
Dynamic \mathcal{P}^n to \mathcal{P}^n Alignment

Amnon Shashua and Lior Wolf*

School of Engineering and Computer Science,
the Hebrew University of Jerusalem,
Jerusalem, 91904, Israel
{shashua, lwolf}@cs.huji.ac.il

We introduce in this chapter A generalization of the classical collineation of \mathcal{P}^n . The generalization allows for a certain degree of freedom in the localization of points in \mathcal{P}^n at the expense of using multiple $m > 2$ views. The degree of freedom per point is governed by an additional parameter $1 \leq k < n$ which stands for the dimension of the subspace in which the individual points are allowed to move while the projective change of coordinates take place. In other words, the point set is not necessarily stationary allowing the configuration to change while the entire coordinate system undergoes a projective change of coordinates. If we denote a change of coordinates as a "view" of the physical set of points, then in this chapter we discuss the multi-view relations that can be determined from observations (views) of a dynamically changing point-configuration — the underlying transformations and how they can be recovered.

For example, for a point set in \mathcal{P}^2 (planar configuration) undergoing linear motion the multiple views of the point set generate a multilinear constraint across three views governed by a $3 \times 3 \times 3$ contravariant tensor \mathcal{H} . The tensor, referred to as *homography tensor*, can be recovered linearly from 26 observations (matching points across three views) and once recovered can be unfolded to yield the global coordinate change (the individual pair of homography matrices). A point set in \mathcal{P}^3 (3D configuration) can undergo motion in a plane or along a line (each point independently). For the line motion, the multi-view constraints are governed by a $4 \times 4 \times 4$ family of contravariant tensors \mathcal{J} that capture the dynamic 3D-to-3D alignment problem. More generally, the family of homography tensors is captured by three parameters: the dimension n of the observation space, the dimension $k < n$ of the subspace along which each point of the point set is allowed to move, and the number of "views" m . Formally, the homography tensors form a $GL(V)$ module, denoted by $V(n, m, k)$, defined by the set of all tensors $v_1 \otimes \dots \otimes v_m \in V^{\otimes m}$ where v_i are n -dimensional

* L. Wolf's current address is at the M.I.T. Center for Computational and Biological Learning (CBCL), Cambridge, MA 02139.

vectors and $\dim \text{Span}\{v_1, \dots, v_m\} \leq k$. We will be interested in the structure and dimension of $V(n, m, k)$.

The notion of using multi-view analysis for non-rigid scenes is interesting and useful on its own right. In a way, this work extends the notion of “stereo triangulation” (a stationary point observed by two or more views), to the notion of “what can be recovered from line of sight measurements only?”. The chapter includes a detailed exposition of these tensors for \mathcal{P}^2 and \mathcal{P}^3 , their properties and applications and derive the dimension of $V(n, m, k)$ in the general case.

1 Introduction

Consider the classic problem of “3D to 3D” alignment of point sets. We are given a set of 3D points P_1, \dots, P_n measured by some device such as a structured light range sensor [14] or a stereo rig of cameras. When the sensor changes its position in space while the 3D points remain stationary, the 3D positions of the measured points P'_1, \dots, P'_n , have undergone a coordinate transformation. In a projective setting, five of these matching pairs in general position are sufficient to recover the 4×4 collineation A such that $AP_i \cong P'_i$, $i = 1, \dots, n$. In a rigid motion setting the coordinate transformation consists of translation and rotation which can be recovered using four matching points; elegant techniques using SVD have been developed for this purpose [4].

In the same vein, consider another popular group of transformations which include the planar collineations between two sets of points on the projective plane \mathcal{P}^2 undergoing a projective mapping. The planar collineations (homographies) are the 3×3 non-singular matrices which map between point sets undergoing a general projectivity. The planar homographies form a fundamental building block in multiple-view geometry in computer vision. The object stands on its own as a point-transfer vehicle for planar scenes (aerial photographs, for example) and in applications of mosaicing, camera stabilization and tracking [8]; a homography matrix is a standard building block in handling 3D scenes from multiple 2D projections: the “plane+parallax” framework [11, 6, 7, 3] uses a homography matrix for setting up a parallax residual field relative to a planar reference surface, and the trifocal tensor of three views is represented by a “homography-epipole” structure whose slices are homography matrices as well [5, 12].

The two examples above, general collineations of \mathcal{P}^2 and \mathcal{P}^3 , readily extend to n -dimensional projective spaces \mathcal{P}^n . A change of coordinates in an n -dimensional projective space \mathcal{P}^n is determined by an $(n+1) \times (n+1)$ matrix. Conversely, given two sets of points in \mathcal{P}^n which result by having one set undergo some collineation, the alignment of the two sets can be achieved by a homography matrix which can be determined uniquely from $n+1$ matching points.

In this chapter we introduce a “dynamic” version of the \mathcal{P}^n to \mathcal{P}^n alignment problem by allowing the individual points of the point set to undergo independent motion within k -dimensional subspaces while the entire point set undergoes a general collineation successively. For example, in the dynamic $\mathcal{P}^2 \rightarrow \mathcal{P}^2$ version, we allow for the possibility that any number of the points may move along straight-line paths during the change of view. A change of view results in a global change of coordinates (a collineation) but while doing so the individual points of the point set have changed relative position to one another. Points that remain in place are called *stationary* and points that move are called *dynamic*. There can be any number of dynamic points — including the possibility that *all* points are dynamic — and the system need not know in advance which of the points are stationary and which are dynamic (an unsegmented configuration). Under these conditions we wish to find the multiple projective coordinate changes from the point-match observations of the point set under successive coordinate changes. We will show that this type of transformation is governed by a $3 \times 3 \times 3$ tensor which captures the multi-view relation of the changing planar point-set. The tensor is formed by a bilinear product of the global pair of homography matrices (responsible for the changes of coordinates between the first view and the other two views). For every triplet of matching points across the three views p, p', p'' the following contravariant relation $p^i p'^j p''^k H_{ijk} = 0$ vanishes. The vanishing constraint provides a linear equation on the elements of the tensor and the global coordinate changes can later be recovered (also linearly) from the tensor.

The dynamic $\mathcal{P}^3 \rightarrow \mathcal{P}^3$ alignment problem can be viewed as a 3D sensor which changes position in 3D space (thus creating global coordinate changes) while the physical points in space undergo independent motion — either along straight line paths or along planar subspaces (or stay put). We will show that this type of transformation is governed by a $4 \times 4 \times 4$ family of tensors which vanishes on each of the matching triplets induced by a physical point under three coordinate systems.

More generally, the family of tensors governing the $\mathcal{P}^n \rightarrow \mathcal{P}^n$ alignment problem is captured by three parameters: the dimension n of the observation space, the dimension $k < n$ of the subspace along which each point of the point set is allowed to move, and the number of “views” m . Formally, these tensors form a $GL(V)$ module, denoted by $V(n, m, k)$, defined by the set of all tensors $v_1 \otimes \cdots \otimes v_m \in V^{\otimes m}$ where v_i are n -dimensional vectors and $\dim \text{Span}\{v_1, \dots, v_m\} \leq k$. We will be interested in the structure and dimension of $V(n, m, k)$.

We will describe in detail the tensor families that are associated with \mathcal{P}^2 and \mathcal{P}^3 , their definition, the way they can be recovered from observations, their properties and applications. The general case will be discussed at a reduced scope where we will address only the dimension of $V(n, m, k)$ (the number independent linear constraints possible for a given value of n, m, k). Other issues which are addressed in derivations for \mathcal{P}^2 and \mathcal{P}^3 (such as mixed

stationary and dynamic motions) are left open in the general case. Part of the material described in Sections 2 and 3 appeared in the proceedings of [13, 16] and the material of Section 4 in the technical report [10].

1.1 Background and Notations

We will be working with the projective space \mathcal{P}^n . A point in \mathcal{P}^n is defined by $n + 1$ numbers, not all zero, that form a coordinate vector defined up to a scale factor. The dual projective space represents the space of hyperplanes which are also defined by a $n + 1$ tuple of numbers. For example, a point p in the projective plane \mathcal{P}^2 coincides with a line s if and only if $p^\top s = 0$, i.e., the scalar product vanishes. In other words, the set of lines coincident with the point p are represented by the coordinate vectors s that satisfy $p^\top s = 0$, and vice versa: a point represented by the coordinate vector p can be thought of as the set of lines through it (a.k.a the pencil of lines through p). A line s going through two points p_1, p_2 is represented by the cross product $s \cong p_1 \times p_2$ where \cong denotes equality up to scale. Likewise, the point of intersection p of the lines s_1, s_2 is represented by $p \cong s_1 \times s_2$. In projective 3D space \mathcal{P}^3 , A point p lies on a plane π if and only if $p^\top \pi = 0$. In other words, in \mathcal{P}^2 points and lines are dual to each other and in \mathcal{P}^3 points and planes are duals to each other — generally, points and hyperplanes are duals.

In projective space any $n + 1$ points in general position (i.e., no subset of n points lie on a hyperplane) can be uniquely mapped into any other $n + 1$ general point configuration. Such a mapping is called a *collineation* and is defined by an invertible $(n + 1) \times (n + 1)$ matrix (also known as the homography matrix) defined up to scale. In particular, the change of coordinates of a planar configuration induced by taking a photograph by a pin-hole camera moving freely in the 3D world is represented by a 3×3 homography matrix, and the change of coordinates of a 3D point configuration caused by the motion of the sensor is represented by a 4×4 homography matrix. If H is a homography matrix (defined by $n + 1$ matching pairs of points), then H^{-T} (inverse transpose) is the dual homography that maps hyperplanes onto hyperplanes.

The projective plane is useful to model the image plane in a pin-hole camera model. Consider a collection of planar points P_1, \dots, P_k in space living on a plane π viewed from two views. The projections of P_i are p_i, p'_i in views 1,2 respectively. Because the collineations form a group, there exists a unique homography matrix H_π that satisfies the relation $H_\pi p_i \cong p'_i$, $i = 1, \dots, k$, and where H_π is uniquely determined by 4 matching pairs from the set of k matching pairs. Moreover, $H_\pi^{-T} s \cong s'$ will map between matching lines s, s' arising from 3D lines living in the plane π . Likewise, $H_\pi^\top s' \cong s$ will map between matching lines from view 2 to view 1.

It will be most convenient to use tensor notations from now on because the material we will be using in this chapter involves coupling together

pairs of collineations into a “joint” object. The distinction of when coordinate vectors stand for points or hyperplanes matters when using tensor notations. A point is an object whose coordinates are specified with superscripts, i.e., $p = (p^0, p^1, \dots, p^n)$, thus p^i stands for the i 'th entry of the vector. These are called contravariant vectors. A hyperplane in \mathcal{P}^n is called a covariant vector and is represented by subscripts, i.e., $s = (s_0, s_1, \dots, s_n)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $p^i s_i = p^0 s_0 + p^1 s_1 + \dots + p^n s_n$. This is known as a contraction. For example, if p is a point incident to a line s in \mathcal{P}^2 , then $p^i s_i = 0$.

