Revisiting Single-View Shape Tensors: Theory and Applications

Anat Levin and Amnon Shashua

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

Abstract. Given the projection of a sufficient number of points it is possible to algebraically eliminate the camera parameters and obtain view-invariant functions of image coordinates and space coordinates. These single view invariants have been introduced in the past, however, they are not as well understood as their dual multi-view tensors. In this paper we revisit the dual tensors (bilinear, trilinear and quadlinear), both the general and the reference-plane reduced version, and describe the complete set of synthetic constraints, properties of the tensor slices, reprojection equations, non-linear constraints and reconstruction formulas. We then apply some of the new results, such as the dual reprojection equations, for multi-view point tracking under occlusions.

1 Introduction

There is a large body of research on multi-view geometry of 3D scenes which has culminated to the point where the issues and solutions are well understood. The body of work on multi-view geometry centers around matching tensors of 2,3,4 views known as multi-view constraints (bifocal, trifocal and quadrifocal tensors) which are borne out of algebraic elimination of the scene geometry (shape) from the 3D-to-2D projection equations given a sufficient number of views [16,20,9,5, 7,10,17,1]. The "dual" form of the elimination process is to eliminate the camera parameters (motion) given a sufficient number of points in a single view with the result of what is known as *single-view shape tensors* [2,21,3] and in a reduced setting where a reference plane is identified in advance is called *parallax geometry* [11,12,4,18].

Multi-view geometry has been put into practice in a variety of applications including 3D reconstruction, novel view synthesis, camera ego-motion, augmented reality and visual recognition by alignment. The multi-point geometry, on the other hand, has been hardly put into use although the topic makes a very appealing case for applications. In many instances, one would like to achieve a direct representation of 3D shape from images without the need to recover the camera geometry as an intermediate step. This includes indexing into a library of objects (cf. [13]), multi-body segmentation (collection of points belong to the same structure when the shape invariants hold), and even for tracking applications

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 399–414, 2002.

(which traditionally use multi-view constraints) where features may get lost due to occlusions and later reappear. A direct shape constraint is advantageous due to the local image support needed to make it work.

In this paper we focus on open issues which remain with shape tensors, such as the number and nature of the *synthetic* constraints, properties of tensor slices, reprojection equations, a full account of the non-linear constraints and reconstruction formulas. We then apply some of the new results, such as the dual reprojection equations, for multi-view point tracking under occlusions.

We will start with a brief description of what is known about these tensors which will create the context for describing in more details our contributions to this topic.

1.1 What Is Known To-Date about Shape Tensors

The basic idea is that points and cameras can be switched (duality principle) and as a result one can obtain exactly the same multi-linear constraints as in the multi-view derivations, where instead of multiple views we have multiple points. Let $P_i = (X_i, Y_i, Z_i, W_i)^\top \in \mathcal{P}^3$ denote points in 3D projective space and let M be a 3×4 projection matrix, thus $p_i \cong MP_i$ where $p_i \in \mathcal{P}^2$ be the corresponding image points in the 2D projective plane. We wish to algebraically eliminate the camera parameters (matrix M) by having a sufficient number of points. This could be done elegantly, and along the way obtain the duality principle, if we first change basis as follows. Let the first 4 points $P_1, ..., P_4$ be assigned $(1,0,0,0), ..., (0,0,0,1)$ and let the image undergo a projective change of coordinates such that the corresponding points $p_1, ..., p_4$ be assigned $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1), e_4 = (1,1,1)$, respectively. Given this setup the camera matrix M contains only 4 non-vanishing entries:

$$M = \begin{bmatrix} \alpha & 0 & 0 & \delta \\ 0 & \beta & 0 & \delta \\ 0 & 0 & \gamma & \delta \end{bmatrix}$$

Let $\hat{M} = (\alpha, \beta, \gamma, \delta) \in \mathcal{P}^3$ be a point (representing the camera) and let $\hat{P}_i$ be the projection matrix:

$$\hat{P}_i = \begin{bmatrix} X_i & 0 & 0 & W_i \\ 0 & Y_i & 0 & W_i \\ 0 & 0 & Z_i & W_i \end{bmatrix}$$

And we have the duality $p_i \cong MP_i = \hat{P}_i \hat{M}$ where the role of the motion (the camera) and shape have been switched. At this point we follow exactly the same steps one does with the multi-view tensors: let l_i, l'_i be two distinct lines passing through the image point p_i, i.e., $p_i^\top l_i = 0$ and $p_i^\top l'_i = 0$, and therefore we have $l_i^\top \hat{P}_i \hat{M} = 0$ and $l_i'^\top \hat{P}_i \hat{M} = 0$. For $i = 5, ..., 8$ we have therefore $E\hat{M} = 0$ where:

$$E = \begin{bmatrix} l_5^\top \hat{P}_5 \\ \cdot \\ l_8^\top \hat{P}_8 \\ l_5'^\top \hat{P}_5 \\ \cdot \\ l_8'^\top \hat{P}_8 \end{bmatrix} \qquad (1)$$

Therefore the determinant of any 4 rows of E must vanish. The choice of the 4 rows can include 2 points, 3 points, or 4 points (on top of the 4 basis points $P_1, ..., P_4$) and each such choice determines a multilinear constraint whose coefficients are arranged in a tensor. For the case of (4+) 2 points, say points p_5, p_6, there is only one such tensor with the bilinear constraint $p_6^\top \mathcal{F} p_5 = 0$ where the 3×3 matrix $\mathcal{F}$ contains the shape parameters of P_5, P_6. In the work of [2,21,3] the properties of $\mathcal{F}$ were derived (there are 4 linear constraints and rank($\mathcal{F}$)=2). The cases of 7 and 8 points were less understood. Clearly, in the case of 7 points, there are three tensors where we choose two rows of a "reference" point (say P_5) and one row from the remaining two points (P_6, P_7). The determinant expansion provides a trilinear constraint of the form $p_5^i l_j^6 l_k^7 \mathcal{T}_i^{jk} = 0$ where p_5 is the reference point, l^6, l^7 are lines through the points p_6, p_7 respectively, and the indices i, j, k follow the covariant-contravariant notations (upper index represents points, lower represent lines) and follow the summation convention (contraction) $u^i v_i = u^1 v_1 + u^2 v_2 + ... + u^n v_n$. Note that since the tensor is contracted by a point (p_5) and a *choice* of line through the remaining two points, then each view contributes 4 linear constraints on the 27 unknowns of the tensor.

