair of views we have two epipoles, one on each image plane). Or, equivalently, if $\tilde{M}_i, \tilde{M}_j$ are the projection matrices, then $\tilde{M}_i null(\tilde{M}_j)$ is the epipole on view i. This definition extends to $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ where the join of the two COPs is an extensor of step 4 (each COP is an extensor of step 2) and its meet with an image plane is an extensor of step $4 + 3 - 5$, i.e., is a line. Thus, the epipoles of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ are *lines* on their respective image planes. This definition, however, does not extend to $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ where the join of two COPs (4+4) fills the entire space $\mathcal{P}^6$. We define instead a "joint epipole", to be described later.

3.1 The Geometry of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$

Recall from the preceding section that one needs at least three views of matching points in order to obtain a constraint (because two image rays always intersect in $\mathcal{P}^4 \rightarrow \mathcal{P}^2$). We also noted in Problem 2 that the multi-linear constraint across three views takes the form of a $3 \times 3 \times 3$ tensor $\mathcal{A}^k_{ij}$ which is contracted by two points and a line. In other words, let p, p', p'' be three matching points along views 1,2,3 and let s'', r'' be any two lines coincident with

p''. The multilinear constraint is expressed as follows:

$$p^i p'^j s''_k \mathcal{A}^k_{ij} = 0,$$

where the index notations follow the covariant-contravariant tensorial convention, i.e., $p^i s_i$ stands for the scalar product $p^\top s$ and superscripts represent points and subscripts represent lines. The entries of the tensor $\mathcal{A}^k_{ij}$ is a multilinear function of the entries of the three projection matrices $\tilde{M}$, $\tilde{M}'$ and $\tilde{M}''$. The constraint itself is a point-point-line constraint, thus a triplet p, p', p'' provides two linear constraints $p^i p'^j s''_k \mathcal{A}^k_{ij} = 0$ and $p^i p'^j r''_k \mathcal{A}^k_{ij} = 0$ on the entries of $\mathcal{A}^k_{ij}$. Therefore, 13 matching triplets are sufficient for a (linear) solution. We will assume from now on that the tensor $\mathcal{A}^k_{ij}$ is given (i.e., recovered from image measurements) and we wish to recover the 3×5 projection matrices $\tilde{M}, \tilde{M}', \tilde{M}''$.

We begin by deriving certain useful properties of the tensor slices from which we could then recover the basic elements (epipoles, homography matrices) of the projection elements.

Claim 2 (point transfer)

$$p^i p'^j \mathcal{A}^k_{ij} \cong p''^k \tag{1}$$

Proof: Follows from the fact that $p^i p'^j s''_k \mathcal{A}^k_{ij} = 0$ for any line s'' coincident with p''. From the covariant-contravariant structure of the tensor, $p^i p'^j \mathcal{A}^k_{ij}$ is a point (contravariant vector), let this point be denoted by q^k. Hence, $q^k s''_k = 0$ for all lines s'' that satisfy $s''_k p''^k = 0$. Thus q and p'' are the same. □

Note that the rays associated with p, p' are extensors of step 3, i.e., a plane. The intersection of those rays is a point (as explained in the preceding section), and thus $p^i p'^j \mathcal{A}^k_{ij}$ is the back-projection onto view 3 (projection of a point is a point). Similarly, let l'' be some line in image 3 (extensor of step 2), thus the image ray associated with a point p' in image 2 and the extensor of step 4 associated with the join of l'' and the COP of camera 3 meet at a line ($3+4-5=2$) and let the projection of this line onto image 1 be denoted by l. The relationship between p', l'', l is captured by the tensor: $p'^j l''_k \mathcal{A}^k_{ij} \cong l_i$.

Claim 3 (homography slice) *Let δ^j be any contravariant vector. The 3×3 matrix $\delta^j \mathcal{A}^k_{ij}$ is a homography matrix (2D collineation) from views 1 to 3 induced by the plane defined by the join of the COP of the second projection matrix and the image point δ in view 2 (i.e., the image ray corresponding to δ).*

Proof: Consider $(\delta^j \mathcal{A}^k_{ij}) p^i = q^k$, from the point transfer equation 1 we have that q is the projection onto view 3 of the intersections of the two planes corresponding to the line of sight p and line of sight δ (recall that each line of sight is a plane in $\mathcal{P}^4$ and that two planes generally intersect at as point). Let π_δ denote the plane associated with the line of sight δ. If we fix δ and vary the point p over image 1, then the resulting points q are projection of points on the plane π_δ onto image 3. Thus the matrix $\delta^j \mathcal{A}^k_{ij}$ is projective transformation from image 1 to image 3 induced by the plane π_δ. □

Note that $\delta^j \mathcal{A}^k_{ij}$ is a linear combination of the three slices $\mathcal{A}^k_{i1}, \mathcal{A}^k_{i2}$ and $\mathcal{A}^k_{i3}$. Thus, in particular a slice (through the "j" index) produces a homography matrix. Likewise, $\delta^i \mathcal{A}^k_{ij}$ is a homography matrix from image 2 to image 3 induced by the plane associates with the image ray of the point δ in image 1.

Now that we have the means to generate homography matrices from the tensor, we are ready to describe the recovery of the epipoles. Let the (unknown) projection matrices be denoted by $\tilde{M}_1$, $\tilde{M}_2$ and $\tilde{M}_3$. Let $e_{ij} = \tilde{M}_i null(\tilde{M}_j)$ be the epipole (a line) as the projection of COP j onto view i.

