On the Reprojection of 3D and 2D Scenes without Explicit Model Selection

Amnon Shashua[1] and Shai Avidan[2]

[1] School of CS and Engineering, Jerusalem 91904, Israel
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

Abstract. It is known that recovering projection matrices from planar configurations is ambiguous, thus, posing the problem of model selection — is the scene planar (2D) or non-planar (3D)? For a 2D scene one would recover a homography matrix, whereas for a 3D scene one would recover the fundamental matrix or trifocal tensor. The task of model selection is especially problematic when the scene is neither 2D nor 3D — for example a "thin" volume in space.
In this paper we show that for certain tasks, such as reprojection, there is no need to select a model. The ambiguity that arises from a 2D scene is orthogonal to the reprojection process, thus if one desires to use multilinear matching constraints for transferring points along a sequence of views it is possible to do so under any situation of 2D, 3D or "thin" volumes.

1 Introduction

There are certain mathematical objects connected with multiple-view analysis which include: (i) homography matrix (2D collineation), and (ii) objects associated with multilinear constraints — fundamental matrix, trifocal and quadrifocal tensors. Given two views of a planar configuration of features (points or lines) it is possible to recover the mapping between the views as a 2D collineation (homography matrix) — a transformation that is also valid when the scene is 3D but the relative camera geometry consists of a pure rotation. On the other hand, when the camera motion is general and the scene consists of a three-dimensional configuration of features, then the valid transformations across a number of views consists of multi-linear relations that perform a variety of point-to-line mappings. The coefficients of the multilinear constraints encode the relative camera geometry, the projection matrices, and form a matrix in two views, a $3 \times 3 \times 3$ tensor in three views and a $3 \times 3 \times 3 \times 3$ tensor in four views.

The objects of multi-view analysis are often used for purposes of reconstruction, i.e., 3D modeling from a collection of views, and for purposes of feature-transfer (reprojection) i.e., predict the image location of a point (line) in some view given its locations in two other views. The reprojection paradigm is useful for feature tracking along image sequences, mosaicing, and image based rendering.

D. Vernon (Ed.): ECCV 2000, LNCS 1842, pp. 936–949, 2000.

Regardless of the application, it seems necessary to know *in advance* whether the scene viewed by the collection of images is 2D or 3D. Because, in case the scene is 2D the multilinear constraints are subject to an ambiguity — rank-6 estimation matrix (instead of 8) for the fundamental matrix and rank-21 (instead of 26) for the trifocal tensor. Hence arises the issue of *model selection*. There has been a large body of research in the general area of model selection for purposes of segmentation (due to shape, motion), and field of view (orthographic versus perspective) [14]. Whatever the scheme of model selection is chosen, it is problematic in the sense that often a decision is to be made in uncertain conditions — in our case, for example, when the scene is neither purely planar nor spans a sufficiently large 3D volume.

In this paper we show that in the case of multilinear constraints, it is not necessary to decide on a *model* i.e., whether a homography matrix is better suited than a fundamental matrix for example, for purposes of reprojection. Our results show that the null space, or the ambiguity space in general, of the estimation of multilinear constraints (fundamental matrix and trifocal tensor) is *orthogonal* to the task of reprojection. In other words, in a situation of three views of a planar scene the 6-dimensional null space of the trifocal tensor estimation is completely admissible for reprojection of features arising from the planar surface. Moreover, generally the space of uncertainty in recovering certain parameters of the tensor due to insufficient "3D volume" of the sampled surface is again orthogonal to reprojection of features arising from the sampled volume.

2 Notations and Necessary Background

We will be working with the projective 3D space and the projective plane. In this section we will describe the basic elements we will be working with (i) homography matrix, (ii) camera projection matrices, (iii) fundamental matrix, (iv) tensor notations, and (v) trifocal tensor.

A point in the projective plane is defined by three numbers, not all zero, that form a coordinate vector defined up to a scale factor. In the projective plane any four points in general position can be uniquely mapped to any other fours points in the projective plane. Such a mapping is called *collineation* and is defined by 3×3 invertible matrices, defined up to scale. These matrices are sometimes referred to as *homographies*. A collineation is defined by 4 pairs of matching points, each pair provides two linear constraints on the entries of the homography matrix. If A is a homography matrices defined by 4 matching pairs of points, then A^{-T} (inverse transpose) is the dual homography that maps lines onto lines.

The projective plane is useful to model the image plane. Consider a collection of planar points $P_1, ..., P_n$ 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 A_π that satisfies the relation $A_\pi p_i \cong p'_i$, $i = 1, ..., n$, and where A_π is uniquely determined by 4 matching pairs from the set of n matching pairs.

A point in 3D projective space is defined by four numbers, not all zero, that form a coordinate vector defined up to a scale factor. A camera projection is a 3×4 matrix which corresponds between points in 3D projective space to points in the projective plane. A useful parameterization (which is the one we adopt in this paper) is to have the 3D coordinate frame and the 2D coordinate frame of view 1 aligned. Thus, in the case we have three views, then the three camera projection matrices between the 3D projective space and the three image planes are denoted by $[I;0],[A;v'],[B;v'']$ associated with views 1,2,3 respectively. These camera matrices are not uniquely defined, as there is a 3-parameter degree of freedom ("gauge" of the system) as $[I;0],[A+v'w^\top;v'],[B+v''w^\top;v'']$ agree with the same image data for all choices of w. The multi-view tensors which we will define next are gauge-invariant, i.e., they are invariant to the choice of w.