Vectors are also called 1-valence tensors. 2-valence tensors (matrices) have two indices and the transformation they represent depends on the covariant-contravariant positioning of the indices. For example, a_i^j is a mapping from points to points (a collineation, for example), and hyperplanes to hyperplanes, because $a_i^j p^i = q^j$ and $a_i^j s_j = r_i$ (in matrix form: $Ap = q$ and $A^T s = r$); a_{ij} maps points to hyperplanes; and a^{ij} maps hyperplanes to points. When viewed as a matrix the row and column positions are determined accordingly: in a_i^j and a_{ji} the index i runs over the columns and j runs over the rows, thus $b_j^k a_i^j = c_i^k$ is $BA = C$ in matrix form.

An outer-product of two 1-valence tensors (vectors), $a_i b^j$, is a 2-valence tensor c_i^j whose i, j entries are $a_i b^j$ — note that in matrix form $C = ba^T$. A 3-valence tensor has three indices, say H_i^{jk} . The positioning of the indices reveals the geometric nature of the mapping: for example, $p^i s_j H_i^{jk}$ must be a point because the i, j indices drop out in the contraction process and we are left with a contravariant vector (the index k is a superscript). Thus, H_i^{jk} maps a point in the first coordinate frame and a hyperplane in the second coordinate frame into a point in the third coordinate frame. A single contraction, say $p^i H_i^{jk}$, of a 3-valence tensor leaves us with a matrix. Note that when p is $(1, 0, 0)$ or $(0, 1, 0)$, or $(0, 0, 1)$ the result is a “slice” of the tensor.

In the projective plane \mathcal{P}^2 we will make use of the “cross-product tensor” ϵ defined next. The cross product (vector product) operation $c = a \times b$ is defined for vectors in \mathcal{P}^2 . The product operation can also be represented as the product $c = [a]_{\times} b$ where $[a]_{\times}$ is called the “skew-symmetric matrix of a ” and has the form:

$$[a]_{\times} = \begin{pmatrix} 0 & -a_2 & a_1 \\ a_2 & 0 & -a_0 \\ -a_1 & a_0 & 0 \end{pmatrix}$$

In tensor form we have $\epsilon_{ijk} a^i b^j = c_k$ representing the cross product of two points (contravariant vectors) resulting in the line (covariant vector) c_k . Similarly, $\epsilon^{ijk} a_i b_j = c^k$ represents the point intersection of the two lines a_i and b_j . The tensor ϵ is defined such that $\epsilon_{ijk} a^i$ produces the matrix $[a]_{\times}$ (i.e., ϵ contains 0, -1 , 1 in its entries such that its operation on a single vector produces the skew-symmetric matrix of that vector).

2 Homography Tensor \mathcal{H} of the Projective Plane

Consider some plane π whose features (points or lines) are projected onto three views and let A be the collineation from view 2 to view 1, and B the collineation from view 3 to 1 (we omit the reference to π in our notation). Let P be some point on the plane π and its projections are p, p', p'' in views 1,2,3 respectively. We consider two possibilities: (i) the point P on the plane π is *stationary*, i.e., the three optical rays from the camera centers to the image points p, p', p'' meet at P , and (ii) the point P *moves* along a straight line (in the plane) path, therefore the three optical rays meet at a line in π instead of a point (see Fig.1). We summarize these two possibilities in the following definition:

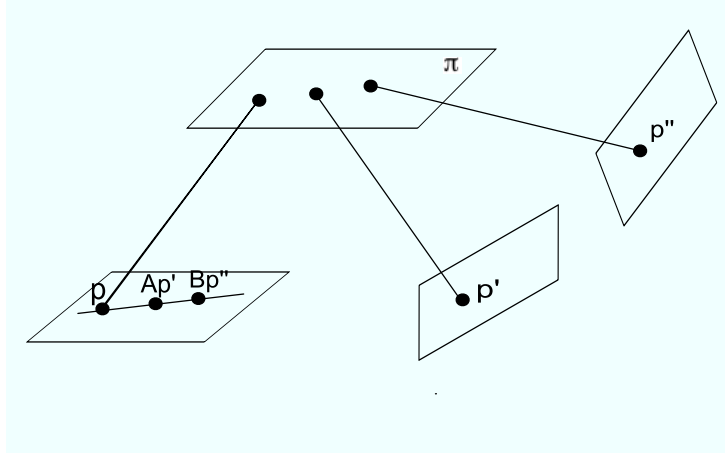


Fig. 1. The homography tensor of \mathcal{P}^2 and moving points. The collineations A, B are from view 2 to 1 and 3 to 1 respectively. If the triplet p, p', p'' are projections of a moving point along a line on π then p, Ap', Bp'' are collinear in view 1. Thus, $p^\top (Ap' \times Bp'') = 0$, or $p^i p'^j p''^k \mathcal{H}_{ijk} = 0$ where $H_{ijk} = \epsilon_{inu} a_j^n b_k^u$.

Definition 1. A triplet of points p, p', p'' are said to be **matching with respect to a stationary point** if they are matching in the usual sense of the term, i.e., the corresponding optical rays meet at a single point. The triplet are said to be **matching with respect to a moving point** if the three optical rays meet at a line on a plane.

The constraint which satisfies both the moving and stationary possibilities is:

$$\det(p, Ap', Bp'') = p^\top (Ap' \times Bp'') = 0.$$

In other words, $\det(p, Ap', Bp'') = 0$ when the rank of the 3×3 matrix $[p, Ap', Bp'']$ is either 1 or 2. The rank is 1 when the point P is stationary

(three optical rays meet at a point) and is 2 when the point P moves along a straight line path (three optical rays meet at a line on π). The constraint $p^\top(Ap' \times Bp'') = 0$ is bilinear in the entries of the unknown collineations A, B and is trilinear in the observations p, p', p'' . Using tensorial notations we can combine the pair of collineations into a single object, a $3 \times 3 \times 3$ tensor, as follows. We define indices i, j, k such that index i runs over view 1, index j runs over view 2 and index k runs over view 3. For example, the operation Ap' is translated to $a_j^i p'^j$ producing a point in view 1. The cross product $Ap' \times Bp''$ is translated to $\epsilon_{inu}(a_j^n p'^j)(b_k^u p''^k)$ where parenthesis are added for clarity only (position of symbols are not important, only position of indices). Taken together we have:

$$p^i \epsilon_{inu}(a_j^n p'^j)(b_k^u p''^k) = 0.$$

After re-arranging the symbol positions we obtain:

$$p^i p'^j p''^k (\epsilon_{inu} a_j^n b_k^u) = 0,$$

where the object in parenthesis is the *homography tensor* of \mathcal{P}^2 , referred to as Htensor:

$$\mathcal{H}_{ijk} = \epsilon_{inu} a_j^n b_k^u \quad (1)$$

whose triple contraction $p^i p'^j p''^k \mathcal{H}_{ijk}$ vanishes on observations p, p', p'' arising from stationary or moving points on the plane π . Each such triplet of matching points provides a linear constraint on the 27 entries of \mathcal{H} , thus 26 matching triplets are necessary to solve for \mathcal{H} uniquely (up to scale).

We see from the above that the tensor \mathcal{H} applies to both stationary and moving points coming from the planar surface π . The possibility of working with stationary and moving elements was first introduced in [1, 2] where it was shown that if a moving point along a general (in 3D) straight path is observed in 5 views, and the camera projection matrices are *known*, then it is possible to set up a linear system for estimating the 3D line. With the Htensor \mathcal{H} , on the other hand, we have no knowledge of the camera projection matrices, but on the other hand we require that the straight paths the points are taking should all be coplanar (what makes it possible to work with 3 views instead of 5 and not require prior information on camera positions). We will address the following issues:

- What are the minimal point configurations that allow a unique solution for \mathcal{H} ? If all points are moving then 26 of them are needed and we will address the issue of necessary point-set configuration. If some of the points are *known* to be stationary how many constraints (i.e. moving points) are minimally necessary for a unique solution? if some of the points are stationary (without the system being told about it) what would be the minimal number of moving points required for a unique solution?
- Contraction properties of \mathcal{H} and the manners in which \mathcal{H} acts as a mapping.
- How to recover the component collineations A, B from \mathcal{H} .

2.1 Recovering Htensor from Image Measurements

The measurements available for recovering \mathcal{H} are triplets of matching points p, p', p'' across the three views and *prior* information whether a triplet arises from a moving or stationary point. Assuming first that all measurements are induced by moving points, a triplet of matching points contribute one linear constraint $p^i p'^j p''^k \mathcal{H}_{ijk} = 0$ on the elements of \mathcal{H} . Therefore 26 triplets are necessary for a unique² solution. The 26 points should be distributed on the plane π in such a way that they cover at least 4 lines in general position, such that no more than 8 points are on the first line, no more than 7 points on the second line, no more than 6 on the third line and no more than 5 on the fourth line. This distribution will guarantee a unique solution for \mathcal{H} (of course more than 26 points are allowed where in that case a least-squares approximation is recovered).

Theorem 1. *A minimal configuration of 26 matching triplets arising from moving points on π are necessary for a unique recovery of \mathcal{H} provided that the distribution of the points are such that the motion trajectories cover at least 4 lines in general position on π and that no more than 8 of the points lie on the first trajectory, no more than 7 on the second trajectory, no more than 6 on the third and no more than 5 points lie on the fourth line trajectory.*

Proof: Consider a line L_1 on the plane π . Let the projections of L onto the three views be denoted by q_1, s_1, r_1 (see Fig. 2). Since each line is determined by two points, we can have at most $2^3 = 8$ linearly independent constraints of the form $p^i p'^j p''^k \mathcal{H}_{ijk} = 0$ where the points p, p', p'' are coincident with the lines q_1, s_1, r_1 respectively. Consider a second line $L_2 \in \pi$ projecting onto lines q_2, s_2, r_2 . Since each of the image lines is spanned by two points, choose one of those points to be the projection of $L_1 \cap L_2$ denoted by p, p', p'' . Among the 8 choices of choosing three points from the three pairs of points, the choice p, p', p'' is already covered by the span of the 8 constraints induced by L_1 — thus we are left with 7 linearly independent constraints in \mathcal{H} . This argument continues by induction over additional lines L_i each inducing one less constraint than the one before it. The process ends with 4 lines inducing $8 + 7 + 6 + 5 = 26$ linearly independent constraints. \square

Next, we consider the contribution of stationary points to the system of linear equations for \mathcal{H} . A stationary point, *known* as such (referred to as *labeled*), contributes 9 linear constraints of rank 7, as follows: let p, p', p'' be a triplet of matching points arising from a known stationary point on π . The rank of the matrix $[p, Ap', Bp'']$ is 1, which in turn translates to the three sets of constraints: $p \times Ap' = 0, p \times Bp'' = 0$ and $Ap' \times Bp'' = 0$. In tensor form, the contractions $p^i p'^j \mathcal{H}_{ijk}, p^i p''^k \mathcal{H}_{ijk}$ and $p'^j p''^k \mathcal{H}_{ijk}$ are null vectors. The 9 constraints are explicitly written below (allow the vector e to vary over the standard basis $(1, 0, 0), (0, 1, 0)$ and $(0, 0, 1)$):

² The dimension of the $\text{GL}(V)$ module $\text{Span}\{p \otimes p' \otimes p'' \in V^{\otimes 3} : \dim \text{Span}\{p, p', p''\} = 2\}$ is 26. The details are in Section 4.