Claim 4 (epipoles) *Let H_{ij}, G_{ij} be two (full-rank) homography matrices from view i to view j induced by two distinct (but arbitrary) planes. The epipole e_{ji} is one of the generalized eigenvectors of H^T_{ij}, G^T_{ij}, i.e., satisfies the equation:*

$$(H^\top_{ij} + \lambda G^\top_{ij}) e_{ji} = 0.$$

Proof: Let H_{ij} be any (full-rank) homography matrix from view i to view j. Thus, H^{-T}_{ij} maps lines (dual space) from view i to view j. Because epipoles are lines in $\mathcal{P}^4 \to \mathcal{P}^2$ geometry, we have $H^{-T}_{ij} e_{ij} \cong e_{ji}$ and conversely $H^\top_{ij} e_{ji} \cong e_{ij}$. Thus, given two such homography matrices, there exists a scalar λ such that $(H^\top_{ij} + \lambda G^\top_{ij}) e_{ji} = 0$.□

Note that from slices of $\mathcal{A}^k_{ij}$ we can obtain three linearly independent homography matrices, thus we can find a unique solution to e_{ji} (each pair of homography matrices produces three solutions). Now that we have the means to recover epipoles and homography matrices we can proceed to the central result which is the reconstruction theorem:

Theorem 1 (reconstruction) *There exists a projective frame for which the first projection matrix takes the form $[I_{3\times3}; 0_{3\times2}]$ and all other projection matrices (of views 2,3,...) take the form:*

$$\tilde{M}_j = [H_j; v_j, v'_j]$$

where H_j is a homography matrix from view 1 to j induced by a fixed (but arbitrary) plane π, and v_j, v'_j are two points on the epipole (a line) e_{j1} on view j (projections of two fixed points in the COP of camera 1 onto view j).

Proof: Consider two views with projection matrices $\tilde{M}_1$ and $\tilde{M}_2$, a point P in space and matching image points

p, p' satisfying $p \cong \tilde{M}_1 P$ and $p' \cong \tilde{M}_2 P$. Let W be a (full-rank) 5×5 matrix representing some arbitrary projective change of coordinates, then $p \cong \tilde{M}_1 W W^{-1} P$ and $p' \cong \tilde{M}_2 W W^{-1} P$, thus we are allowed to choose W at will because reconstruction is only up to a projectivity in $\mathcal{P}^4$. Let C, C' be two points spanning the COP of camera 1, i.e., two points spanning the null space of $\tilde{M}_1$, thus $\tilde{M}_1 C = 0$ and $\tilde{M}_1 C' = 0$. Let $W = [U, C, C']$ for some 5×3 matrix U chosen such that $\tilde{M}_1 U = I_{3\times3}$. Clearly, $\tilde{M}_1 W = [I_{3\times3}; 0_{3\times2}]$.

Let U be chosen to consist of the first 3 columns of the matrix:

$$U = \begin{bmatrix} \tilde{M}_1 \\ C_\pi \end{bmatrix}^{-1}_{1-3}$$

where the subscript 1–3 signals that we are taking only columns 1–3 from the 5×5 matrix, and C_π is the 2×5 matrix defining the plane π, i.e., $C_\pi P = 0$ for all $P \in \pi$. Recall that a plane in $\mathcal{P}^4$ is the intersection (meet) of two hyperplanes (extensor of step 4) because $4+4-5 = 3$, thus a plane is defined by a 2×5 matrix whose rows represent the hyperplanes. We have that $\tilde{M}_1 U = I_{3\times3}$. Consider

$$\tilde{M}_2 W = \tilde{M}_2 [U, C, C'] = [\tilde{M}_2 U, v, v']$$

where $v = \tilde{M}_2 C$ and $v' = \tilde{M}_2 C'$ are two points on the epipole e_{21}. Recall that $e_{21} = \tilde{M}_2 null(\tilde{M}_1)$ and $null(\tilde{M}_1)$ is spanned by C, C'. What is left to show is that $\tilde{M}_2 U$ is a homography matrix H_π from view 1 to 2 induced by the plane π. This is shown next.

We have that

$$\begin{bmatrix} \tilde{M}_1 \\ C_\pi \end{bmatrix} P = \begin{pmatrix} \tilde{M}_1 P \\ C_\pi P \end{pmatrix} \cong \begin{pmatrix} p \\ 0 \\ 0 \end{pmatrix} \qquad \forall P \in \pi$$

From which we obtain:

$$\tilde{M}_2 U p = \tilde{M}_2 \begin{bmatrix} \tilde{M}_1 \\ C_\pi \end{bmatrix}^{-1} \begin{pmatrix} p \\ 0 \\ 0 \end{pmatrix} = \tilde{M}_2 P \cong p'$$

Thus, we have shown that $\tilde{M}_2 U p \cong p'$ for all matching points arising from points $P \in \pi$. □

Taken together, by using the homography slices of the tensor we can recover $\tilde{M}_2$. The third projection matrix $\tilde{M}_3$ can be recovered (linearly) from the tensor and $\tilde{M}_1, \tilde{M}_2$ because the tensor is a multi-linear form whose entries are multi-linear functions of the three projection matrices. Finally, it is not difficult to see that the family of homography matrices (as a function of the position of the plane π) has the general form with 7 degrees of freedom:

$$H_{\pi_1} = \lambda H_{\pi_2} + v n^T + v' n'^T,$$

where λ, n, n' are general.