At this point the literature becomes incomplete — clearly, we expect three views to be sufficient for recovering the tensor (because of duality with the multi-view trilinear tensor) thus the coefficients of the tensor must satisfy internal linear constraints. In [2,21,3] the way around this was to find out using Grobner basis with computer algebra tools that there are only 11 parameters (up to scale) which form 4 trilinear equations. The number of parameters is indeed 11 (as we will see later), but the tensor has been lost in all of this. In [8], in attempt to summarize the topic, have noticed that there are internal linear constraints, which they called "synthetic constraints" (which we will touch upon later). However, they did not provide the exact number of such constraints (which is indeed 16, leaving 11 parameters up to scale, as we shall see later). Moreover, the following issues remained open: (i) non-linear constraints on the tensor (there should be 4), (ii) tensor slices (from which we obtain "reprojection equation", homography slices, dual epipoles, etc.), and (iii) reconstruction of shape from the tensor slices' properties (dual epipole, dual homography).

The case of 8 points is open to a large extent. This case is dual to the quadrifocal multi-view tensor, thus by choosing one row from each point we obtain a vanishing determinant involving 4 points which provides 16 constraints (per view) $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ for the 81 coefficients of the tensor $\mathcal{Q}^{ijkt}$. Again, one expects two views to suffice, therefore the quadlinear tensor must contain many internal linear constraints. In [3] through the use of Grobner basis with computer algebra tools it was found that there are 22 quadlinear constraints (per view) with 41 coefficients, thus two views would suffice. It is unclear where this result comes from, in fact (as we will show later) the quadlinear tensor has 81 coefficients (just like its dual brother in the multi-view case) but there are 58 synthetic linear constraints — therefore we have 23 parameters up to scale. The first view provides 12 constraints, and the second view provides 11 constraints (thus two views are sufficient). As for non-linear constraints, there are 13 of them.

Next, consider the case in which a plane has been identified in advance and has been "stabilized" across the sequence of views. This reduced setting has been coined "parallax geometry" and has the advantage of making a clear geometric picture of the basic building block of the dual geometry [11,12,4]. Existing work focus on the geometric interpretation (the dual epipole) and in [12] the trilinear constraint was derived geometrically (and has been shown to be bilinear in this setting).

We will show that, beyond the geometrical interpretation, the real advantages lies elsewhere. First, this setting corresponds to having the first 4 points $P_1, ..., P_4$ to have the coordinates $(1,0,0,0), (0,1,0,0), (0,0,1,0), (1,1,1,0)$ which is appropriate when $P_1, ..., P_4$ are indeed coplanar. The tensors are the same, what changes is the number of synthetic constraints — for the bilinear tensor we have 5 constraints, for the trilinear tensor we have 21 constraints, and for the quadlinear tensor 72 constraints (the latter requires tools from representation theory). More importantly, these tensors do not have any non-linear constraints — thus making them appealing in practice. The lack of non-linear constraints is at the heart of the recent result in [18] showing that factorization is possible in this context.

Finally, we have conducted real imagery experiments which highlight the use of some of the new discoveries — such as the reprojection equations — and which covers a number of tracking applications and multi-body motion segmentation.

2 Synthetic Constraints in the General Case

The multi-point tensors are derived from the vanishing 4×4 determinants of E (eqn. 1). Because of duality, we obtain exactly the same tensorial forms as in the multi-view case. The difference is that the projection matrices $\hat{P}_i$ are sparse and as a result one obtains additional constraints, which following [8] we will call *synthetic* constraints, which we will now analyze.

Consider the camera projection matrix M_j be constructed such that the j'th column is e_j (the j'th standard basis vector) and the remaining entries vanish. We have then $M_j P$ is either e_j or vanishes for all choices of P. Let l_i, l'_i, $i = 5, ..., 8$, be lines through e_j, therefore

$$l_i^\top M_j P = l_i^\top \hat{P} \hat{M}_j = 0$$
$$l_i'^\top M_j P = l_i'^\top \hat{P} \hat{M}_j = 0$$

for all points P, and dually for all projection matrices $\hat{P}$. Therefore the 4×4 determinants of E vanish regardless of $\hat{P}_i$. For example, in the case of 6 points (choose two rows from p_5 and two rows from p_6) we obtain $e_j^\top \mathcal{F} e_j = 0$, $j = 1, ..., 4$. Therefore, we have 4 synthetic constraints on $\mathcal{F}$, i.e., the 6-point tensor is represented by $9 - 4 = 5$ parameters up to scale as already pointed out in [2, 21,3].

