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 $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 4, 5, 6$ 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 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 $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 4, 5, 6$ 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. 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.

*The length of this paper was considerably reduced to fit the length guidelines of this proceedings. The full-length version of this work can be found in http://www.cs.huji.ac.il/∼ shashua/papers/pkp2 − journal.pdf

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

We will describe below a number of different applications for values of $k = 4, 5, 6$ (for $k = 3$ see [14]). These applications include multi-body segmentation (we call "segmentation tensors") and multiple linearly moving points.

2.1 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. The first application consists of 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 1 (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:

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

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 2 (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 3 (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^1_j, $j = 0, 1, 2$, provide the image points $p_j \cong M^1_j P$. Likewise, if P is on the second body then $p_j \cong M^2_j P$. 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:

$$M^1_j \cong [\, A^j_\infty v^1_j \,] \quad M^2_j \cong [\, A^j_\infty v^2_j \,].$$

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:

$$M_j \cong [\, A^j_\infty v^1_j v^2_j \,].$$

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.

2.2 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 4 (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 2, the j'th projection matrix $\tilde{M}_j$ has the form $[M_j, jM_j dP_1, jM_j dP_2]$ 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 $p'' \cong C(p_1, p_2, p_3)^\top + D(p_4, p_5, p_6)^\top$. □

2.3 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 5 (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 $P_i + j dP_i$. 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 $\mathcal{B}_{ijk}^q$, i.e., $p^i p'^j p''^k l'''_q \mathcal{B}_{ijk}^q = 0$. 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.

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 $\mathcal{P}^4 \rightarrow \mathcal{P}^2$. The detailed derivation of $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ can be found in the full-length version of this work [14].

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", described in [14].

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 1 that the multi-linear constraint across three views takes the form of a $3 \times 3 \times 3$ tensor $\mathcal{A}_{ij}^k$ 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:

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

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 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 $p^i p'^j s''_k \mathcal{A}_{ij}^k = 0$ 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 (linear) solution. We will assume from now on that the tensor $\mathcal{A}_{ij}^k$ 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)

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

Proof: Follows from the fact that $p^i p'^j s''_k \mathcal{A}_{ij}^k = 0$ for any line s'' coincident with p''. From the covariant-contravariant structure of the tensor, $p^i p'^j \mathcal{A}_{ij}^k$ 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 $p^i p'^j \mathcal{A}_{ij}^k$ 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: $p'^j l''_k \mathcal{A}_{ij}^k \cong l_i$.

Claim 3 (homography slice) *Let δ^j be any contravariant vector. The 3×3 matrix $\delta^j \mathcal{A}_{ij}^k$ 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 $(\delta^j \mathcal{A}_{ij}^k) p^i = q^k$, 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 $\delta^j \mathcal{A}_{ij}^k$ is projective transformation from image 1 to image 3 induced by the plane π_δ. □

Note that $\delta^j \mathcal{A}_{ij}^k$ is a linear combination of the three slices $\mathcal{A}_{i1}^k, \mathcal{A}_{i2}^k$ and $\mathcal{A}_{i3}^k$. Thus, in particular a slice (through the "j" index) produces a homography matrix. Likewise, $\delta^i \mathcal{A}_{ij}^k$ 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_{ij}^T, G_{ij}^T, i.e., satisfies the equation:*

$$(H_{ij}^\top + \lambda G_{ij}^\top)e_{ji} = 0.$$

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

Note that from slices of $\mathcal{A}_{ij}^k$ 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 $[I_{3\times 3}; 0_{3\times 2}]$ 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 $\tilde{M}_1 U = I_{3\times 3}$. Clearly, $\tilde{M}_1 W = [I_{3\times 3}; 0_{3\times 2}]$.

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 $\tilde{M}_1 U = I_{3\times 3}$. 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 $P \in \pi$. $\square$

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 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:

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

The case of $\mathcal{P}^6 \rightarrow \mathcal{P}^2$ and segmentation tensors are described in [14].

4 Experiments

We describe an experiment for one of the applications in this paper, the 3D segmentation tensor (Problem 3). 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 the $3 \times 3 \times 3$ tensor $\mathcal{A}_{ij}^k$ 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) shows one of the three views, Fig. 1b shows the points which were tracked along the sequence and used for recovery of the tensor. Fig. 1c shows the 4 labeled points (on the background body) used to segment the entire scene, and Fig. 1d shows the segmentation result — all points on the background body were correctly classified as such.

5 Summary

We have introduced 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	**??**
$\mathcal{P}^4$	2D constant velocity tensor	3^3	2.1
$\mathcal{P}^4$	3D segmentation tensor	3^3	2.1
$\mathcal{P}^4$	3D constant collinear velocity	3^3	2.1
$\mathcal{P}^5$	3D constant coplanar velocity	3^3	2.2
$\mathcal{P}^6$	3D constant velocity tensor	3^4	2.3

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 $k = 4$, see [14] for $k = 6$) 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.

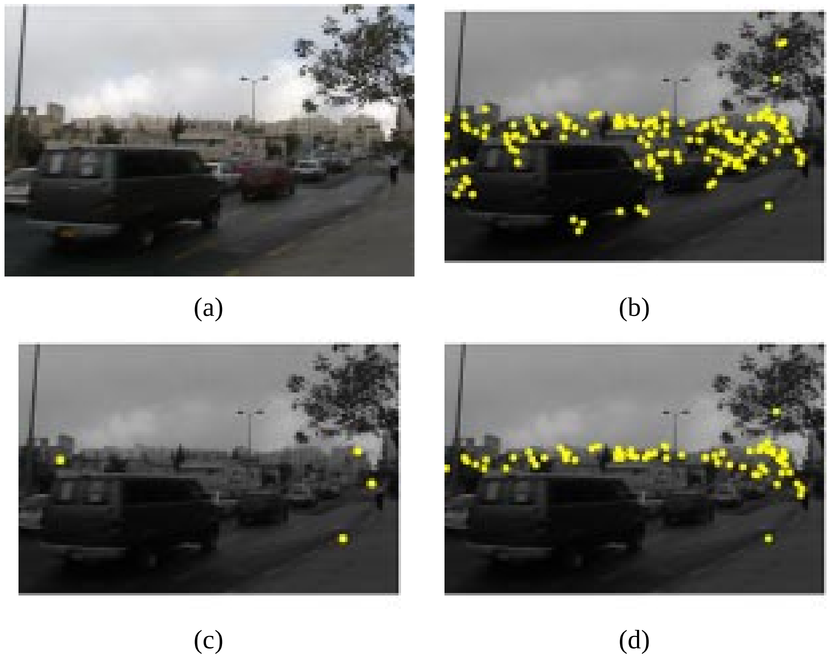

Figure 1. 3D segmentation tensor experiment. (a) first view out of a sequence of three views of a scene consisting of two bodies moving in relative translation to one another: the vehicles and the static background. (b) points (unsegmented) matched across the three views for the purpose of recovering the tensor $\mathcal{A}_{ij}^{k}$. (c) four points (manually chosen) on one body (the background) completely determine the segmentation of the scene into two bodies (display (d)).

[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.

[14] http://www.cs.huji.ac.il/∼ shashua/papers/pkp2 − journal.pdf