3.2 The Geometry of $\mathcal{P}^6 \to \mathcal{P}^2$

In $\mathcal{P}^6 \to \mathcal{P}^2$ three image rays always intersect. This is because two extensors of step 5 in $\mathcal{P}^6$ intersect in an extensor of step of at least $5 + 5 - 7 = 3$, and an extensor of step 3 intersects an extensor of step 5 in a point. Thus we need more then three views of matching points in order to obtain a constraint. This agrees with the result we have noted in Problem 6 — a multi-linear constraint across four images $\mathcal{B}^l_{ijk}$ which is contracted by three points and a line.

Let p, p', p'', p''' be four matching points along views 1,2,3,4 and let s''', r''' be any two lines coincident with p'''. The multilinear constraint is expressed as follows:

$$p^i p'^j p''^k s'''_l \mathcal{B}^l_{ijk} = 0,$$

The entries of the tensor $\mathcal{B}^l_{ijk}$ are multilinear functions of the entries of the four projection matrices $\tilde{M}_1, \tilde{M}_2, \tilde{M}_3$ and $\tilde{M}_4$. The constraint itself is a point-point-point-line constraint, thus a triplet p, p', p'', p''' provides two linear constraints $p^i p'^j p''^k s'''_l \mathcal{B}^l_{ijk} = 0$, and $p^i p'^j p''^k r'''_l \mathcal{B}^l_{ijk} = 0$, on the entries of $\mathcal{B}^l_{ijk}$. Therefore, 40 matching triplets are sufficient for a (linear) solution. We will assume from now on that the tensor $\mathcal{B}^l_{ijk}$ was already recovered from image measurements and we wish to recover the 3×7 projection matrices $\tilde{M}_1, \tilde{M}_2, \tilde{M}_3, \tilde{M}_4$. As in the case of $\mathcal{P}^4 \to \mathcal{P}^2$, we will make use of tensor slices while recovering some basic elements of the projective settings. Note that for some of those elements, like homography matrices from view 2 to view 3, we will resort to permuted tensors, i.e., where the matches are for example point-point-line-point ($\mathcal{B}^k_{ijl}$). These permuted tensors can be recovered from exactly the same image measurements.

Claim 5 (point transfer)

$$p^i p'^j p''^k \mathcal{B}^l_{ijk} \cong p'''^l \tag{2}$$

Proof: Follows from the fact that $p^i p'^j p''^k s'''_l \mathcal{B}^l_{ijk} = 0$ for any line s''' coincident with p'''. From the covariant-contravariant structure of the tensor, $p^i p'^j p''^k \mathcal{B}^l_{ijk}$ is a point (contravariant vector), let this point be denoted by q^l. Hence, $q^l s'''_l = 0$ for all lines s''' that satisfy $s'''_l p'''^l = 0$. Thus q and p''' are the same point. □

The rays associated with p, p', p'' are extensors of step 5, which as explained in the preceding section intersect at a point, and thus $p^i p'^j p''^k \mathcal{B}^l_{ijk}$ is the back-projection onto view 4. Similarly, let l''' be some line in image 4. The image rays associated with a point p', p'' in images 2 and 3 and the extensor of step 6 associated with the join of l''' and the COP of camera 4 meet at a line ($(5+5-7)+6-7 = 2$) and let the projection of this line onto image 1 be denoted by l. The relationship between p', p'', l''', l is captured by the tensor: $p'^j p''^k l'''_l \mathcal{B}^l_{ijk} \cong l_i$.

Claim 6 (homography slice) *Let γ^j and δ^k be any contravariant vectors. The 3×3 matrix $\gamma^j \delta^k \mathcal{B}^l_{ijk}$ is a homography matrix (2D collineation) from views 1 to 4 induced by the plane defined by the intersection of image rays of γ and δ.*

Proof: Consider $(\gamma^j \delta^k \mathcal{B}^l_{ijk}) p^i = q^l$, from the point transfer equation 2 we have that q is the projection onto view 4 of the intersections of the three rays of sight corresponding to p, γ, δ. (recall that each ray of sight is an extensor of step 5 in $\mathcal{P}^6$ and that three such extensors generally intersect at a point). Let $\pi_{\gamma\delta}$ denote the plane associated with the intersection of the rays of sight of γ and δ. If we fix γ and δ and vary the point p over image 1, then the resulting points q are projection of points on the plane $\pi_{\gamma\delta}$ onto image 3. Thus the matrix $\gamma^j \delta^k \mathcal{B}^l_{ijk}$ is projective transformation from image 1 to image 4 induced by the plane $\pi_{\gamma\delta}$. $\square$

Likewise, $\gamma^i \delta^j \mathcal{B}^l_{ijk}$ is a homography matrix from image 3 to image 4, and $\gamma^i \delta^k \mathcal{B}^l_{ijk}$ is an homography matrix from image 2 to image 4.

The next item on the list of elementary building blocks for reconstruction of projection matrices are the epipoles. However, there are no epipoles in $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ because the join of two COPs (each is a step 4 extensor) fills up the entire space $\mathcal{P}^6$. We define instead the notion of "Joint Epipole" as follows:

Definition 1 (Joint Epipoles) *Let C_{ij} be the intersection (meet) of the centers of two projection matrices $\tilde{M}_i$ and $\tilde{M}_j$:*

$$C_{ij} \cong null(\tilde{M}_i) \wedge null(\tilde{M}_j).$$

C_{ij} is a point because $4+4-7=1$. Let c^k_{ij} be the projection of C_{ij} onto the k'th view, i.e., $c^k_{ij} \cong \tilde{M}_k C_{ij}$. We refer to c^k_{ij} the joint epipole in image k of the COPs of the projection matrices $\tilde{M}_i, \tilde{M}_j$.