In the case of 7 points, say p_5 is the reference point, thus we have the multilinear constraint $p_5^i l_j^6 l_k^7 \mathcal{T}_i^{jk} = 0$ where l^6, l^7 are lines through the points p_6, p_7 respectively. Let l_j, l'_k be a line through $e = e_j$, then from the above we have that $e^i l_j l'_k \mathcal{T}_i^{jk} = 0$ which provides 4 constraints (because there are two choices for lines l_j and two choices for lines l'_k). Therefore we have 16 synthetic constraints (because e ranges over $e_1 = (1,0,0), ..., e_4 = (1,1,1)$. We have arrived to the result:

Claim. In the case of 7 points, each of the three $3 \times 3 \times 3$ trilinear tensors contract on a point (the reference point) and two lines coincident with the remaining two points. The choice of the reference point determines the tensor in question. Each of these tensors has 16 internal linear constraints, thus leaving 11 parameters up to scale. Each view contributes 4 linear constraints on the tensor in question, thus 3 views are necessary for a linear solution.

The number of parameters a 7-point configuration carries is 3+3 = 6 (because P_5 can be set arbitrarily, say $P_5 = (1,1,1,1)$ and each additional point carries 3 parameters). *We therefore expect 4 non-linear constraints on each of the tensors.* We will return to this issue later after we study the tensor slices.

In the case of 8 points, we have a single $3 \times 3 \times 3 \times 3$ tensor $\mathcal{Q}^{ijkt}$ responsible for the 16 quadlinear constraints $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ (we have a choice of 2 lines for each point, thus 16 constraints). From the discussion above, if all the lines are coincident with $e = e_j$ the constraint holds for all quadlinear tensors (i.e., apply to all space points). Therefore, for $e = e_1$ we have 16 synthetic constraints. For $e = e_2$ we will have 15 constraints because the line between e_1 and e_2 is already covered by the previous 16 constraints. Likewise, each additional point provides one less constraint, thus we have a total of $16 + 15 + 14 + 13 = 58$ synthetic constraints. The first view will contribute 12 constraints (the lines through $p_5, ..., p_8$ passing through e_j are already spanned by the synthetic constraints), and the second view will contribute 11 constraints (because the lines through the four points in view 1 and the four points in view 2 are spanned by the 12 constraints from view 1). Therefore, we have $58 + 12 + 11 = 81$ which provides sufficient constraints to solve for the quadlinear tensor. To summarize:

Claim. In the case of 8 points, there is a single quadlinear tensor of size $3^4 = 81$. The tensor has 58 linear constraints, thus is defined by 23 parameters up to scale. Each view contributes 16 linear constraints on the tensor of which 12 are independent for the first view and 11 are independent for the second view.

An 8-point configuration is determined by $3 + 3 + 3 = 9$ parameters, thus we expect 13 non-linear constraints.

3 Synthetic Constraints with a Reference Plane

When a plane is identified in advance and stabilized we find a different set of synthetic constraints. The first 4 basis points $P_1, ..., P_4$ are assigned the coordinates

$(1,0,0,0),(0,1,0,0),(0,0,1,0),(1,1,1,0)$ which is appropriate when $P_1,...,P_4$ are indeed coplanar. The resulting camera matrix becomes:

$$M = \begin{bmatrix} \delta & 0 & 0 & \alpha \\ 0 & \delta & 0 & \beta \\ 0 & 0 & \delta & \gamma \end{bmatrix}$$

and the resulting projection matrix $\hat{P}_i$ becomes:

$$\hat{P} = \begin{bmatrix} W_i & 0 & 0 & X_i \\ 0 & W_i & 0 & Y_i \\ 0 & 0 & W_i & Z_i \end{bmatrix}$$

And as in the general case we have the duality $p_i \cong MP_i = \hat{P}_i\hat{M}$ where the role of the motion (the camera) and shape have been switched. The matrix E is identical to eqn. 1 thus the tensors are exactly the same. What is different is the number of synthetic constraints (and also geometric interpretation which we will address later).

Since $P_1,...,P_4$ are coplanar we have the constraint $P_i^\top n = 0$, $i = 1,...,4$ and, due to our choice of coordinates, $n = (0,0,0,1)^T$. Consider the family of camera matrices $M = un^\top$ for all choices of $u = (u_1,u_2,u_3)$. In other words, the 4'th column of M consists of the arbitrary vector u and all other entries vanish. Thus we have that MP either vanishes or is equal to u (up to scale) for all P. Let l_i, l'_i be lines through u, therefore

$$l_i^\top M_j P = l_i^\top \hat{P}\hat{M}_j = 0$$
$$l_i'^\top M_j P = l_i'^\top \hat{P}\hat{M}_j = 0$$

for all points P, and dually for all projection matrices $\hat{P}$. Therefore the 4×4 determinants of E vanish regardless of $\hat{P}_i$. For example, in the case of 6 points (choose two rows from p_5 and two rows from p_6) we obtain $u^\top \mathcal{F} u = 0$ for all choices of u, thus the matrix $\mathcal{F}$ is skew-symmetric and in turn is defined by 3 parameters (as already pointed out in [11,12,4]).