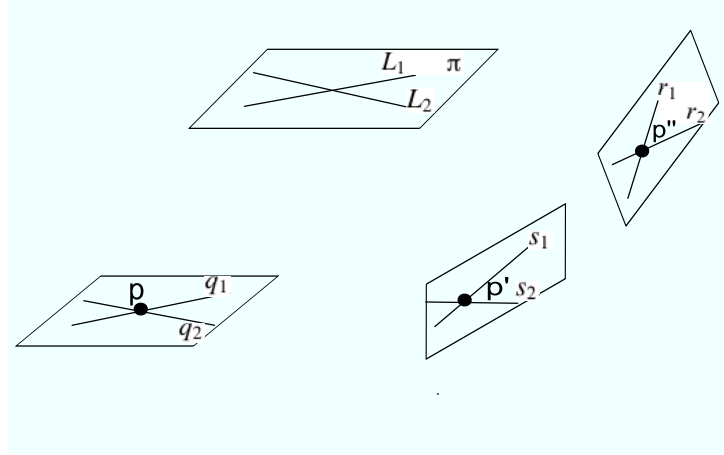


Fig. 2. A straight line path L_1 can induce at most 8 independent linear constraints as the projections q_1, s_1, r_1 in the three views are determined by two points each. A second straight line path L_2 can contribute at most 7 independent constraints since the constraint $p^i p'^j p''^k \mathcal{H}_{ijk} = 0$ induced by the projection of the intersection $L_1 \cap L_2$ onto p, p', p'' is spanned by the 8 constraints from L_1 .

$$\begin{aligned}
 p^i p'^j e^k \mathcal{H}_{ijk} &= 0 \quad \forall e \\
 p^i e^j p''^k \mathcal{H}_{ijk} &= 0 \quad \forall e \\
 e^i p'^j p''^k \mathcal{H}_{ijk} &= 0 \quad \forall e
 \end{aligned} \tag{2}$$

Note that the constraint $p^i p'^j p''^k \mathcal{H}_{ijk} = 0$ is in the span of the three sets of constraints — thus making a total of 7 linearly independent constraints (a system of 9 linear equations of rank 7). We thus arrive to the conclusion:

Proposition 1. *The matching triplets induced by four labeled stationary points in general position on π provide a unique solution for \mathcal{H} .*

We consider next the contribution of *unlabeled* stationary points. A stationary point can provide 9 constraints (of rank 7) provided it is known to be stationary — otherwise it provides only a single constraint. Consider the case where all the measurements arise from unlabeled stationary points. It is easy to see that the rank of the estimation matrix for \mathcal{H} is at most 10 (compared to 26 when moving points are used). Each row of the estimation matrix for \mathcal{H} is some “constraint tensor” G^{ijk} such that $G^{ijk} \mathcal{H}_{ijk} = 0$. It is sufficient to prove this statement for the case where $A = B = I$ (the identity matrix) — because all other cases are transformed into this one by local change of coordinates. In the case $A = B = I$, G^{ijk} is a symmetric tensor, i.e., remains the same under permutation of indices — hence contains only 10 different groups of indices

111, 222, 333, 112, 113, 221, 223, 331, 332, 123

up to permutations. Generally speaking, the m -fold symmetric powers $\text{Sym}^m V$ of an n -dimensional vector space V is a vector space of dimension $\binom{n+m-1}{m}$ (substitute $n = 3, m = 3$ to get 10). We arrive to the following conclusion:

Proposition 2. *In a collection of unlabeled matching triplets, there could be at most 10 of which are induced by stationary points. In other words, there should be at least 16 moving points an an input collection of unlabeled points for a unique linear solution for \mathcal{H} .*

Finally, we consider the situation of a *mixed* labeled and unlabeled triplets. Consider the case where $x \leq 4$ of the triplets are labeled as arising from stationary points. We saw above that a labeled stationary point is equivalent to 7 constrains, however some of those constraints may be already included in the span of the unlabeled stationary points. The theorem below addresses the question of how many matching triplets arising from moving points are necessary given that $x \leq 4$ matching triplets are labeled as stationary? Clearly, when $x = 4$ there is no need for further measurements, but when $x < 4$ we obtain the following result:

Theorem 2. *In a situation of matching triplets arising from a mixture of stationary and moving points, let $x \leq 4$ be the number of matching triplets that are known a priori to arise from stationary points. To obtain a unique linear solution for \mathcal{H} , the minimal number of matching triplets arising from moving points is $16 - 4x$ and at most $10 - 3x$ can be (unlabeled) stationary points.*

Proof: Each row of the estimation matrix for \mathcal{H} is some “constraint tensor” G^{ijk} such that $G^{ijk}H_{ijk} = 0$. It is sufficient to prove this statement for the case where $A = B = I$ (the identity matrix) — because all other cases are transformed into this one by local change of coordinates. Therefore, a stationary point induces a symmetric tensor $G^{ijk} = p^i p^j p^k$. The case $x = 0$ was discussed above with the conclusion that a minimal of 16 moving points are required.

Consider the case $x = 1$, i.e., one of the matching triplets contributed 9 constraints of rank 7:

$$\begin{aligned} p^i p^j e_1^k H_{ijk} = 0 & \quad p^i e_1^j p^{''k} H_{ijk} = 0 & \quad e_1^i p^j p^{''k} H_{ijk} = 0 \\ p^i p^j e_2^k H_{ijk} = 0 & \quad p^i e_2^j p^{''k} H_{ijk} = 0 & \quad e_2^i p^j p^{''k} H_{ijk} = 0 \\ p^i p^j e_3^k H_{ijk} = 0 & \quad p^i e_3^j p^{''k} H_{ijk} = 0 & \quad e_3^i p^j p^{''k} H_{ijk} = 0, \end{aligned}$$

where e_1, e_2, e_3 are the standard basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Add the three constraints in the first row:

$$E^{ijk} = p^i p^j e_1^k + p^i e_1^j p^k + e_1^k p^j p^k$$

Then, E^{ijk} is a symmetric tensor and thus spanned by the 10-dimensional subspace of the unlabeled stationary points. Likewise, the constraint tensors

resulting from adding the constraint of the second and third row above are also symmetric. Taken together, 3 out of the 7 constraints contributed by a labeled stationary point are already accounted for by the space of unlabeled stationary points. Therefore, each labeled stationary point adds only 4 linearly independent constraints. \square

2.2 Contraction Properties of \mathcal{H} and Recovery of A, B

We turn our attention next to single and double contractions of the Htensor — what can be extracted from them and what is their geometric significance. Those contractions will hold the key for decoupling the collineations A, B from \mathcal{H} .

The double contractions perform mapping operations. Consider for example $p^i p'^j H_{ijk}$, which by the index arrangements, must be a contravariant vector (a line in \mathcal{P}^2) denoted by l'' . Since the remaining index is k , l'' is a line in view 3. Consider the line $L \in \pi$ defined by the projection p, p' in views 1,2. Since $p^i p'^j p''^k H_{ijk} = 0$ for all points p'' in view 3 which are the projections from L , we conclude that l'' is the projection of L onto view 3.

The single contractions produce matrices which form the key for decoupling the collineations A, B from \mathcal{H} . Consider, for example, $\delta^k H_{ijk}$ for some contravariant vector (a point in view 3) δ . The result is a matrix E with index structure suggesting it maps points to lines (a correlation matrix) and between views 1,2. By substitution in the definition of \mathcal{H} we obtain:

$$\delta^k H_{ijk} = \epsilon_{inu} a_j^n (b_k^u \delta^k) = [B\delta]_{\times} A.$$

Let $E = [B\delta]_{\times} A$ and note that the point $\mu = B\delta$ is the matching point to δ in view 1, i.e., it is the projection onto view 1 of the point defined by the intersection of the plane π with the optical ray associated with δ (see Fig. 3). The matching points to δ , the points $\mu = B\delta$ and $\eta = A^{-1}B\delta$, can be recovered directly from E since:

$$\begin{aligned} E^{\top} \mu &= -A^{\top} [\mu]_{\times} \mu = 0 \\ E \eta &= [\mu]_{\times} A \eta \cong [\mu]_{\times} \mu = 0. \end{aligned}$$

The matrix E forms a point to line mapping from view 2 to view 1, as follows. Consider any point p' in view 2, then $E p' = p'^j \delta^k \mathcal{H}_{ijk}$ is the projection of the line in π , defined by the optical rays associated with δ and p' , onto view 1. Therefore, any point p coincident with the projected line will satisfy $p^{\top} E p' = 0$. We conclude that the bilinear form $p^{\top} E p' = 0$ is satisfied for all pairs of p, p' which are on matching lines through the fixed points μ, η (see Fig. 3).

Finally, the collineation A can be recovered from single contractions by the fact that $A^{\top} E$ is a skew-symmetric matrix:

$$A^{\top} E + E^{\top} A = A^{\top} [\mu]_{\times} A - A^{\top} [\mu]_{\times} A = 0,$$

which provides 6 linearly independent equations on the entries of A . By taking δ to range over the standard basis $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ we obtain three slices of \mathcal{H} denoted by E_1, E_2, E_3 each producing 6 linear equations on A — taken together A can be recovered linearly from the slices of \mathcal{H} . Likewise, B can be recovered from the slices $\delta^j \mathcal{H}_{ijk}$ in the same manner, and the collineation $A^{-1}B$ (between views 2 and 3) from the slices $\delta^i \mathcal{H}_{ijk}$. These findings are summarized in the theorem below:

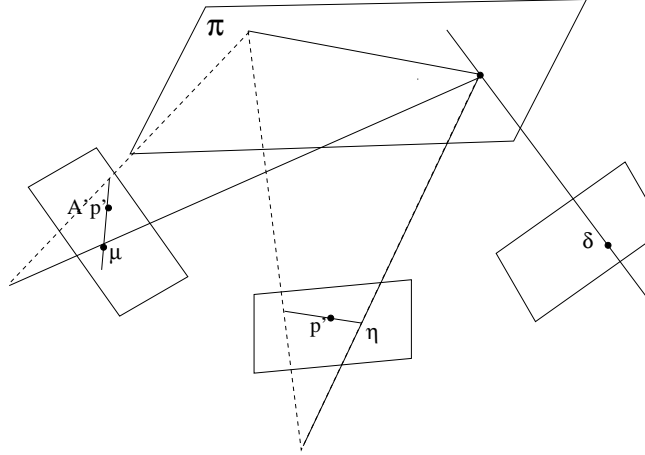


Fig. 3. A single contraction, say $\delta^k \mathcal{H}_{ijk}$, is a mapping E between views 1,2 from points to concurrent lines. The null spaces of E and E^T are the matching points μ, η of δ in views 1,2. The image points p' are mapped by E to the lines $Ap' \times \mu$ and the image points p are mapped by E^T to the lines $A^{-1}p \times \eta$ in view 2. The bilinear relation $p^T E p' = 0$ is satisfied for all pairs of p, p' on matching lines through the fixed points μ, η .