Just as with epipoles in $\mathcal{P}^3 \rightarrow \mathcal{P}^2$, the joint epipoles are mapped to each other via homography matrices (which in turn are obtained from the homography slices of the tensor).

Claim 7 (Joint Epipoles) *Let $H_{il} = \gamma^j \delta^k \mathcal{B}^l_{ijk}$ be a homography matrix from view 1 to view 4, obtained by slicing the tensor $\mathcal{B}^l_{ijk}$, then: $Hc^1_{23} \cong c^4_{23}$*

Proof: The homography matrix $\gamma^j \delta^k \mathcal{B}^l_{ijk}$ from view 1 to view 4 is induced by the plane defined by the intersection of the rays of sights associated with the points γ and δ (see above). Each ray of sight (extensor of step 5) contains its projection center, hence the plane of intersection of two image rays must contain the point C_{23} (which is the intersection of both projection centers of views 2,3) — regardless of the choice of γ, δ. So any homography of this form H would satisfy $Hc^3_{12} \cong c^4_{12}$. $\square$

From the result above, and similarly to $\mathcal{P}^4 \rightarrow \mathcal{P}^2$, it is clear the joint epipoles are generalized eigenvectors of homography matrices obtained by slicing the tensor.

Now that we have the means to recover epipoles and homography matrices we can proceed to the (first) reconstruction theorem.

Theorem 2 (Reconstruction I) *There exists a projective frame for which the first projection matrix takes the form $[I_{3\times3}; I_{3\times3}; 0_{3\times1}]$ and all other projection matrices (of views 2,3,4,...) take the form:*

$$\tilde{M}_j = [H_j; G_j; v_j]$$

where H_j is a homography matrix from view 1 to j induced by a fixed (but arbitrary) plane π, G_j is a homography matrix from view 1 to j induced by another fixed arbitrary plane σ and v_j is the projection of a fixed arbitrary point contained in the first camera center to image j.

Proof: Reconstruction in $\mathcal{P}^6$ is given up to a 7×7 projective transformation W. Let C be a point inside the COP of camera 1, i.e., any point which satisfies $\tilde{M}_1 C = 0$. Let $W = [U, V, C]$ for some 5×3 matrices U and V chosen such that $\tilde{M}_1 U = \tilde{M}_1 V = I_{3\times3}$. Clearly, $\tilde{M}_1 W = [I_{3\times3}; I_{3\times3}; 0_{3\times1}]$.

Let U be chosen to consist of the first 3 columns of the matrix:

$$U = \left[\begin{array}{c} \tilde{M}_1 \\ C_\pi \end{array} \right]^{-1}_{1-3}$$

where the subscript 1–3 signals that we are taking only columns 1–3 from the inverted 7×7 matrix, and C_π is the 4×7 matrix defining the plane π, i.e., $C_\pi P = 0$ for all $P \in \pi$. Recall that a plane in $\mathcal{P}^6$ is dual to an extensor of step four and thus is defined by the intersection (meet) of four hyperplanes, i.e a plane is defined by a 4×7 matrix whose rows represent these hyperplanes. We have that $\tilde{M}_1 U = I_{3\times3}$. Likewise, let

$$V = \left[\begin{array}{c} \tilde{M}_1 \\ C_\sigma \end{array} \right]^{-1}_{1-3}$$

where C_σ is the 4×7 matrix representing the plane σ. Consider

$$\tilde{M}_2 W = \tilde{M}_2 [U, V, C] = [\tilde{M}_2 U, M_2 V, v]$$

where $v = \tilde{M}_2 C$. What is left to show is that $\tilde{M}_2 U$ is a homography matrix H_π from view 1 to 2 induced by the plane π, and that $\tilde{M}_2 V$ is a homography matrix H_σ from view 1 to 2 induced by the plane σ. The proof of this is very similar to what was done in the proof of Theorem 1. $\square$

This reconstruction theorem is not ready yet for practical use because one needs homographies of two planes from view 1 and view 2, and homographies for the same planes from view 1 to view 3. One also needs the projection to views 2 and 3 of the same point C in the first camera center. (The fourth camera can then be recovered linearly from the tensor $\mathcal{B}_{ijk}^{l}$ - which is multilinear in the entries of the camera matrices).

Although it is fairly easy to find homography matrices between any two views (simply take slices of the tensors), it is difficult finding homographies of some fixed plane across three views. We will show later that it is possible to select a canonical coordinate system which allows choosing homography matrices between two views only (instead of across three views). As a preparation for this, we define next the "correlation slices" of the tensor:

Claim 8 (correlation slices) *$\gamma^i \delta_l \mathcal{B}_{ijk}^{l}$ is a mapping (correlation matrix) from points in the second view to a line in the third view (or from points in the third view to lines in the second view). This mapping is associated with the extensor of step 4 defined by the intersection of an extensor of step 5 with an extensor of step 6 ($5+6-7=4$). The step 5 extensor is the ray of sight associated with γ (in view 1). The step 6 extensor is the join of the line in the 4'th image plane δ and the projection center (extensor of step 4) of the forth camera.*