In the case of 7 points, say p_5 is the reference point, thus we have the multi-linear constraint $p_5^i l_j^6 l_k^7 \mathcal{T}_i^{jk} = 0$ where l^6, l^7 are lines through the points p_6, p_7 respectively. Let l_j, l'_k be a line through u, then from the above we have that $u^i l_j l'_k \mathcal{T}_i^{jk} = 0$ *for all choices of* u *where* $l^\top u = 0$ *and* $l'^\top u = 0$. The number of synthetic constraints is the dimension of $\mathcal{T}_i^{jk}$. The issue of dimension is dual to the issue of recovering the multi-view tensor from a coplanar scene. Let p, p', p'' be matching points across 3 views and let A, B be the homography matrices $p' \cong Ap$ and $p'' \cong Bp$ (since the scene is coplanar). Let s, r be lines through p', p'' respectively, thus the measurements for the multi-view trilinear tensor come from $p^i s_j r_k \mathcal{T}_i^{jk} = 0$ for all choices of p while $s^\top(Ap) = 0$ and $r^\top(Bp) = 0$. Since the choice of A, B does not affect the issue of dimension, we may as well set $A = B = I$ and we have exactly the same situation as in the multi-point tensor described above. The issue of dimension for planar configuration for the multi-view trilinear tensor was resolved in [19] with the answer of 21. As a result, we can conclude that there are 21 synthetic constraints for the multi-point trilinear tensor. Moreover, the remaining parameters $27 - 21 - 1 = 5$ is exactly what is required to describe a configuration of 7 points, 4 of which are coplanar: P_6 adds only 2 parameters (because the first 5 points do not provide a full projective basis because the 4'th point is spanned by the first three, thus carries only 2 parameters, therefore they provide only 14 degrees of freedom instead of 15), and P_7 adds the usual 3 parameters. Note that since we are left with 6 parameters (up to scale) and each view contributes 4 linear equations on the tensor, only 2 views are necessary (instead of 3 which was required in the general case). Any of the two views adds only 3 constraints, instead of the expected 4, since it's possible to choose: $l^6 = p_6 \times p_5$ and $l^7 = p_7 \times p_5$. Than p_5, l^6, l^7 is a configuration of points and two lines through the point, those expressed already by the synthetic constraints. We summarize:

Claim. In the case of 7 points where the first 4 are coplanar, each of the three trilinear tensors has 21 internal linear constraints of the form $u^i l_j l'_k \mathcal{T}_i^{jk} = 0$ for all choices of u where the lines l, l' are coincident with u. These constraints are all the constraints on the tensor, there are no other non-linear constraints. Finally, only two views are required in order to solve for the tensor.

In the case of 8 points, the dimension analysis is more subtle and requires different tools. The quadlinear tensor is defined exactly as in the general case: we have a single $3 \times 3 \times 3 \times 3$ tensor $\mathcal{Q}^{ijkt}$ responsible for the 16 quadlinear constraints $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ (we have a choice of 2 lines for each point, thus 16 constraints). From the discussion above, the four lines contracted by the tensor are all coincident with the arbitrary point u. Therefore, the question is what is the dimension of the set of constraints $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ where the lines are arbitrary but form a 2-dimensional subspace? Let

$$V = \{v_1 \otimes v_2 \otimes v_3 \otimes v_4 | \dim \mathrm{Span}\{v_1, ..., v_4\} \leq 2\}$$

where $v_1, ..., v_4$ are vectors in R^3. Our question regarding the number of synthetic constraints is equivalent to the question of *what is the dimensions of* V? The answer is 72 and is derived as follows.

Let $\lambda = (\lambda_1, ..., \lambda_4)$, be a partition of 4, i.e., $\lambda_1 \geq \lambda_2 \geq \lambda_2 \geq \lambda_4$ and $\sum_i \lambda_i = 4$. A diagram associated with λ has 4 rows of left-aligned boxes with λ_i boxes in row i. Let f_λ be the number of ways to fill the diagram of λ with the numbers from 1 to 4, such that all rows and columns are increasing. Let (i, j) denote the coordinates of the boxes of the diagram where $i = 1, .., 4$ denotes the row number and j denotes the column number, i.e., $j = 1, ..., \lambda_i$ 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. Then,

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

where the product of the hook-lengths is over all boxes of the diagram. Let d_λ denote the number of ways to fill the diagram with the numbers from 1 to 3, such that all rows are non-decreasing and all columns are increasing. We have:

$$d_\lambda = \prod_{(i,j)} \frac{3-i+j}{h_{ij}}.$$

The definitions above can be found in [6]. With this in mind we have (proof is omitted) that:

$$\dim V = \sum_{\lambda_3=\lambda_4=0} f_\lambda d_\lambda.$$

We have therefore only three partitions: $\lambda = (4), (2,2), (3,1)$ to consider. Thus, $f_{(4)} = 1, d_{(4)} = 15, f_{(2,2)} = 2, d_{(2,2)} = 6, f_{(3,1)} = 3$ and $d_{(3,1)} = 15$. Therefore, $\dim V = 15 + 12 + 45 = 72$.

Note that there no more non-linear constraints because we are left with $81 - 1 - 72 = 8$ parameters which is exactly the number required to represent a configuration of 8 points in which the first 4 are coplanar: P_6 contributes 2 parameters, P_7, P_8 contribute 3 parameters each. Finally, although we are left with $81 - 72 - 1 = 8$ parameters, we still need two views: the first contributes 5 constraints and the second 4 constraints (proof omitted). We summarize:

Claim. In the case of 8 points where the first 4 are coplanar, the quadlinear tensor has 72 internal linear constraints. These constraints are all the constraints on the tensor, there are no other non-linear constraints.

4 Tensor Slices and Properties

The dual tensors are borne out of exactly the same multi-linear forms as the multi-view tensors — the differences lie in the fact that the projection matrices $\hat{P}_i$ are sparse and thus additional constraints are imposed (like the synthetic constraints described above). Moreover, because the multilinear forms are the same we should expect to have a "dual" of each of the basic elements one encounters in multi-view analysis: image ray, epipolar line, epipoles and homography matrix. These duals exist both in the general case and in the special case of stabilized reference plane. The duals of the homography matrices are the most important because they are the key for obtaining the source of the non-linear constraints for the trilinear tensor (as pointed out in [1] for the multi-view trilinear tensor).

We will derive the dual elements, the reprojection equation from the trilinear tensor, the homography slices of the trilinear tensor, the non-linear constraints from the homography slices, the breakdown of the trilinear tensor into a epipole-homography structure, and reconstruction of space points. We will switch back and forth between the general case and the case of stabilized reference plane (which we will refer to as the "reduced case").

We will assume, without proof due to lack of space, that indeed the family of all camera matrices projecting a fixed set of 4 copalanr points from $\mathcal{P}^3$ to $\mathcal{P}^2$, have a common stabilized plane — thus they differ from one another only in the location of the center of projection.