Theorem 3. *Each of the contractions*

$$\delta^k \mathcal{H}_{ijk} \tag{3}$$

$$\delta^j \mathcal{H}_{ijk} \tag{4}$$

$$\delta^i \mathcal{H}_{ijk} \tag{5}$$

represents a point-to-line (correlation) mapping between views $(1, 2)$, $(1, 3)$ and $(2, 3)$ respectively. By setting δ to be $(1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$ we obtain three different slicings of the tensor: denote the slices of $\delta^i \mathcal{H}_{ijk}$ by the matrices G_1, G_2, G_3 , the slices of $\delta^j \mathcal{H}_{ijk}$ by the matrices W_1, W_2, W_3 , and the slices of $\delta^k \mathcal{H}_{ijk}$ by the matrices E_1, E_2, E_3 . Then these slices provide sufficient (and over-determined) linear constraints for the constituent homography matrices A, B and for $C = A^{-1}B$:

$$CG_i^\top + G_iC^\top = 0, \quad (6)$$

$$BW_i^\top + W_iB^\top = 0, \quad (7)$$

$$AE_i^\top + E_iA^\top = 0, \quad (8)$$

for $i = 1, 2, 3$.

In summary, the homography tensor in \mathcal{P}^2 applies to both cases: optical rays meet at a single point (matching points with respect to a stationary point) and optical rays meet at a line on π (matching points with respect to a moving point). In the case where no distinction can be made to the source of a matching triplet p, p', p'' (stationary or moving) then we have seen that in a set of at least 26 such matching triplets, 16 of them *must* arise from moving points. In case that a number $x \leq 4$ of these triplets are *known a-priori* to arise from stationary points, then $16 - 4x$ must arise from moving points. Once \mathcal{H} is recovered from image measurements it forms a mapping of both moving and stationary points and in particular can be used to *distinguish* between moving and stationary points (a triplet p, p', p'' arising from a stationary point is mapped to null vectors $p^i p'^j H_{ijk}, p^i p''^k H_{ijk}$ and $p'^j p''^k H_{ijk}$). The Htensor can be useful in practice to handle situations rich in dynamic motion seen from a monocular sequence — some experiments are shown in Section 6.

We will next describe the homography tensors of \mathcal{P}^3 where points lie in the 3D projective space, the collineations which are responsible for the coordinate changes are 4×4 matrices and the points are allowed to move along straight lines or planar subspaces while coordinate changes take place.

3 Homography Tensors of \mathcal{P}^3

We consider stepping up one dimension, namely, the point configuration lies in \mathcal{P}^3 the collineations are 4×4 matrices and the dimension in which the points are allowed to move while the global collineations take place are $k = 1, 2, 3$ where $k = 1$ stands for stationary points, $k = 2$ stands for motion along a straight line path and $k = 3$ stand for motion along a planar subspace. We will focus below on the constraint of straight line motion and stationary points $k = 2, 1$ which induce a $4 \times 4 \times 4$ homography tensor. The situation of planar dynamic motion $k = 3$ induces a 4^4 tensor which we will not consider in detail here and leave it for the discussion on general dynamic alignment in Section 4.

Let X be some stationary point in 3D space with coordinate vector P . Let P' be the coordinate representation of the point X at some other time instant (i.e., the measurement sensor has changed its viewing position) and let P'' be the coordinate representation of X at a third time instant. Let A, B be the collineations mapping the second and third coordinate representations back to the first representation, i.e., $P \cong AP'$ and $P \cong BP''$.

If the point X happens to *move* along some straight-line path during the change of coordinate systems, then P, AP', BP'' do not coincide but they form a rank-2 matrix (see Fig. 4):

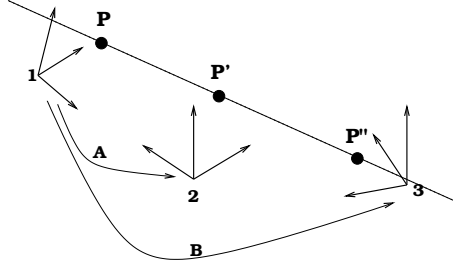


Fig. 4. The points P, P' and P'' are measured at three time instants from different viewing positions of the sensor, i.e., each point is given in a different coordinate system. While the measuring device changes position, the physical point in space moves along a straight line path. In other words, the rank of the 4×3 matrix $[P, AP', BP'']$ is 2 for a moving point and 1 for a stationary point. The 4×4 matrices A, B are responsible for the change of coordinate system back to the starting position.

$$\text{rank} \begin{pmatrix} | & | & | \\ P & AP' & BP'' \\ | & | & | \end{pmatrix} = 2$$

And for every column vector V we have

$$\det \begin{pmatrix} | & | & | & | \\ P & AP' & BP'' & V \\ | & | & | & | \end{pmatrix} = 0 \quad (9)$$

Note that because V is spanned by a basis of size four, we can obtain at most four linearly independent constraints on some object consisting of A, B from a triplet of matching points P, P', P'' . Note also that the null vector of a 4×3 matrix can be represented by the 3×3 determinant expansion. For example, let X, Y, Z be three column vectors in a 4×3 matrix, then the vector $W = (w_1, \dots, w_4)$ representing the plane defined by the points X, Y, Z is

$$\begin{aligned} w_1 &= \det \begin{pmatrix} x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix} & w_2 &= -\det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix} \\ w_3 &= \det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_4 & y_4 & z_4 \end{pmatrix} & w_4 &= -\det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} \end{aligned}$$

We can write the relationship between W and X, Y, Z as a tensor operation as follows:

$$w_i = \epsilon_{ijkl} x^j y^k z^l$$

where the entries of ϵ consist of $+1, -1, 0$ in the appropriate places. We will refer to ϵ as the “cross-product” tensor. Note that the determinant of a 4×4 matrix whose columns consist of $[X, Y, Z, T]$ can be compactly written as

$$t^i x^j y^k z^l \epsilon_{ijkl}.$$

Using the cross-product tensor we can write the constraint (9) as follows:

$$\begin{aligned} 0 &= \det \begin{pmatrix} | & | & | & | \\ P & AP' & BP'' & V \\ | & | & | & | \end{pmatrix} \\ &= P^i (\epsilon_{ilmu} (a_j^l P'^j) (b_k^m P''^k) v^u) \\ &= P^i P'^j P''^k (\epsilon_{ilmu} a_j^l b_k^m v^u) \end{aligned}$$

Note that the tensor form allows us to separate the measurements P, P', P'' from the unknowns A, B (and vector V), and we denote the expression in parentheses as:

$$\mathcal{J}_{ijk} = \epsilon_{ilmu} a_j^l b_k^m v^u \quad (10)$$

as the the homography tensor of \mathcal{P}^3 . Note that for every choice of the vector V we get an Htensor. As previously mentioned, since V is spanned by a basis of dimension four there are at most four such tensors; each tensor is defined by the constraints

$$P^i P'^j P''^k \mathcal{J}_{ijk} = 0.$$

These are linear constraints on the 64 elements of the Htensor. Since there are four Htensors compatible with the observations, the linear system of equations for solving for \mathcal{J} from the matching triplets P, P', P'' has a four-dimensional null space. The vectors of the null space are spanned by the Htensors. In practical terms, given $N \geq 60$ matching triplets P, P', P'' , each triplet contributes one linear equation $P^i P'^j P''^k \mathcal{J}_{ijk} = 0$ for the 64 entries of \mathcal{J} . The eigenvectors associated with the four smallest eigenvalues of the estimation matrix are *the Htensors of the dynamic 3D-to-3D alignment problem*. We summarize this in the following theorem:

Theorem 4 (Htensors in \mathcal{P}^2). *Each matching triplet P, P', P'' arising from a dynamic point contributes one linear equation $P^i P'^j P''^k \mathcal{J}_{ijk} = 0$ to a $4 \times 4 \times 4$ tensor \mathcal{J} . Any $N \geq 60$ matching triplets in general position provide an estimation matrix for \mathcal{J}_{ijk} with a four-dimensional null space. The 60 points should be distributed along at least 10 lines, five of which can hold up to eight dynamic points, and the remaining five up to four dynamic points.*

In the remainder of this section we will discuss (i) tensor slices and the extraction of the constituent collineations A, B from the four Htensors, (ii) the use of Htensors for direct mapping between coordinate systems (without extracting A, B along the way), (iii) the use of Htensors to distinguish between dynamic and stationary points, and (iv) the relationship between the number of stationary and dynamic points for estimating the Htensors in unsegmented and segmented configurations.

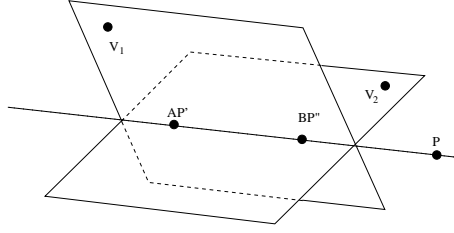


Fig. 5. The points AP', BP'' and V define a plane. AP', BP'' and V' define another plane. The line of intersection of these planes contains P .

3.1 Tensor Slices and the Extraction of the Collineations A, B

The role of \mathcal{J} is symmetric with respect to the position of the points P, P', P'' (this is true for every purely covariant or contravariant tensor, unlike the mixed covariant-contravariant tensor). It is therefore sufficient to investigate $P'^j P''^k \mathcal{J}_{ijk}$ as one of the tensor double-contractions; the others, $P^i P''^k \mathcal{J}_{ijk}$ and $P^i P'^j \mathcal{J}_{ijk}$, follow by symmetry.

Consider any Htensor with its associated vector V . Recall that from observations we can recover four Htensors which span the null space of the measurement matrix — each Htensor has a different vector V associated with it. We will describe next how to recover the vector V , referred to as the "principal point" of the tensor, from the Htensor.