Proof: $\gamma^i p^j q^k \delta_l \mathcal{B}_{ijk}^{l} = 0$ iff the lines of sight associated with γ, p, q and the step 6 extensor associated with δ all intersect in at least one point. Fixing γ and δ we get a fixed extensor of step $5+6-7=4$. The equation $p^j q^k (\gamma^i \delta_l \mathcal{B}_{ijk}^{l}) = 0$ implies that the lines of sight associated with p^j and q^k intersect that extensor at a single point. The line of sight associated with p^j intersects that fixed extensor in an extensor of step $4+5-7=2$ — which is a line. Every point q^k on the projection of that line onto view three has to satisfy $p^j q^k (\gamma^i \delta_l \mathcal{B}_{ijk}^{l}) = 0$, hence the projection of this line is $p^j (\gamma^i \delta_l \mathcal{B}_{ijk}^{l})$. $\square$

This correlation matrix can be seen as the "Fundamental matrix" of the extensor of step four space, where the effective "camera centers" are the intersection of the COP of the $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ projection matrices with that space.

Using the correlation slices introduced above we wish to describe a homography matrix H from view 1 onto view 3 associated with a plane which is *contained* in the second view projection center (which is a step 4 extensor). Let $Q_1 = p^j s_l \mathcal{B}_{ijk}^{l}$ and $Q_2 = q^j s_l \mathcal{B}_{ijk}^{l}$ be the correlation matrices described above — each is associated with an extensor of step 4. Generally, two extensors of step 4 intersect (meet) at a point ($4+4-7=1$), however in this particular case since the image line s is shared among the two extensors, their meet is a step 3 extensor (a plane). To see why this is so, let $\mathcal{Q}_1$ be the step 4 extensor associated with the correlation matrix Q_1, and let $\mathcal{Q}_2$ be the step 4 extensor associated with the correlation matrix Q_2. Let $\hat{p}, \hat{q}, \hat{s}$ be the embedded image points and lines in $\mathcal{P}^6$. We have:

$$\begin{aligned} \mathcal{Q}_1 &= (c_2 \vee \hat{p}) \wedge (c_4 \vee \hat{s}) \\ \mathcal{Q}_1 &= (c_2 \vee \hat{q}) \wedge (c_4 \vee \hat{s}) \end{aligned}$$

where c_2, c_4 are the step 4 extensors representing the projection centers of view 2,4 respectively; and "$\vee$" denotes the join operation and "$\wedge$" denotes the intersection (meet) operation. Because the step 6 extensor $c_4 \vee \hat{s}$ is shared, and also noting that $(c_2 \vee \hat{p}) \wedge (c_2 \vee \hat{q}) = c_2$ because p, q are points in view 2, then

$$\begin{aligned} \mathcal{Q}_1 \wedge \mathcal{Q}_2 &= (c_2 \vee \hat{p}) \wedge (c_2 \vee \hat{q}) \wedge (c_4 \vee \hat{s}) \\ &= c_2 \wedge (c_4 \vee \hat{s}) \end{aligned}$$

Therefore, $\mathcal{Q}_1 \wedge \mathcal{Q}_2$ is the intersection of a step 4 and step 6 extensors, which is a plane ($4+6-7=3$) *contained in the center of projection c_2 of view 2*. Since Q_1, Q_2 are the mappings from view 1 to view 3 induced by the step 4 extensors $\mathcal{Q}_1, \mathcal{Q}_2$ respectively[1], the mapping $Q_1 p \times Q_2 p$ from view 1 to view 3 is a homography induced by the plane $\mathcal{Q}_1 \wedge \mathcal{Q}_2$. The homography matrix H can be recovered directly (linearly) from the matrices Q_1, Q_2 by noting that $Q_1^T H$ and $Q_2^T H$ are anti-symmetrical — thus providing 6 linear constraints each for H.

Now that we have a tool for the recovery of homography matrices which lie inside projection matrix centers we can proceed to the second (simplified) reconstruction theorem:

Theorem 3 (reconstruction II) *There exists a projective frame for which the first and second projection matrices take the form*

$$\begin{aligned} \tilde{M}_1 &\cong [\, I_{3\times 3} \quad 0_{3\times 3} \quad 0_{3\times 1} \,] \\ \tilde{M}_2 &\cong [\, 0_{3\times 3} \quad I_{3\times 3} \quad 0_{3\times 1} \,] \end{aligned}$$

and all other projection matrices (of views 3,4,...) take the form:

$$\tilde{M}_j \cong [\, H_{1j} \quad H_{2j} \quad c_{12}^{j} \,]$$

where H_{1j} is a homography matrix from view 1 to view j induced by a plane π which is contained in the second projection matrix center, H_{2j} is a homography matrix from view 2 to view j induced by a plane σ which is contained in the first projection matrix center, and c_{12}^{j} is the joint epipole, i.e., the projection onto view j of the intersection point of the projection centers of views 1,2.

Proof: Consider three views with projection matrices $\tilde{M}_j$, $j=1,2,3$, a point $P \in \mathcal{P}^6$ in space and matching image points p, p', p'' satisfying $p \cong \tilde{M}_1 P, p' \cong \tilde{M}_2 P$ and

[1] Such a mapping must be a correlation by definition because the image ray of view 1 intersects the step 4 extensor at a line ($5+4-7=2$) whose projection onto view 3 is a line.