Let $null(M)$ be the projection center. Note that in the general case $null(M) = (1/\alpha, 1/\beta, 1/\gamma, -1/\delta)^T$, whereas in the reduced case $null(M) = (\alpha, \beta, \gamma, -\delta)^T$. This indicates something of importance: in the general case, when $\hat{M} = (\alpha, \beta, \gamma, \delta)^T$ varies along a linear subspace (a line or a plane), $null(M)$ varies along an algebraic *surface* (non-linear), whereas in the reduced case, $null(M)$ varies along a linear subspace of the same dimension. This is the key for the simple geometric interpretation of the elements (like image ray and epipole) in the reduced case (as introduced in [11]). Nevertheless, all the elements exist and are well defined in the general case as well.

The dual epipoles. In the multi-view context the epipoles are $M_i null(M_j)$ which is the projection of the j'th camera center onto the i'th image plane. Likewise, because of the duality $MP = \hat{P}\hat{M}$, the *dual epipole is defined by* $\hat{P}_i null(\hat{P}_j)$. In the case of 6 points bilinear tensor we should have two dual epipoles: $e_{65} = \hat{P}_6 null(\hat{P}_5)$ and $e_{56} = \hat{P}_5 null(\hat{P}_6)$.

Claim. $\hat{P}_6 null(\hat{P}_5)$ is the projection of P_6 via the camera whose center is at P_5.

Proof: Recall that $null(\hat{P}_5) = (1/X_5, 1/Y_5, 1/Z_5, -1/W_5)$. From duality we have:

$$\hat{P}_6(null(\hat{P}_5)) = \begin{bmatrix} 1/X_5 & 0 & 0 & -1/W_5 \\ 0 & 1/Y_5 & 0 & -1/W_5 \\ 0 & 0 & 1/Z_5 & -1/W_5 \end{bmatrix} P_6$$

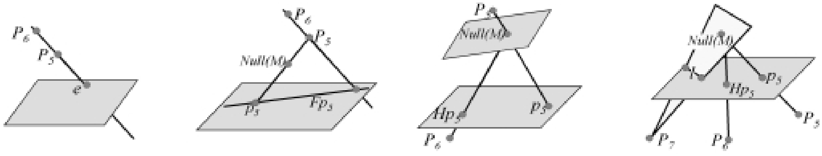
where the camera projection has its center equal to P_5. □

Claim. $e_{65} = null(\mathcal{F})$ and $e_{56} = null(\mathcal{F}^\top)$

Proof: Consider a camera matrix M whose null space is P_6. Such a camera maps P_6 to 0. Therefore, for any point $p_6 \in \mathcal{P}^2$, the lines l_6, l'_6 passing through p_6 satisfies: $l_6^\top MP_6 = 0$ and $l_6'^\top MP_6 = 0$. Therefore, if we take p_5 to be the projection of P_5 via M, than $p_6^\top \mathcal{F} p_5 = 0$, for all choices of $p_6 \in \mathcal{P}^2$. This of course implies that $\mathcal{F} p_5 = 0$. □

In the reduced case the two epipoles must coincide since $\mathcal{F}$ is a skew-symmetric matrix (thus $null(\mathcal{F}) = null(\mathcal{F}^\top)$). The epipole $e = e_{65} = e_{56}$ was coined the "dual epipole" in [11]. Since in the reduced case the image plane is stabilized, the dual epipole is simply the intersection of the line $\overline{P_5 P_6}$ with the stabilized image plane (see Fig. 1a). Note however, that this is true only for the reduced case. In general there are two dual epipoles which we will denote as "left" and "right" dual epipoles.

The dual image ray. Let l, l' be two lines coincident with the image point p. Then, $l^\top MP = 0$ and $l'^\top MP = 0$, therefore P has a 1-parameter degree of freedom, i.e., it is determined up to a line which is defined by the intersection of the two planes $l^\top M$ and $l'^\top M$. This line is the image ray corresponding to p. The same applies in the dual: the camera vector $\hat{M}$ is determined up to a line — the line passing through p and $null(\hat{P})$ (defined by $l^\top \hat{P}\hat{M} = 0$ and $l'^\top \hat{P}\hat{M} = 0$). This is defined as the *dual image ray*. Note that in the general case this mean that the camera center varies along a 1-parameter curve (non-linear). In the reduced case, however, $null(M)$ (the camera center) varies along a line — the line $\overline{Pp}$.



(a) (b) (c) (d)

Fig. 1. Stabilized reference plane. (a) the dual epipole. (b) The dual epipolar line $\mathcal{F}p_5$. (c) Homography duality. (d) Homography slice of the dual trilinear tensor.

Claim. In the reduced case, $p \cong \hat{P}\hat{M}$ constrains $null(M)$ to vary along the line $\overline{Pp}$, which is the line passing through p and P.

Proof: Note that for any line l that pass trough p

$$0 = l^\top \hat{P}\hat{M} = l^\top \begin{bmatrix} -W & 0 & 0 & X \\ 0 & -W & 0 & Y \\ 0 & 0 & -W & Z \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ -\delta \end{pmatrix}$$

where the camera projection on the right-hand side has its center at P, and $null(M) = (\alpha, \beta, \gamma, -\delta)^T$. Therefore, $null(M)$ is on $\overline{Pp}$.□

The dual epipolar line. Recall that the dual fundamental matrix $\mathcal{F}$ is a result of eliminating $\hat{M}$ from the equations $p_5 \cong \hat{P}_5\hat{M}$ and $p_6 \cong \hat{P}_6\hat{M}$. Recall also that $p_5 \cong \hat{P}_5\hat{M}$ constrains $\hat{M}$ to lie on the dual image ray of p_5. The projection of P_6, $MP_6 = \hat{P}_6\hat{M}$ is a line (lines are preserved under projection) in the image corresponding to all the vectors $\hat{M}$ that vary along the dual image ray of p_5. This line, $\mathcal{F}p_5$ is defined as the *dual epipolar line* . Note that in the reduced case, if $\hat{M}$ varies along the dual image ray of p_5, then $null(M)$ varies along the line $\overline{P_5p_5}$. Thus we obtain the simple geometric interpretation [11] drawn in Fig. 1b.