Consider the plane π defined by $\pi_i = P'^j P''^k \mathcal{J}_{ijk}$ and which contains the three points V, AP' and BP'' :

$$\pi_i = P'^j P''^k \mathcal{J}_{ijk} = \epsilon_{ilmu} (a_j^l P'^j) (b_k^m P''^k) v^u,$$

which by definition of the cross-product tensor provides the plane associated with the three points acted upon by ϵ . By varying P' and P'' we obtain a star of planes all coincident with the point V . As a result, the principal point V of the tensor can be recovered by taking three double slices of the tensor and finding their intersection.

We next recover the line in space coincident with the points AP' and BP'' . Consider two Htensors denoted by \mathcal{J}^1 and \mathcal{J}^2 (recall that we have four Htensors at our disposal). The intersection of the planes $P'^j P''^k \mathcal{J}_{ijk}^1$ and $P'^j P''^k \mathcal{J}_{ijk}^2$ is the line passing through AP' and BP'' (see Fig. 5).

The collineations A, B can be recovered (linearly) from the matrices resulting from single contractions of the Htensors. A single contraction $H_{ij} = P''^k \mathcal{J}_{ijk}$ is a 4×4 matrix H that maps points to planes. As mentioned above, $P'^j H_{ij} = P'^j P''^k \mathcal{J}_{ijk}$ is the plane passing through V, AP', BP'' ; thus by varying P' one obtains a pencil of planes coincident with the line through V and BP'' . Hence the rank of the matrix H must be 2.

Because HP' is the plane through V, AP', BP'' , we have $P'^\top A^\top HP' = 0$ for every choice of P' . Therefore $A^\top H$ is a skew-symmetric matrix and thus

provides ten linear constraints for A . By varying P'' and thus obtaining other H-matrices $P''^k \mathcal{J}_{ijk}$ we can obtain more constraints on A but this is not sufficient to obtain a unique solution for A . A unique solution requires the H-matrix of at least another Htensor because the principal point must vary as well. Likewise, one can recover B from the contractions $P'^j \mathcal{J}_{ijk}$ by varying P' and taking at least two Htensors.

3.2 Direct Mapping

We can use the Htensor to map points between the coordinate frames without the need to extract the collineations A and B .

Consider for example the direct mapping $P \cong BP''$ between the third and the first coordinate frames. The contraction $\gamma^j P''^k \mathcal{J}_{ijk}$ for some arbitrary vector γ is a plane in 3D containing the points BP'' , $A\gamma$ and V (the principal point of \mathcal{J}), all represented in the first coordinate frame. By varying γ over the standard basis, and taking the four different Htensors (so that V also varies), we get a collection of 16 planes. These planes intersect in the point $P \cong BP''$. It is sufficient to use a subset of these planes (at least three) as long as not all of them are generated using the same Htensor or the same γ .

As a result, the Htensor can play the same role as a collineation (i.e., direct) mapping between coordinate frames. The direct mapping can be used, for example, to distinguish between *stationary* and *dynamic* points. If P is equal to the direct mapping BP'' , then the corresponding physical point X is stationary; otherwise (ignoring noise considerations), X is dynamic.

The segmentation of stationary and dynamic points can be achieved in other ways as well. For example, from (9) we know that for a stationary point X with coordinate vectors P, P' and P'' in the three frames, any double contraction vanishes:

$$P^i P'^j \mathcal{J}_{ijk} = P'^j P''^k \mathcal{J}_{ijk} = P^i P''^k \mathcal{J}_{ijk} = 0.$$

Hence a vanishing double contraction (under all three possibilities) indicates a stationary point. In practice, since the double contraction provides only an algebraic (rather than geometric) measure of error, better segmentation results are achieved by measuring the distance between the point P and the direct mapping BP'' .

3.3 Constraints from Stationary Points

We have seen that a matching triplet P, P' and P'' satisfies the Htensor constraint

$$P^i P'^j P''^k \mathcal{J}_{ijk} = 0$$

regardless of whether the corresponding physical point X is moving along a straight-line path (dynamic) or is stationary. For a dynamic point, the rank of the 4×3 matrix $[P, AP', BP'']$ is 2 and for a stationary point the rank is

1. In other words, admissible measurements for recovering the Htensors come from dynamic and stationary points alike. The natural question is, how much alike? — i.e., can all the measurements arise (unknowingly) from stationary points? If not, what is the maximal number number of stationary points after which the contributions of additional stationary points become redundant? These questions are exactly the same as those addressed for \mathcal{H} in the context of \mathcal{P}^2 .

The contribution of unlabeled stationary points, i.e., recovering \mathcal{J} from constraints $P^i P'^j P''^k \mathcal{J}_{ijk} = 0$ where the triplet P, P', P'' are induced by stationary points only, can fill up a 20 dimensional subspace only (out of 60). Without loss of generality we can assume that $A = B = I$ which in turn makes each constraint $G^{ijk} \mathcal{J}_{ijk} = 0$ where $G^{ijk} = P^i P^j P^k$ is a symmetric tensor (remains the same under permutation of indices). The 3-fold symmetric powers $\text{Sym}^3 V$ of a 4-dimensional vector space V is $\binom{4+3-1}{3} = 20$. In other words, there are only 20 different groups of indices:

$$111, 222, 333, 444, 112, 113, 114, 221, 223, 224,$$

$$331, 332, 334, 441, 442, 444, 123, 124, 134, 234.$$

This analysis is summarized in the theorem below:

Theorem 5. *The constraints $P^i P'^j P''^k \mathcal{J}_{ijk} = 0$ made solely from stationary points span at most a 20-dimensional space.*

Consequently, in the unsegmented situation when stationary and dynamic points are treated alike, it is not possible to obtain a unique solution from stationary points alone; one needs at least 40 dynamic points in the collection of $N \geq 60$ matching triplets. We consider next the contribution arising from *labeled* stationary points, i.e., how many constraints would a triplet P, P', P'' contribute if it were known that the corresponding physical point X is stationary? In this case, for every $\delta_{4 \times 1}$ and for every V , the determinant

$$\det \begin{pmatrix} | & | & | & | \\ P & AP' & B\delta & V \\ | & | & | & | \end{pmatrix}$$

vanishes. Since this is true for every pair of the three points, then for each of the four Htensors we get:

$$\begin{aligned} P^i P'^j e_1^k \mathcal{J}_{ijk} = 0 & \quad P^i e_1^j P''^k \mathcal{J}_{ijk} = 0 & \quad e_1^i P'^j P''^k \mathcal{J}_{ijk} = 0 \\ P^i P'^j e_2^k \mathcal{J}_{ijk} = 0 & \quad P^i e_2^j P''^k \mathcal{J}_{ijk} = 0 & \quad e_2^i P'^j P''^k \mathcal{J}_{ijk} = 0 \\ P^i P'^j e_3^k \mathcal{J}_{ijk} = 0 & \quad P^i e_3^j P''^k \mathcal{J}_{ijk} = 0 & \quad e_3^i P'^j P''^k \mathcal{J}_{ijk} = 0 \\ P^i P'^j e_4^k \mathcal{J}_{ijk} = 0 & \quad P^i e_4^j P''^k \mathcal{J}_{ijk} = 0 & \quad e_4^i P'^j P''^k \mathcal{J}_{ijk} = 0, \end{aligned} \tag{11}$$

where e_1, e_2, e_3, e_4 are the standard basis $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)$. Note that the constraint

$$P^i P'^j P''^k \mathcal{J}_{ijk} = 0$$

can be spanned by each row separately, hence the rank of the above system is at most 10. We thus arrive to the conclusion:

Theorem 6. *A labeled stationary point can provide at most 10 linearly independent constraints for the solution of \mathcal{J} .*

These constraints came from one stationary point, but how many of them are spanned by the subspace of constraints obtained from unlabeled stationary points? This question is answered next:

Theorem 7. *Out of the ten linearly independent constraints arising from a labeled stationary point, four lie in the rank-20 subspace spanned by unlabeled stationary points and six lie in the subspace spanned only by dynamic points.*

Proof: Again, it is sufficient to prove this theorem for the case where $A = B = I$. In this case a stationary point satisfies $P \cong P' \cong P''$.

We look at the 12 constraints of rank 10 described in (11). Adding the three constraints in the first row gives

$$G^{ijk} = P^i P^j e_1^k + P^i e_1^j P^k + e_1^k P^j P^k$$

which is a symmetric tensor and thus is spanned by the 20-dimensional subspace of the unlabeled stationary points.

Similarly, the constraint tensors resulting from adding the other three rows are also symmetric. One can verify that except for those four constraints (and the ones they span) there are no other symmetric constraints.

Taken together, four out of the ten constraints contributed by a labeled stationary point lie in the subspace of unlabeled stationary points and six constraints lie in the subspace of dimension 40 spanned by dynamic points. \square

As a corollary, we can deduce that 7 labeled stationary points are necessary to fill up the 60 dimensional subspace necessary for a solution for \mathcal{J} . Since the 10 constraints contributed by a labeled stationary point include 4 which are spanned by the subspace of unlabeled stationary points, then 5 labeled stationary points will fill up the 20-dimensional subspace of unlabeled stationary point. Each additional labeled stationary point can contribute at most 6 linearly independent constraints.

Corollary 1. *A minimum of 7 labeled stationary points are necessary for a unique (up to a 4-dimensional solution space) solution for \mathcal{J} .*

Note that we used the term "unique" for the solution of \mathcal{J} (despite the fact that \mathcal{J} can be recovered only up to a 4-fold linear subspace) due to the fact that the collineations A, B can be recovered uniquely from the 4-dimensional \mathcal{J} tensor space.

Finally, we consider the situation of a *mixed* labeled and unlabeled triplets. Consider the case where $x \leq 7$ of the triplets are labeled as arising from stationary points. The corollary below addresses the question of how many

matching triplets arising from moving points are necessary given that $x \leq 7$ matching triplets are labeled as stationary? Clearly, when $x = 7$ there is no need for further measurements, but when $x < 7$ we obtain the following result:

Corollary 2. *In a situation of matching triplets arising from a mixture of stationary and moving points, let $x \leq 7$ be the number of matching triplets that are known a priori to arise from stationary points. To obtain a unique linear solution for \mathcal{J} (up to a 4-dimensional solution space), the minimal number of unlabeled matching triplets required is:*

$$\left\{ \begin{array}{ll} 60 - 10x & x \leq 5 \\ 4 & x = 6 \\ 0 & x = 7 \end{array} \right\},$$

out of which $40 - 6x$, $x < 7$, should be dynamic and at most $20 - 4x$, $x \leq 5$, could be unlabeled stationary points.