$p'' \cong \tilde{M}_3 P$. Since reconstruction is determined up to a projectivity, let W be a (full-rank) 7×7 matrix representing some arbitrary projective change of coordinates (we are allowed to choose W at will). Let C be the point of intersection of the projection centers of views 1,2 (each is a step 4 extensor, thus they intersect at a point because $4+4-7=1$), thus $\tilde{M}_1 C = 0$ and $\tilde{M}_2 C = 0$ and $\tilde{M}_3 C = c^3_{12}$ (the joint epipole). Let π be some plane contained in $null(\tilde{M}_2)$ and let σ be some plane contained in $null(\tilde{M}_1)$. Let C_π be the 4×7 matrix defining the plane π, i.e., $C_\pi P = 0$ for all $P \in \pi$; and let C_σ be the 4×7 matrix defining the plane σ. Let $W = [U, V, C]$ where U, V are 7×3 matrices defined as follows.

$$U = \begin{bmatrix} \tilde{M}_1 \\ C_\pi \end{bmatrix}^{-1}_{1-3} \quad V = \begin{bmatrix} \tilde{M}_2 \\ C_\sigma \end{bmatrix}^{-1}_{1-3}$$

where the subscript 1–3 signals that we are taking only columns 1–3 from the inverted 7×7 matrix. We have that $\tilde{M}_1 U = I_{3\times 3}$ and $\tilde{M}_2 V = I$. Moreover, the columns of U consist of points on π and since π is contained in $null(\tilde{M}_2)$ we have that $\tilde{M}_2 U = 0$; and likewise $\tilde{M}_1 V = 0$. To see why this is so, recall that

$$\begin{bmatrix} \tilde{M}_1 \\ C_\pi \end{bmatrix} P = \begin{pmatrix} \tilde{M}_1 P \\ C_\pi P \end{pmatrix} \cong \begin{pmatrix} p \\ 0 \\ 0 \end{pmatrix} \qquad \forall P \in \pi$$

from which we obtain that $Up = P$, i.e., U maps the first image plane onto the plane π. Thus, in particular the columns of U are points on π. Taken together, we have that for these choices of planes π and σ, the first two projection matrices are:

$$\begin{aligned} \tilde{M}_1 W &\cong [\, I_{3\times 3} \quad 0_{3\times 3} \quad 0_{3\times 1} \,] \\ \tilde{M}_2 W &\cong [\, 0_{3\times 3} \quad I_{3\times 3} \quad 0_{3\times 1} \,] \end{aligned}$$

We show next that $\tilde{M}_3 U$ is a homography matrix from view 1 to 3 induced by π. Recall that Up is a point $P \in \pi$, thus $\tilde{M}_3 U p = \tilde{M}_3 P \cong p''$ where p, p'' are projections of a point in π. Similarly, Vp' is a point $P \in \sigma$, thus $\tilde{M}_3 V p' = \tilde{M}_3 P \cong p''$ where p', p'' are projections of a point on σ. Taken together, we have

$$\tilde{M}_3 W \cong [\, H_{13} \quad H_{23} \quad c^3_{12} \,] \,.$$

Putting together the correlation slices and the reconstruction theorem above, we see that for reconstruction of projection matrices all we need to do is to choose 2 correlation slices from which H_{13} is recovered (linearly), and choose another pair of correlation slices from which H_{23} is recovered. Then, by using homography slices we can recover the joint epipole c^3_{12} and we have thus created $\tilde{M}_3$. The fourth projection matrix $\tilde{M}_4$ can be recovered (linearly) from the tensor and the three projection matrices.

3.3 Reconstruction of the $\mathcal{P}^3 \rightarrow \mathcal{P}^2$ Camera Matrices

Given that we have recovered the projection matrices $\tilde{H}_j$, $j = 1, 2, 3$, of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$, and the projection matrices $\tilde{M}_j$, $j = 1, 2, 3, 4$ of $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ we wish to recover the original 3×4 camera matrices up to a 3D Affine ambiguity. The special structure of the matrices $\tilde{H}$ and $\tilde{M}$ — they have repeated scaled columns — provides us with linear constraints on a the coordinate change in $\mathcal{P}^k \rightarrow \mathcal{P}^2$ which will transform the recovered matrices $\tilde{H}$ and $\tilde{M}$ to the admissible ones we are looking for.

In the case of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$, since the third column of $\tilde{H}_j$ is unconstrained, the family of collineations of $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ that leave the structural form intact is organized as follows:

$$\begin{pmatrix} a & b & e & 0 & 0 \\ c & d & f & 0 & 0 \\ 0 & 0 & g & 0 & 0 \\ 0 & 0 & h & a & b \\ 0 & 0 & i & c & d \end{pmatrix}$$

Note that we have 9 degrees of freedom up to scale, which means we have 8 free parameters — 2 more than what is allowed for a 2D affinity. The extra degrees of freedom could be compensated for by applying another transformation of the form:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \hat{h} & 1 & 0 \\ 0 & 0 & \hat{i} & 0 & 1 \end{pmatrix}.$$

The unknown variables $\hat{h}$ and $\hat{i}$ can be solved using a single static point, as follows. Let $\widehat{H}_j$ be the projection matrices up to the unknown correction $\hat{h}$ and $\hat{i}$. Let H_j to be the left 3×3 part of $\widehat{H}_j$. Let p_1, p_2 be a matching pair in views 1,2 of a static point. Then,

$$p_2 \cong H_2 \begin{bmatrix} 1 & 0 & \hat{h} \\ 0 & 1 & \hat{i} \\ 0 & 0 & 1 \end{bmatrix} H_1^{-1} p_1$$