Homography Duality. The term "dual homography" is already used in classic projective geometry, thus we refer to the dual case of the homography matrix as "Homography duality". Recall that in multi-view, the homography matrix H_π induced by a plane π, defines $H_\pi p$ as the projection of P onto the second image plane, where P is the intersection point of the image ray corresponding to p with π. In the dual case, $H_\pi p_5$ is the projection of P_6 when the vector (point) $\hat{M}$ is at the intersection of the dual image ray of p_5 and the plane π. Let $[n^\top, \lambda]$ be the normal to the plane π, and denote by π' the plane defined by: $[n^\top, -\lambda]$. In the reduced case, constraining $\hat{M}$ to lie on π is equivalent to constraining $null(M)$ to lie on π'. So $null(M)$ is at the intersection of the line $\overline{P_5p_5}$ with π', thereby providing a simple geometric interpretation — see Fig. 1c.

The next three claims are provided without proofs due to lack of space:

Claim. Let $p_6^\top \mathcal{F} p_5 = 0$. Then $\mathcal{F}^\top H_\pi$ is skew-symmetric for all choices of planes π.

Claim 4 is crucial for later on when we discuss the source of the non-linear constraints of the dual trilinear tensor.

Claim. $[e_{65}]_\times H_\pi \cong \mathcal{F}$ for all π.

4.1 Trilinear Tensor Properties

Given the elements introduced above we will describe the source of the non-linear constraints of the dual trilinear tensor, but first we will introduce a useful equation:

Claim (dual reprojection equation). Let l^6 be a line coincident with the point p_6. Then,

$$p_5^i l_j^6 \mathcal{T}_i^{jk} \cong p_7 \tag{2}$$

Proof: $p_5^i l_j^6 \mathcal{T}_i^{jk}$ is a point (contravariant vector) q. Since $p_5^i l_j^6 l_k^7 \mathcal{T}_i^{jk} = 0$ for all lines l^7 coincident with p_7, then l^7 and q are coincident — hence, $q = p_7$.□

The reprojection equation maps the point p_5 and any line through p_6 onto the point p_7. It is dual to the multi-view reprojection where matching points in views 1,2 are mapped onto the matching point in view 3. The dual reprojection equation can be used for purposes of tracking (6 points predicting the position of the 7'th) and will be detailed further in the experimental section.

Homography slice. A single covariant contraction of the tensor produces a homography duality (just like in the multi-view tensor [16]).

Claim. The matrix $\delta_k \mathcal{T}_i^{jk}$ is a homography duality induced by the plane π defined by $null(\hat{P}_7)$ and the line δ in the image.

Proof: Recall that $p_5^i l_j^6 l_k^7 \mathcal{T}_i^{jk} = 0$ constrains $\hat{M}$ to lie at the intersection of the dual image ray of p_5, the plane $null(\hat{P}_6) \vee l^6$ and the plane $null(\hat{P}_7) \vee l^7$ (since $l^i \hat{P}_i \hat{M} = 0$). By a single contraction $\delta_k \mathcal{T}_i^{jk}$ we therefore constrain $\hat{M}$ to lie on the plane $null(\hat{P}_7) \vee \delta$.□

Note that in the reduced case, this also constrains $null(M)$ to lie on the plane $P_7 \vee \delta$ (see claim 4), thus we obtain the geometric interpretation shown in Fig. 1d.

Non-linear Constraints. We should expect 4 non-linear constraints on the trilinear tensor. Consider three homography slices: $\mathcal{T}_i^{j1}, \mathcal{T}_i^{j2}, \mathcal{T}_i^{j3}$ and denote them as H_1, H_2, H_3. From Claim 4 we have that $H_1^\top \mathcal{F}$ provides 6 linear constraints on $\mathcal{F}$ and so do $H_2^\top \mathcal{F}$ and $H_3^\top \mathcal{F}$ — taken together 18 linear constraints. Choose 8 of these constraints, then the entries of $\mathcal{F}$ are represented by 8×8 determinant expansions from the 8×9 estimation matrix constructed from the 8 constraints. Each of these determinant expansions is a polynomial in the entries of $\mathcal{T}_i^{jk}$. The remaining 10 constraints are of rank 4 because only 4 constraints are required to specify $\mathcal{F}$ $(9-1-4=4)$. Therefore by substituting the representation of the entries of $\mathcal{F}$ as determinant expansions in the remaining 10 constraints (choose any 4 of them) we obtain 4 polynomials in the entries of $\mathcal{T}_i^{jk}$.

Reconstruction. All the information required for the 3D reconstruction task, is contained in the dual epipoles. In the case of the 6-point tensor, Let $\mathcal{F}$ be recovered from image measurements (we need at least 4 views of the 6 points). The dual epipoles satisfy $\mathcal{F}e_{65}=0$ and $\mathcal{F}^\top e_{56}=0$. Point P_5 can be assigned $(1,1,1,1)$ (to complete the projective basis) and we are left with recovering point P_6. Recall: $e_{65}\cong \hat{P}_6 null(\hat{P}_5)=(X_6-W_6, Y_6-W_6, Z_6-W_6)^\top$ and $e_{65}=\hat{P}_5 null(\hat{P}_6)=((X_6-W_6)/X_6, (Y_6-W_6)/Y_6, (Z_6-W_6)/Z_6)^\top$.