4 Homography Tensors for \mathcal{P}^n

The tensors \mathcal{H} and \mathcal{J} we have encountered so far belong to the general class of tensors defined as follows. Let $V(n, m, k)$, where $n > k$, be a $GL(V)$ module defined by the set of all tensors $v_1 \otimes \cdots \otimes v_m \in V^{\otimes m}$ where $v_i \in V$ are n -dimensional vectors and $\dim \text{Span}\{v_1, \dots, v_m\} \leq k$. What is the structure and dimension of $V(n, m, k)$? In the terminology of the previous sections, we considered the space \mathcal{P}^{n-1} , the number of views to be m and the motion of the dynamic points are limited to a k -dimensional subspace. Thus we have encountered $V(3, 3, 2)$ and $V(3, 3, 1)$ which stand for dynamic and stationary points in \mathcal{P}^2 , and encountered $V(4, 3, 2)$ and $V(4, 3, 1)$ which stand for dynamic motion along straight lines and stationary points in \mathcal{P}^3 . To generalize the construction of homography tensors to \mathcal{P}^n we need to find out:

1. The dimension of $V(n, m, k)$. Namely, given linear constraints generated by a multilinear form over the m -fold Htensor from known observations of m points moving inside k -dimensional subspaces what would be the maximal space those measurements could fill. For example, for $V(3, 3, 2)$ the maximal space is 26, which means we can obtain a unique solution for the $3 \times 3 \times 3$ Htensor, but for $V(4, 3, 2)$ the maximal dimension is 60 which means we can pin-point the $4 \times 4 \times 4$ tensor up to a 4-fold linear space. This will be the focus of this section.
2. Is the dimension of $V(n, m, k)$ sufficient for uniquely recovering the $m - 1$ individual collineations? and how to recover those collineations using the tensor slices? For example, we saw that the two collineations A, B can be recovered uniquely from \mathcal{H} and also uniquely from \mathcal{J} even though \mathcal{J} cannot be uniquely recovered from the measurements (in other words, we recovered A, B from the 4-dimensional linear space of solutions for \mathcal{J}). This generalization is an open for future research.

3. What are the constraints contributed from a labeled $k' < k$ -dimensional point? from example we saw that the stationary points $k' = 1$ for $V(3, 3, 2)$ contribute 7 independent constraints, and 10 independent constraints for $V(4, 3, 2)$. This is left open for future research.
4. What would be the dimension of the space covered by *mixed* observations, i.e., from labeled $k' < k$, and unlabeled points from k and $k' < k$. For example, we saw that the the labeled stationary $k' = 1$ points provide only 4 new constraints as 3 of the 7 provided by labeled stationary points constraints are included in the space of dimension $V(3, 3, 1)$ covered by unlabeled stationary points. This topic is left for future research.

We will focus below on the first item above which is the dimension of $V(n, m, k)$. The simple cases are $\dim V(n, m, 1) = \binom{n+m-1}{m}$ (because $V(n, m, 1) = \text{Sym}^m V$) and $\dim V(n, m, m-1) = n^m - \binom{n}{m}$ which arises by naive introspection. For example, $\dim V(3, 3, 2) = 26$ which means that the Htensor requires 26 matching triplets across three views of a dynamic planar configuration for a unique solution ($27 - 26 = 1$), whereas if all the measurements arise from “stationary” points then $\dim V(3, 3, 1) = 10$. Likewise, $\dim V(4, 3, 2) = 60$ which means the Jtensors are spanned by 4 tensors ($64 - 60 = 4$) and 60 matching triplets of 3D points across changes of coordinate systems of a dynamic 3D configuration are required for a solution, and if all the measurements arise from stationary points then $\dim V(4, 3, 1) = 20$.

We will show next that the question of structure and dimension of the $GL(V)$ module $V(n, m, k)$ can be generally solved by counting irreducibles using the tools of *Representation Theory* [15]. The notations and a brief primer on representation theory can be found in the Appendix. The central result of this section is proving that:

$$V(n, m, k) = \bigoplus_{\lambda_{k+1}=0} S_{\lambda}(V)^{\oplus f_{\lambda}} ,$$

and in particular

$$\dim V(n, m, k) = \sum_{\lambda_{k+1}=0} f_{\lambda} \dim S_{\lambda}(V) .$$

Where λ is a partition of m , the direct sum is over all partitions with at most k parts, f_{λ} is the number of standard tableaux on λ , and $S_{\lambda}(V)$ is Schur’s module.

The mathematics of representation theory may be somewhat unfamiliar as it was so far not in use in computer vision literature, yet it uncovers some beautiful connections between the recent new efforts of extending the envelope of Structure from Motion (SFM) theory and applications to non-rigid scenes and the representations of finite groups and of $GL(V)$ on the m -fold tensor product.

5 The Structure of $V(n, m, k)$

We would like to prove the following claim:

Claim.

$$V(n, m, k) = \bigoplus_{\lambda_{k+1}=0} S_\lambda(V)^{\oplus f_\lambda} .$$

In particular

$$\dim V(n, m, k) = \sum_{\lambda_{k+1}=0} f_\lambda s_\lambda .$$

Proof: Suppose $\lambda \vdash m$ and $\lambda_{k+1} = 0$. Let t be the tableau given by $t(i, j) = \sum_{l=1}^{i-1} \lambda_l + j$. Noting that $V(n, r, 1) = \text{Sym}^r V$ it follows that

$$\begin{aligned} V^{\otimes m} \cdot a_t &= \text{Sym}^{\lambda_1} V \otimes \cdots \otimes \text{Sym}^{\lambda_k} V \\ &= V(n, \lambda_1, 1) \otimes \cdots \otimes V(n, \lambda_k, 1) \subset V(n, m, k) . \end{aligned}$$

Therefore,

$$S_t(V) = V^{\otimes m} \cdot a_T \cdot b_T \subset V(n, m, k) \cdot b_T \subset V(n, m, k)$$

hence,

$$\bigoplus_{\lambda_{k+1}=0} S_\lambda(V)^{\oplus f_\lambda} \subset V(n, m, k) .$$

To show the other direction let (\cdot, \cdot) be a hermitian form on V and let the induced form on $V^{\otimes m}$ be given by

$$(u_1 \otimes \cdots \otimes u_m, v_1 \otimes \cdots \otimes v_m) = \prod_{i=1}^m (u_i, v_i) .$$

Note that

$$\begin{aligned} &(u_1 \wedge \cdots \wedge u_m, v_1 \otimes \cdots \otimes v_m) \\ &= \frac{1}{m!} (u_1 \wedge \cdots \wedge u_m, v_1 \wedge \cdots \wedge v_m) \\ &= \frac{1}{m!} \det[(u_i, v_j)]_{i,j=1}^m . \end{aligned}$$

Let $\lambda \vdash m$ with $\lambda_{k+1} \neq 0$, then the conjugate partition $\mu = (\mu_1 \geq \mu_2 \geq \cdots \geq \mu_t)$ satisfies $\mu_1 \geq k+1$. Let $l_j = \sum_{r=1}^j \mu_r$ and let t be the tableau given by $t(i, j) = l_{j-1} + i$. Then

$$\begin{aligned} S_t(V) &= V^{\otimes m} \cdot a_t \cdot b_t \subset V^{\otimes m} \cdot b_t \\ &= \wedge^{\mu_1} V \otimes \cdots \otimes \wedge^{\mu_t} V . \end{aligned}$$

Suppose now that $v_1, \dots, v_m \in V^{\otimes m}$ satisfy $\dim \text{Span}\{v_1, \dots, v_m\} \leq k$. Then $v_1 \wedge \cdots \wedge v_{\mu_1} = 0$ therefore for any $u_1, \dots, u_m \in V$

$$\begin{aligned} & ((u_1 \otimes \cdots \otimes u_m) \cdot b_T, v_1 \otimes \cdots \otimes v_m) = \\ & \prod_{r=1}^l \frac{1}{\mu_r!} \left(\bigwedge_{i=l_{r-1}+1}^{l_r} u_i, \bigwedge_{i=l_{r-1}+1}^{l_r} v_i \right) = 0 . \end{aligned}$$

It follows that $V(n, m, k)$ is orthogonal to

$$\bigoplus_{\lambda_{k+1} \neq 0} \mathcal{S}_\lambda(V)^{\oplus f_\lambda}$$

hence,

$$\dim V(n, m, k) \leq \dim \bigoplus_{\lambda_{k+1}=0} \mathcal{S}_\lambda(V)^{\oplus f_\lambda} .$$

□

Claim 5 can be used to give explicit formulas for $\dim V(n, m, k)$ when either k or $m - k$ are small. In the later case we write

$$\dim V(n, m, k) = n^m - \sum_{\lambda_{k+1} \neq 0} f_\lambda d_\lambda(n)$$

and note that the partitions of m with $\lambda_{k+1} \neq 0$ correspond to all partitions of all numbers up to $m - k - 1$.

Examples: to calculate $\dim V(n, m, m - 1)$ note that only $\lambda = (1^m)$ must be excluded, thus:

$$f_{(1^m)} = 1 , \quad d_{(1^m)}(n) = \binom{n}{m}$$

hence,

$$\dim V(n, m, m - 1) = n^m - \binom{n}{m} .$$

To calculate $\dim V(n, m, m - 2)$ we must exclude, in addition to the above, the partition $(2, 1^{m-2})$, thus:

$$f_{(2, 1^{m-2})} = m - 1 , \quad d_{(2, 1^{m-2})}(n) = (m - 1) \binom{n + 1}{m}$$

hence,

$$\dim V(n, m, m - 2) = n^m - \left[\binom{n}{m} + (m - 1)^2 \binom{n + 1}{m} \right] .$$

To calculate $\dim V(n, m, m - 3)$ we must exclude, in addition to the above, the partitions $(3, 1^{m-3})$ and $(2^2, 1^{m-4})$, thus:

$$f_{(3, 1^{m-3})} = \binom{m - 1}{2} , \quad d_{(3, 1^{m-3})}(n) = \binom{m - 1}{2} \binom{n + 2}{m}$$

$$f_{(2^2, 1^{m-4})} = \frac{m(m-3)}{2},$$

$$d_{(2^2, 1^{m-4})}(n) = \frac{(m-3)n}{2} \binom{n+1}{m-1}$$

Hence,

$$\dim V(n, m, m-3) = n^m - \left[\binom{n}{m} + \right.$$

$$(m-1)^2 \binom{n+1}{m} +$$

$$\left. \binom{m-1}{2}^2 \binom{n+2}{m} + \frac{m(m-3)^2 n}{4} \binom{n+1}{m-1} \right].$$

6 Experiments and Applications

We start with an experiment for separating dynamic from stationary points from a planar configuration. The projections of a planar configuration are governed by collineations. The conventional way to separate the moving from the stationary points is to treat the dynamic points as outliers and use robust estimation to recover the collineations [9]. Using homography tensors we can treat the dynamic and stationary points alike and recover the governing Htensor \mathcal{H} instead.