This gives us two linear equations for solving $\hat{h}$ and $\hat{i}$. The resulting homography matrices (up to a 2D Affine ambiguity) are:

$$H_1, H_2 \begin{bmatrix} 1 & 0 & \hat{h} \\ 0 & 1 & \hat{i} \\ 0 & 0 & 1 \end{bmatrix}, H_3 \begin{bmatrix} 1 & 0 & 2\hat{h} \\ 0 & 1 & 2\hat{i} \\ 0 & 0 & 1 \end{bmatrix}$$

In the case of $\mathcal{P}^6 \rightarrow \mathcal{P}^2$, the reconstruction of $\tilde{M}_j$ satisfying the structural constraints up to a 3D Affine ambiguity proceeds along similar lines. The ambiguity matrix is of this form:

$$\begin{pmatrix} a & b & c & j & 0 & 0 & 0 \\ d & e & f & k & 0 & 0 & 0 \\ g & h & i & l & 0 & 0 & 0 \\ 0 & 0 & 0 & m & 0 & 0 & 0 \\ 0 & 0 & 0 & n & a & b & c \\ 0 & 0 & 0 & o & d & e & f \\ 0 & 0 & 0 & p & g & h & i \end{pmatrix}$$

This kind of matrices is an Affine transformation on the left 3×4 part of the projection matrix from $\mathcal{P}^6$ to $\mathcal{P}^2$, but it is a different Affine transformation for every view.

Here again we can take the first recovered camera matrix to be the left part of the transformed projective camera matrix. We have to find only some transformation of the form:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \hat{n} & 1 & 0 & 0 \\ 0 & 0 & 0 & \hat{o} & 0 & 1 & 0 \\ 0 & 0 & 0 & \hat{p} & 0 & 0 & 1 \end{pmatrix}$$

Assuming that we know one static point, we can extract eight *linear* constraints on the unknowns $\hat{n}, \hat{o}, \hat{p}$ of the form:

$$\det \left(\begin{bmatrix} l_0^T R_0 \\ l_1^T R_1 \begin{bmatrix} 1 & 0 & 0 & \hat{n} \\ 0 & 1 & 0 & \hat{o} \\ 0 & 0 & 1 & \hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ l_2^T R_2 \begin{bmatrix} 1 & 0 & 0 & 2\hat{n} \\ 0 & 1 & 0 & 2\hat{o} \\ 0 & 0 & 1 & 2\hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ l_3^T R_3 \begin{bmatrix} 1 & 0 & 0 & 3\hat{n} \\ 0 & 1 & 0 & 3\hat{o} \\ 0 & 0 & 1 & 3\hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{bmatrix} \right) = 0$$

Where R_i are the left parts of the transformed projective camera matrices, and l_i are lines through the tracked static point. The final cameras would be:

$$R_0, R_1 \begin{bmatrix} 1 & 0 & 0 & \hat{n} \\ 0 & 1 & 0 & \hat{o} \\ 0 & 0 & 1 & \hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix}, R_2 \begin{bmatrix} 1 & 0 & 0 & 2\hat{n} \\ 0 & 1 & 0 & 2\hat{o} \\ 0 & 0 & 1 & 2\hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix}, R_3 \begin{bmatrix} 1 & 0 & 0 & 3\hat{n} \\ 0 & 1 & 0 & 3\hat{o} \\ 0 & 0 & 1 & 3\hat{p} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3.4 Reconstruction of Segmentation Tensors

The stage of the reconstruction of the underlying structure is (as noted above) application dependent. For reconstruction in the case of the segmentation tensor, we do not have any special information about structure of the projection matrices. Here we may use some known points on one object in order to reconstruct in 2D/3D.

In the planar segmentation tensor case we know that the space in $\mathcal{P}^3$ spanned by points on one object is a space of rank 3. From 3 point matches in two images (or even point-line matches), we can reconstruct 3 points in that rank 3 subspace of $\mathcal{P}^3$. Note that using the $\mathcal{P}^4$ to $\mathcal{P}^2$ projection matrices we've recovered earlier, we do not need a forth basis point in order to determine the projection of each point in this space to the images. Hence we compute the homography of the first object is achieved. Now that we know the homographies of the first object, segmentation is possible, so we can determine the homography of the second object from its points. The next stage is to find a transformation that will make the first 2 columns of the homographies identical. The resulting solution would be the real homographies up to an Affine transformation.