Therefore, taking the ratio (component-wise) of e_{56}/e_{65} would produce X_6, Y_6, Z_6 up to a common scale, and W_6 can be recovered by substitution in e_{65} thus obtaining a linear solution for P_6 (up to scale). This reconstruction approach easily generalizes to the trilinear and quadlinear tensors. First one recovers the dual epipoles from the tensor and the reconstruction proceeds from there in a similar manner as with the 6-point tensor.

Regarding the reconstruction from the stabilized reference plane tensor, we first recall that P_6 has only two parameters. On the other hand, in the reduced case the left and right epipoles coincide, so recovering the dual epipole from $\mathcal{F}$ provides only two constraints on P_6, and this is the most one can expect.

In the trilinear tensor case, the points P_6, P_7 contains 5 parameters. Therefore considering only the epipoles $e_{56}=e_{65}$ and $e_{57}=e_{75}$ provides only 4 constraints, which is not sufficient. Therefore, one should use the third epipole $e_{76}=e_{67}$ as well. Note that in the reduced case, the epipole associated with the points P_i, P_j has the form: $e_{ij}\cong(X_i-X_j, Y_i-Y_j, Z_i-Z_j)^\top$ (assuming that $W_i=W_j=1$). So the three epipoles provides us with the system:

$$
\begin{gathered}
e_{56}(3)(X_6-1)=e_{56}(1)(Z_6-1)\\
e_{56}(3)(Y_6-1)=e_{56}(2)(Z_6-1)\\
e_{57}(3)(X_7-1)=e_{57}(1)(Z_7-1)\\
e_{57}(3)(Y_7-1)=e_{57}(2)(Z_7-1)\\
e_{67}(3)(X_6-X_7)=e_{67}(1)(Z_6-Z_7)\\
e_{67}(3)(Y_6-Y_7)=e_{67}(2)(Z_6-Z_7)
\end{gathered}
$$

where $e_{ij}(k)$ is the k'th coordinate of the epipole e_{ij}. Though this system contains 6 equations, it has one trivial solution $X_6=Y_6=Z_6=X_7=Y_7=Z_7=1$. If the system has a non-trivial solution, it must be of rank 5 most. Therefore the system enables the recovery of P_6, P_7 up to one degree of freedom, as expected. Similarly, we can achieve 3D reconstruction from the quadlinear tensor, using 4 out of the 6 epipoles. Note, as pointed out in [18], that in the reduced case it possible to collect any number of images and perform factorization — however, points on the reference plane would not be reconstructed (and points near the reference plane would be subject to unstable reconstruction). Thus, the reconstruction formulas described above are of interest as they hold generally.

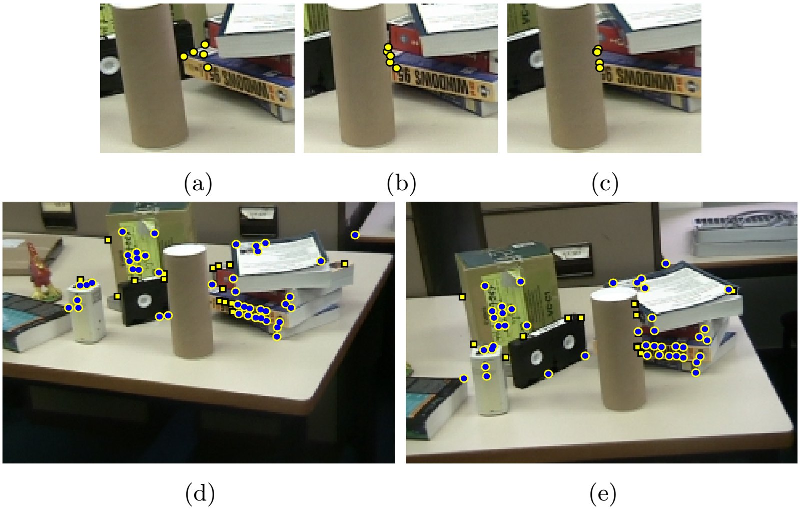
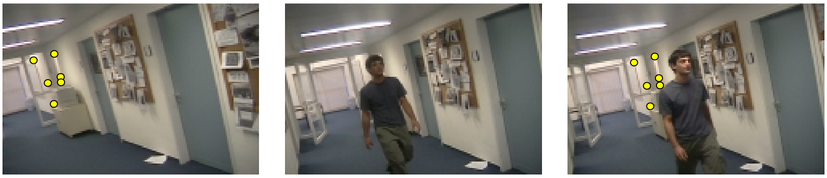


Fig. 2. Separating badly tracked points due to occlusions. (a,b) two images of a sequence of 100 — one of the problematic areas is zoomed in. (c) the 6-point tensor evaluated for all points. The badly tracked points are clustered to the right and (except one mistake) are well separated from the good points. (d,e) two views of the sequence with the labeling of good and bad points.

5 Experiments

We describe three applications. The first is occlusion detection in a point tracking experiment. Fig. 2(a,b) displays two images from a 100 frames sequence, and a set of points that were located and tracked by an automatic tracker (openCV's [15] pyramid LK tracker). In the first image the tracker located some points on the edge of the books, but as the cylinder starts covering those edges the tracked points on the edges start moving as well and loose their accuracy. We use the 6-point tensor to evaluate configurations of 6 points in the following manner. The scene contained roughly 50 feature points which were automatically detected and tracked. For each of those points, 50 quintets of points were randomly selected. For each such sextet (the tested point + the quintet), we computed robustly (using LMeds [14]) the 6-points tensor over the entire sequence. For badly tracked points, we expect that for each sextet, the evaluation error (the contraction of the tensor with points p_5, p_5) will be high. For a good point, we except that out of the 50 random sets, there would be enough good sets, so that the sum of errors achieved this way will be significantly lower than the sum of errors achieved for a badly tracked point. Fog. 2(c) displays the error graph — one can observe a good separation between the badly tracked points (clustered on the right) and the good points (modulo one exception). Fig. 2(d-e) presents two images from the sequence with the labeling of good and bad points.