The point configuration is illustrated in Fig. 6. The moving points were part of 4 remote controls that were in motion while the camera changed position from one view to the next. The points were tracked along the three views, without knowledge what was stationary and what was moving. The triplet of matching points were fed into a least-square estimation for \mathcal{H} . We then checked the error of reprojection on the stationary points —these were at sub-pixel level as can be seen in Fig. 6h — and the accuracy of the line trajectory of the moving points. Because the moving points were clustered on only 4 objects (the remote controls), then the accuracy was measured by “eye-balling” the parallelism of the trajectories of all points within a moving object. The lines are closely parallel as can be seen in Fig. 6f. The Htensor can also be used to segment the scene into stationary and moving points — this is shown in Fig. 6e.

To illustrate the use of the homography tensor \mathcal{J} , consider the problem of 3D reconstruction of an object which extends beyond the field of view of the sensor. For this purpose we can use a stereo rig, that contains a texture pattern projector for obtaining matching points on textureless areas of the object. Because the field of view of the cameras does not cover the entire object, the stereo rig must acquire images from multiple viewing positions. Each image provides a 3D patch of the object and the goal is to “stitch” these

patches together by aligning their coordinate systems. In other words, we must recover the relative 3D motion of the rig. The problem is conventional if the texture projection is stationary (i.e., remains in place while the rig changes position); but here the projector moves with the rig. In this domain, the dynamic points are the points arising from the projected texture and the stationary points arise from texture markings on the object’s surface. Hence, if the rig moves in a piecewise straight-line path and the object is polyhedral, Htensor theory is an appropriate tool for aligning the coordinate systems of the 3D patches.

Once the Htensor \mathcal{J} are recovered one can align the reconstructed patches using two different approaches. The first approach is to align all the patches to one coordinate frame using direct mapping (Section 3.2) or by recovering the transformations A and B (Section 3.1). The second approach is to first segment the tracked points into stationary and dynamic points. Then, using only the stationary points, we can recover the collineation between the coordinate frames A and B .

We apply the Htensor to the scene with multiple objects shown in Fig 7. Most of the objects are textureless but there are stationary features throughout the scene. A texture was projected and 236 features were tracked between the images in each stereo pair and across the three stereo pairs. The feature set contains both stationary and dynamic points. It can be seen from the last row of Fig. 8 that the correct motion was captured because the stationary points were stabilized whereas the dynamic points are moving on straight line paths. The last image shows the segmented stationary points. Note that in our framework we use only projective reconstruction and we do not use any calibration. If Euclidean reconstruction is desired, a 4×4 projective-to-Euclidean transformation can be applied later on.

7 Summary

We have introduced in this chapter the m -view analogue of the classical collineation (homography matrix). The extension from 2 to m views introduces an additional parameter $k < n$ which endows the individual points of the point configuration, which is being transformed projectively from view to view, with the ability to become "dynamic". The value of k stands for the dimension of the subspace in which the individual points are allowed to move while the projective change of coordinates take place. For example, when $k = 1$ the points are not allowed to move (are stationary) just like with conventional collineations, and when $k = 2$ the individual points are allowed to move along straight line paths, and so forth.

The m -view tensors for \mathcal{P}^n and for $k < n$, referred to as homography tensors, were developed in detail for the case $n = 3, 4$ and the case $k = 2, 1$ — which are instances of practical value for applications. In the derivation of the homography tensor the following issues need to be addressed (i) the

maximal space contributed by dynamic points of sub-dimension k , (ii) number of constraints contributed by *mixed* points where some are labeled to move in k -subspace and some are unlabeled, and (iii) the use of the homography tensor as a mapping and the recovery of the individual projective mappings between views from the elements of the tensor.

Those issues were covered in detail for $n = 3, k = 2, 1$ (the \mathcal{H} tensor for planar configurations) and for $n = 4, k = 2, 1$ (the \mathcal{J} tensor for 3D configurations). For general n, m, k we have covered only the first issue above, that of dimension of the $\text{GL}(V)$ module $V(n, m, k)$ associated with the question of how many independent linear constraints are possible for a given value of n, m, k .

As for applications, we presented two instances, in 2D and 3D, of the problem of recovering the global alignment under dynamic motion. Without homography tensors, a recovery of alignment requires the use of statistical methods of sampling where the points undergoing dynamic motion are considered as outliers — whereas with the homography tensors both stationary and moving points can be considered alike and part of a global transformation which can be recovered analytically from observations (matching points across m views).

Generally, the homography tensors can be used to recover linear models under linear uncertainty. This generalization is quite straightforward, although the size of the resulting tensors grows exponentially. The use of such tensors in dimensions larger than \mathcal{P}^3 ($n > 4$) is not straightforward and is left for further research.

References

1. S. Avidan and A. Shashua. Trajectory triangulation of lines: Reconstruction of a 3d point moving along a line from a monocular image sequence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 1999.
2. 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.
3. A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *Proceedings of the European Conference on Computer Vision*, Freiburg, Germany, 1998. Springer, LNCS 1407.
4. G.H. Golub and C.F. Van Loan. *Matrix computations*. John Hopkins University Press, 1989.
5. R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.
6. M. Irani and P. Anandan. Parallax geometry of pairs of points for 3D scene analysis. In *Proceedings of the European Conference on Computer Vision*, LNCS 1064, pages 17–30, Cambridge, UK, April 1996. Springer-Verlag.
7. M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multiview parallax geometry and applications. In *Proceedings of the*

- European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.
8. M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization'. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, Washington, June 1994.
 9. P. Meer, D. Mintz, D. Kim, and A. Rosenfeld. Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1):59–70, 1991.
 10. A. Shashua, R. Meshulam, L. Wolf, A. Levin, and G. Kalai. On representation theory in computer vision problems. Technical report, School of Computer Science and Eng., The Hebrew University of Jerusalem, July 2002.
 11. A. Shashua and N. Navab. Relative affine structure: Canonical model for 3D from 2D geometry and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):873–883, 1996.
 12. A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *Proceedings of the International Conference on Computer Vision*, June 1995.
 13. 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.
 14. C.C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry and Remote Sensing, 1980.
 15. W.Fulton and J.Harris. *Representation Theory: a First Course*. Springer-Verlag, 1991.
 16. Lior Wolf, A. Shashua, and Y. Wexler. Join tensors: on 3d-to-3d alignment of dynamic sets. In *Proceedings of the International Conference on Pattern Recognition*, Barcelona, Spain, September 2000.

A Representation Theory Digest

In this section we briefly recall some relevant facts concerning the representation theory of the general linear group. For a thorough introduction see [15].

Let V be a finite n -dimensional vector space over the complex numbers. The collection of invertible $n \times n$ matrices is denoted by $GL(n)$ which is the group of automorphisms of V denoted by $GL(V)$. The vector space $V^{\otimes m}$ (m -fold tensor product) is spanned by decomposable tensors of the form $v_1 \otimes \cdots \otimes v_m$, where the vectors v_i are in V . Hence the dimension of $V^{\otimes m}$ is n^m . The vector space $V^{\oplus m}$ is the m -fold direct sum of V , thus is of dimension nm .

The *exterior* powers $\wedge^m V$ of V , $n \geq m$, is the vector space spanned by the $m \times m$ minors of the $n \times m$ matrix $[v_1, \dots, v_m]$ where the vectors v_i are in V . Hence the dimension of $\wedge^m V$ is $\binom{n}{m}$. The exterior powers are the images of the map $V^{\times m} \rightarrow V^{\otimes m}$ given by

$$(v_1, \dots, v_m) \rightarrow \sum_{\sigma \in S_m} \text{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(m)}$$

where S_m denotes the symmetric group (of *permutations* of m letters).

The *symmetric powers* $\text{Sym}^m V$ are the images of the map $V^{\times m} \rightarrow V^{\otimes m}$ given by

$$(v_1, \dots, v_m) \rightarrow \sum_{\sigma \in S_m} v_{\sigma(1)} \otimes \dots \otimes v_{\sigma(m)}$$

Hence the vector space $\text{Sym}^m V$ is of dimension $\binom{n+m-1}{m}$. Note that,

$$V \otimes V = \text{Sym}^2 V \oplus \wedge^2 V$$

with the appropriate dimension: $n^2 = \binom{n+1}{2} + \binom{n}{2}$. This decomposition into irreducibles (see later) is not true for $V^{\otimes m}$, $m > 2$. The remainder of this section is devoted to the necessary notation for representing $V^{\otimes m}$ as a decomposition of irreducibles.

A *representation* of a group G on a complex finite dimensional space U is a homomorphism G to $GL(U)$ - the group of linear automorphisms of U . The action of $g \in G$ on $u \in U$ is denoted by $g \cdot u$. The G -module U is *irreducible* if it contains no non-trivial G -invariant subspaces. Any finite dimensional representation of a compact group G can be decomposed as a direct sum of irreducible representations. This basic property called *complete reducibility* also holds for all holomorphic representations of the general linear group $GL(V)$.

The main focus of Section 4 is the space

$$V(n, m, k) = \text{Span}\{v_1 \otimes \dots \otimes v_m \in V^{\otimes m} : \dim \text{Span}\{v_1, \dots, v_m\} \leq k\}.$$

Since $V(n, m, k)$ is invariant under the $GL(V)$ action given by $g \cdot v_1 \otimes \dots \otimes v_m = g(v_1) \otimes \dots \otimes g(v_m)$ it is natural to study its structure by decomposing it into irreducible $GL(V)$ -modules.

The description of the finite dimensional irreducible representations (irreps) of $GL(V)$ depends on the combinatorics of partitions and Young diagrams which we now describe:

A *partition* of m is an ordered set $\lambda = (\lambda_1, \dots, \lambda_k)$ such that $\lambda_1 \geq \dots \geq \lambda_k \geq 1$ and $\sum \lambda_i = m$. A partition is represented by its *Young diagram* (also called *shape*) which consists of k left aligned rows of boxes with λ_i boxes in row i . The *conjugate partition* $\mu = (\mu_1, \dots, \mu_r)$ to a partition λ is defined by interchanging rows and columns in the Young diagram — or without reference to the diagram, μ_i is the number of terms in λ that are greater than or equal to i .