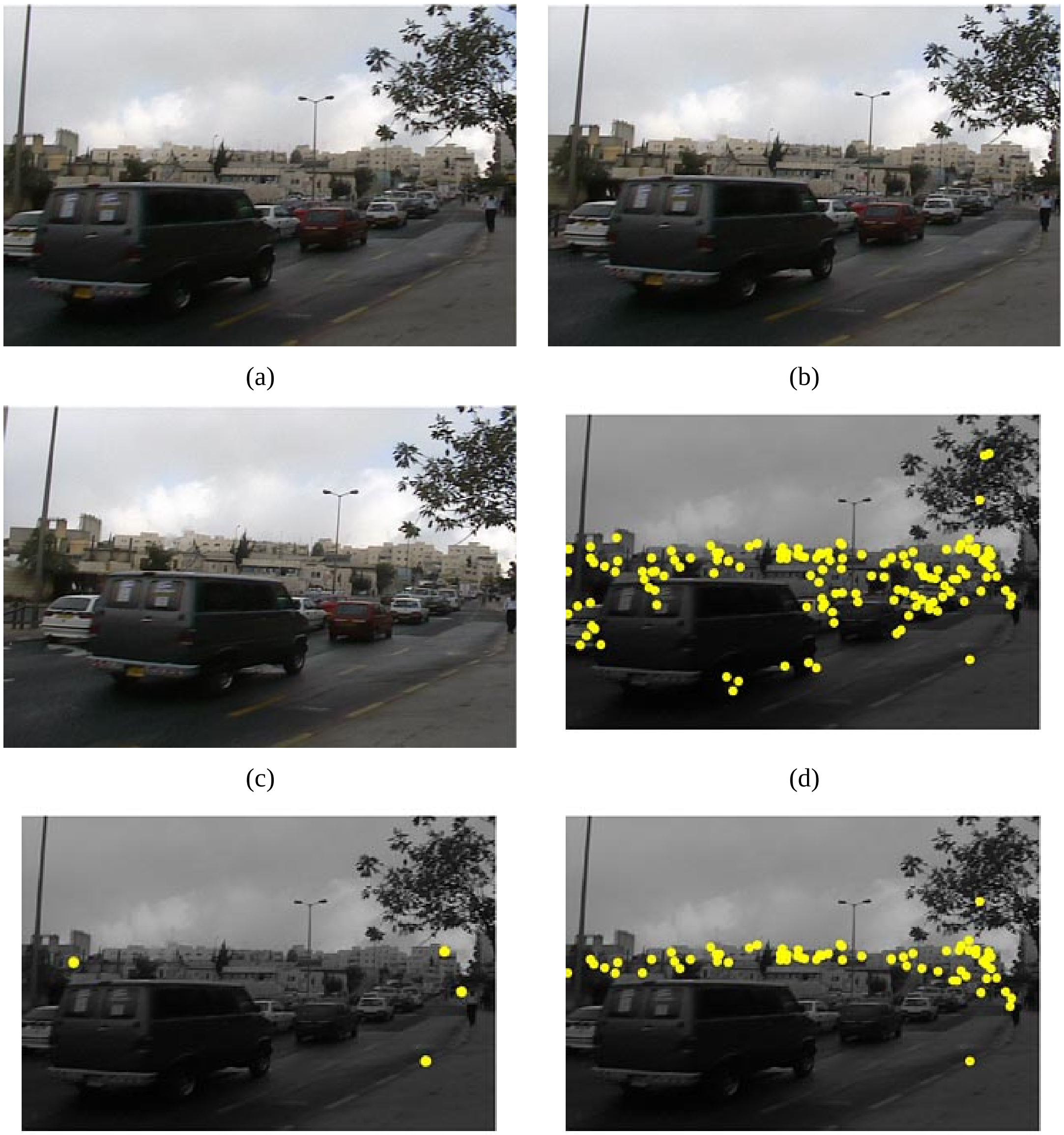
The segmentation tensor for the 3D case is similar. Here we are going to have to use 4 point matches from one object in order to recover the set of cameras for the first object over time. These cameras would be defined up to a projective transformation. Segmentation would now give us points on the second object, from which recovery of the motion of the second camera is possible. Aligning these sets of cameras would give us a common Affine reconstruction. Note that both sets of cameras agree on the homography at infinity. Thus the recovery of that homography can be achieved for example by intersecting epipolar lines.

The case of the constant velocity in 3D going in one direction is similar to the case of the 3D segmentation tensor. Note that recovery of the image projections of the common direction in 3D can be achieved, although we can not use this information as one of our 4 points. This is because this point has more then one reconstruction in $\mathcal{P}^4$ from it's point matches (as a static point, or as pure motion, or any combination of the two).

4 Experiments

We describe an experiment for one of the applications in this paper, the 3D segmentation tensor (Problem 4). Recall that we observe views of a scene containing two bodies moving in relative translation to one another. The $\mathcal{P}^4 \rightarrow \mathcal{P}^2$ problem formulation requires a matching set of at least 13 points across 3 views where the points come from both bodies in an unsegmented fashion. The triplets of matching points are used to construct a $3 \times 3 \times 3$ tensor such that with the segmentation of 4 points on one of the bodies one can then segment the entire scene.

The scene in the experiment, displayed in Fig. 1, consists of a rigid background (first body) and a foreground consisting of a number of vehicles moving cohesively together (second body). Image points were identified and tracked using openCV's [11] KLT [9] tracker. Fig. 1(a-c) shows the three views , Fig. 1d shows the points which were tracked along the sequence and used for recovery of the tensor. Fig. 1e shows the 4 labeled points (on the background body) used to segment the entire scene, and Fig. 1f shows the segmentations result — all point on the background body were correctly classified as such.



(a) (b) (c) (d) (e) (f)

Figure 1: 3D segmentation tensor experiment. See text for details.

5 Summary

This paper has two parts. In Section 2 we have shown that multi-view constraints of scenes containing multiple linearly moving points can be derived by "lifting" the non-rigid 3D phenomena into a rigid configuration in a higher dimensional space of $\mathcal{P}^k$. And to that end we have presented 6 applications for various values of k ranging from 3 to 6.

In the second part of the paper (Section 3) we worked out the details of describing and recovering $3 \times (k+1)$ projection matrices (for $k = 4, 6$) from the multi-view tensors' slices, and the details of recovering the 3×4 original camera matrices from the projection matrices.

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] 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.

[4] O.D. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[5] 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.

[6] 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.

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

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

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

[10] 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.

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

[12] A. Shashua and Lior 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*, Dublin, Ireland, June 2000.

[13] 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.