(a) (b) (c)

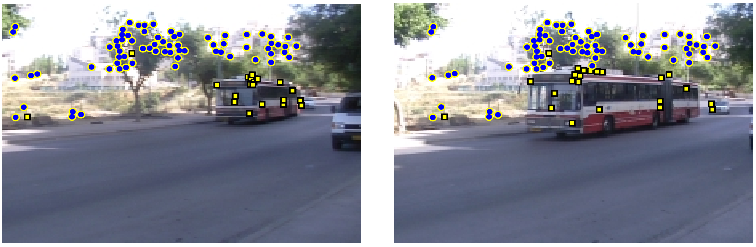
Fig. 3. Using the dual trilinear tensor to resume tracking of lost points. The images are part of a sequence of 160 images in which roughly 50 points were detected in frame 1 and tracked throughout the sequence. The points overlaid in (a) are those which were lost as some point due to the passing person (b) and recovered in (c).

The second application using the 7-point dual tensor reprojection equation (eqn. 2) describes a tracking recovery due to interference. In other words, situations in which the tracker looses a point due to low confidence yet in future frames the point reappears and one would like to resume tracking it. Fig. 3(a-c) shows a sample of 3 images, out of a sequence of 160. Roughly 50 points were automatically located and tracked in the beginning of the sequence. Some of those points were occluded due to the presence of a moving person and the goal was to recover the lost points once the occlusion was removed. For each lost point we wish to recover, we randomly chose a set of sextets of points out of the set of points that were successfully tracked all other the movie. For each such septet (the sextet+the point we wish to recover), we compute a 7-points dual trilinear tensor, using the images at the beginning of the movie, where we had all the 7 points. After the 7th point was lost, we used the computed shape tensor and the projections of the first 6 points in order to recover the 7th point. The recovered points that are shown in Fig. 3(c), are the average of the results that were achieved using the set of sextets we chose.

The third application is motion segmentation. Fig. 4 shows two images out of a sequence of 40 images of a moving bus (taken from a moving camera) where roughly 95 points were automatically located and tracked along the sequence. The segmentation technique uses the 6-point dual tensor in a manner similar to the first tracking application.

6 Summary

Single-view shape invariants were introduced in the past, however, much of the underlying constraints and forms remained incomplete. In this paper we have introduced a full account of the dual tensors which have exactly the same form as the multi-view tensors but with additional constraints borne out of the special structure of the dual projection matrices. The first difference lies with the internal constraints (synthetic constraints) — we have shown that the trilinear



(a) (b)

Fig. 4. Two images out of a sequence of 40 images of a moving bus taken from a moving camera. The 6-point dual tensor was used for segmentation.

tensor has 16 of them and the quadlinear tensor has 58. The nature of these constraints change when a plane has been identified in advance and stabilized (the reduced case). There we have shown that the trilinear tensor has 21 constraints and the quadlinear tensor 72 constraints. We have introduced the dual of the tensor slices, the reprojection equation, the dual epipoles, the homography duality, the non-linear constraints of the trilinear tensor, and reconstruction from dual epipoles.

Given the understanding of the internal constraints and the introduction of the dual reprojection equation we made use of the shape tensors for two applications of tracking and an application of 2-body segmentation. The hope is that with a better understanding of the underlying internal structure of these shape invariants the applications (which have so far been a few) using these invariants would increase as well.

References

1. N. Canterakis. A minimal set of constraints for the trifocal tensor. In *Proceedings of the European Conference on Computer Vision*, Dublin, Ireland, June 2000.
2. S. Carlsson. Duality of reconstruction and positioning from projective views. In *Proceedings of the workshop on Scene Representations*, Cambridge, MA., June 1995.
3. S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from multiple images. *International Journal of Computer Vision*, 27(3), 1998.
4. A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.
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. W. Fulton and J. Harris. *Representation Theory.* Springer, 1991.

7. R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.
8. R.I. Hartley and A. Zisserman. *Multiple View Geometry.* Cambridge University Press, 2000.
9. A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the International Conference on Computer Vision*, pages 1058–1063, Cambridge, MA, June 1995.
10. A. Heyden. A common framework for multiple view tensors. In *Proceedings of the European Conference on Computer Vision*, pages 3–19, Freiburg, Germany, June 1998.
11. 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.
12. 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.
13. D.W. Jacobs. Space efficient 3D model indexing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 439–444, 1992.
14. P. Meer, D. Mintz, D. Kim and A. Rosenfeld. Robust regression methods for computer vision: A review. *International Journal of Computer Vision* 6(1), 1991.
15. Open source computer vision library *http://www.intel.com/research/mrl/research/cvlib/*
16. 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.
17. A. Shashua and Lior Wolf. On the structure and properties of the quadrifocal tensor. In *Proceedings of the European Conference on Computer Vision*, Dublin, Ireland, June 2000.
18. C. Rother and S. Carlsson. Linear Multi View Reconstruction and Camera Recovery. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.
19. G. Stein and A. Shashua. On degeneracy of linear reconstruction from three views: Linear line complex and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):244–251, 1999.
20. B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995.
21. D. Weinshall, M. Werman, and A. Shashua. Duality of multi-point and multi-frame geometry: Fundamental shape matrices and tensors. In *Proceedings of the European Conference on Computer Vision*, LNCS 1065, pages 217–227, Cambridge, UK, April 1996. Springer-Verlag.