An assignment of the numbers $\{1, \dots, m\}$ to each of the boxes of the diagram of λ , one number to each box, is called a *tableau*. A tableau in which all the rows and columns of the diagram are increasing is called a *standard tableau*. We denote by f_λ the number of standard tableaux on λ , i.e., the number of ways to fill the young diagram of λ with the numbers from 1 to m , such that all rows and columns are increasing. Let (i, j) denote the coordinates of

the boxes of the diagram where $i = 1, \dots, k$ denotes the row number and j denotes the column number, i.e., $j = 1, \dots, \lambda_i$ in the i 'th row. The *hook length* h_{ij} of a box at position (i, j) in the diagram is the number of boxes directly below plus the number of boxes to the right plus 1 (without reference to the diagram, $h_{ij} = \lambda_i + \mu_j - i - j + 1$). Then,

$$f_\lambda = \frac{m!}{\prod_{(i,j)} h_{ij}}$$

where the product of the hook-lengths is over all boxes of the diagram. We denote by $d_\lambda(n)$ the number of *semi-standard tableaux* which is the number of ways to fill the diagram with the numbers from 1 to n , such that all rows are non-decreasing and all columns are increasing. We have:

$$d_\lambda(n) = \prod_{(i,j)} \frac{n - i + j}{h_{ij}}.$$

Let S_m denote the symmetric group on $\{1, \dots, m\}$. The *group algebra* $\mathbb{C}S_m$ is the algebra spanned by the elements of S_m

$$\mathbb{C}G = \left\{ \sum_{\sigma \in S_m} \alpha_\sigma \sigma \mid \alpha_\sigma \in \mathbb{C} \right\}$$

where addition and multiplication are defined as follows:

$$\alpha \left(\sum_{\sigma \in S_m} \alpha_\sigma \sigma \right) + \beta \left(\sum_{\sigma \in S_m} \beta_\sigma \sigma \right) = \sum_{\sigma \in S_m} (\alpha \alpha_\sigma + \beta \beta_\sigma) \sigma$$

and

$$\left(\sum_{\sigma \in S_m} \alpha_\sigma \sigma \right) \left(\sum_{\tau \in S_m} \beta_\tau \tau \right) = \sum_{g \in S_m} \left(\sum_{g=\sigma\tau} \alpha_\sigma \beta_\tau \right) g$$

for $\alpha, \beta, \alpha_\sigma, \beta_\sigma \in \mathbb{C}$.

Let t be a tableau on λ (a numbering of the boxes of the diagram) and let $P(t)$ denote the group of all permutations $\sigma \in S_m$ which permute only the rows of t . Similarly, let $Q(t)$ denote the group of permutations that preserve the columns of t . Let a_t, b_t be two elements in the group algebra $\mathbb{C}S_m$ defined as:

$$a_t = \sum_{g \in P(t)} g, \quad b_t = \sum_{g \in Q(t)} \text{sgn}(g)g.$$

The group algebra $\mathbb{C}S_m$ acts on $V^{\otimes m}$ on the right by permuting factors, i.e., $(v_1 \otimes \dots \otimes v_m) \cdot \sigma = v_{\sigma(1)} \otimes \dots \otimes v_{\sigma(m)}$. For a general shape λ and a tableau t on λ the image of $a_t, V^{\otimes m} \cdot a_t$, is the subspace:

$$V^{\otimes m} \cdot a_t = \text{Sym}^{\lambda_1} V \otimes \dots \otimes \text{Sym}^{\lambda_k} V \subset V^{\otimes m}$$

and the image of b_t is

$$V^{\otimes m} \cdot b_t = \wedge^{\mu_1} V \otimes \dots \otimes \wedge^{\mu_r} V \subset V^{\otimes m}$$

where μ is the conjugate partition to λ . The *Young symmetrizer* is defined by $c_t = a_t \cdot b_t \in \mathbb{C}S_m$. The image of the Young symmetrizer

$$S_t(V) = V^{\otimes m} \cdot c_t$$

is the *Schur Module* associated to t and is an irreducible $GL(V)$ -module. The isomorphism type of $S_t(V)$ depends only on the shape λ so we may write $S_t(V) = S_\lambda(V)$. It turns out that all the polynomial irreps of $GL(V)$ are of the form $S_\lambda(V)$ for some m and a partition $\lambda \vdash m$.

Let \mathcal{T}_λ denote the set of standard tableaux on λ then the direct sum decomposition of $V^{\otimes m}$ into irreducible $GL(V)$ -modules is given by

$$\begin{aligned} V^{\otimes m} &= \bigoplus_{\lambda \vdash m} \bigoplus_{t \in \mathcal{T}_\lambda} S_t(V) \cong \\ &\bigoplus_{\lambda \vdash m} \mathcal{S}_\lambda(V)^{\oplus f_\lambda} . \end{aligned}$$

Since $d_\lambda(n) = \dim \mathcal{S}_\lambda(V)$ it follows that

$$\dim V^{\otimes m} = n^m = \sum_{\lambda \vdash m} d_\lambda(n) f_\lambda .$$

For example, consider $n = m = 3$, i.e., $V \otimes V \otimes V$ where $\dim V = 3$. There are three possible partitions λ of 3 — these are (3) , $(1, 1, 1)$ and $(2, 1)$. From the above, $S_{(3)}(V) = \text{Sym}^3 V$ and $S_{(1,1,1)}(V) = \wedge^3 V$. There are two, $f_{(2,1)} = 2$, standard tableaux for $\lambda = (2, 1)$ and these are 123 and 132 (numbering of boxes left to right and top to bottom). There are eight, $d_{(2,1)}(3) = 8$, semi-standard tableaux which are: 112, 113, 122, 123, 132, 133, 223 and 233. We have the decomposition:

$$V \otimes V \otimes V = \text{Sym}^3 V \oplus \wedge^3 V \oplus (S_{(2,1)}(V))^{\oplus 2}$$

with the appropriate dimensions: $27 = 10 + 1 + (8 + 8)$.

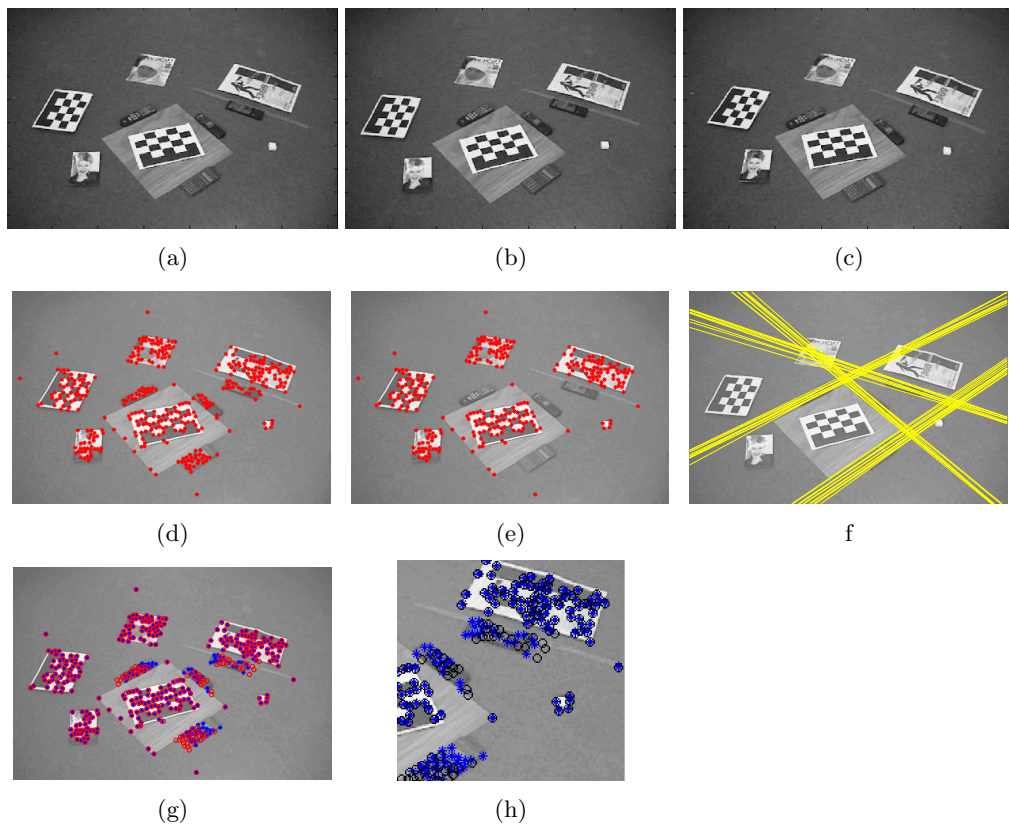


Fig. 6. (a),(b),(c) three views of a planar scene with 4 remotes moving on straight lines (d) The first view with the points that were tracked across the sequence. These points were used for computing the homography tensor \mathcal{H} in a least-squares manner. (e) Segmentation: the homography tensor was used to choose the stationary points. Only the stationary points are shown. (f) Trajectory lines: the homography tensor was used to calculate the trajectory lines. In this figure we see the trajectory lines in the third image. (g) Reprojection: Using the homography tensor we reprojected the points in view 1 to view 3. The reprojected points are shown as circles. The Tracked points as stars. (h) A zoom of the previous image.



(a) Left view, time 1

(b) Right view, time 1



(c) Left view, time 2

(d) Right view, time 2



(e) Left view, time 3

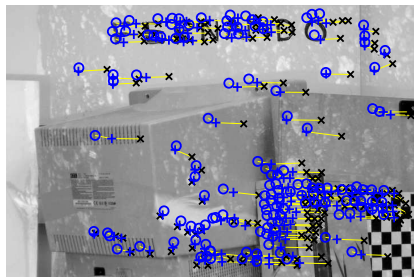
(f) Right view, time 3

Fig. 7. A pair of views from a stereo rig taken at three time instants. The rig is moving with the texture pattern. The scene therefore contains both stationary and dynamic points.

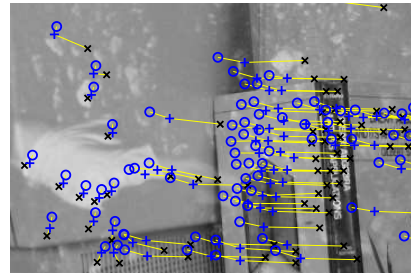


(a) Left-hand image of first pair

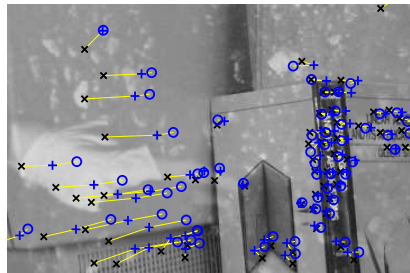
(b) Right-hand image of first pair



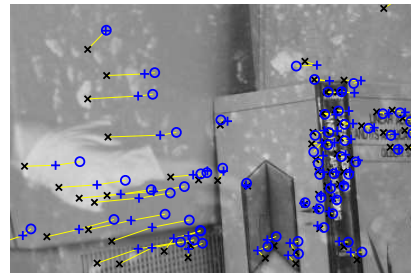
(c) Tracked points, shown on (a)



(d) Zoomed part of (c)



(e) Stabilized points, shown on (a)



(f) Zoomed part of (e)



(g) Segmentation of moving/static points

Fig. 8. Application of the Htensor \mathcal{J} to 3D reconstruction. Row 1 displays two images from one stereo pair. The images show the projected texture. The stereo rig and the projector are moved together at subsequent time instants (not shown). Row 2 displays the tracked points. Some of the points are stationary features (physical objects) and some are from the projected texture. Row 3 displays the points after the motion was canceled with the Htensor. Notice that points that are stationary were stabilized, meaning that the Htensor captured the correct 3D motion. (g) shows the stationary points, which were identified by the Htensor.