The 3×3 principle minor of the camera matrix, under this kind of parameterization, is a homography matrix. The choice of gauge parameters determine the position of the plane associated with the homography — the reference plane. In particular, the space of all homography matrices between views 1,2 (up to scale) is $A + v'w^\top$.

The simplest multi-view tensor is the fundamental matrix $F = [v]_x A$ whose entries are the coefficients of the bilinear matching constraint $p'^\top F p = 0$, where p, p' are matching points in views 1,2 respectively. Note that F is gauge invariant as $[v]_x(A + v'w^\top) = [v]_x A$.

It will be most convenient to use tensor notations from now on because the multi-view tensors couple together pieces from different projections matrices into a "joint" object. When working with tensor objects the distinction of when coordinate vectors stand for points or lines matters. A point is an object whose coordinates are specified with superscripts, i.e., $p^i = (p^1, p^2, p^3)$. These are called contravariant vectors. A line in $\mathcal{P}^2$ is called a covariant vector and is represented by subscripts, i.e., $s_j = (s_1, s_2, s_3)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $p^i s_i = p^1 s_1 + p^2 s_2 + p^3 s_3$. 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 (lines in $\mathcal{P}^2$) to hyperplanes, because $a_i^j p^i = q^j$ and $a_i^j s_j = r_i$ (in matrix form: $Ap = q$ and $A^\top s = r$); 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 $C = ba^\top$. 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 line in the second coordinate frame into a point in the third coordinate frame. The "trifocal" tensor in multiple-view geometry is an example of such a tensor. 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.

The $3 \times 3 \times 3$ trifocal tensor is defined below:

$$\mathcal{T}_i^{jk} = v'^j b_i^k - v''^k a_i^j. \tag{1}$$

The elements of the tensor are coefficients of trilinear constraints on triplets of matching points across the three views. Let p, p', p'' be a matching triplet of points, i.e., they are projections of some point in 3D. Let s be some line coincident with p', i.e., $s_j p'^j = 0$, and let r be some line through p''. Then,

$$p^i s_j r_k \mathcal{T}_i^{ij} = 0. \tag{2}$$

Because p' is spanned by two lines (say, the horizontal and vertical scan lines) and r as well, a triplet p, p', p'' generate 4 "trilinearities" each is a linear constraint on the elements of the tensor. Thus 7 matching points (or more) are sufficient to solve for the tensor. Note that the trifocal tensor is also gauge invariant as:

$$\begin{aligned}
\mathcal{T}_i^{jk} &= v'^j b_i^k - v''^k a_i^j & (3)\\
&= v'^j (b_i^k + w_i v''^k) - v''^k (a_i^j + w_i v'^j) & (4)\\
&= \mathcal{T}_i^{jk} + w_i v'^j v''^k - w_i v'^j v''^k & (5)\\
&= \mathcal{T}_i^{jk} & (6)
\end{aligned}$$

Once the trifocal tensor is recovered from image measurements (matching triplets, or matching lines, or matching points and lines) the task of "reconstruction" is to extract the camera projection matrices (up to a choice of gauge parameters) from the tensor. We will not discuss this here. The task of "reprojection" is to predict (or "back-project") the location of p'' using the matching pair p, p' and the tensor. This is done simply as:

$$p^i s_j \mathcal{T}_i^{jk} \cong p''^k$$

and since there are two choices for s we have a redundant system for extracting p''.

These were the necessary details we need for the rest of the paper. More details on the trifocal tensor can be found in the review [10] and in (not exhaustive) [9,4,11,12,3,5,16,6].

3 Reprojection of a Planar Surface from Multilinear Constraints

In case the scene is indeed 3D there is a one-to-one mapping between tensors that satisfy Eqn. 1 and tensors that satisfy Eqn. 2. However, when the scene is planar then a tensor that satisfies Eqn. 2 does not necessarily satisfy Eqn. 1, as we will see now.

Consider a collection of matching point triplets p, p', p'' of a **planar** scene π in views 1,2,3, respectively. Because the scene is planar there exist a unique homography matrix A from views 1 to 2, i.e., $Ap \cong p'$ and a unique homography matrix B from views 1 to 3, i.e., $Bp \cong p''$. Let δ, μ be arbitrary vectors, then the tensor

$$\mathcal{T}_i^{jk} = \delta^j b_i^k - \mu^k a_i^j, \tag{7}$$

satisfies the trilinearity - Eqn. 2. To see why this is so, note that $s^\top Ap = 0$ and $r^\top Bp = 0$ for triplets p, p', p'' arising from the plane π. We have therefore:

$$\begin{aligned} p^i s_j r_k \mathcal{T}_i^{jk} &= p^i s_j r_k (\delta^j b_i^k - \mu^k a_i^j) \\ &= (s_j \delta^j)(r_k b_i^k p^i) - (r_k \mu^k)(s_j a_i^j p^i) \\ &= (s^\top \delta)(r^\top Bp) - (r^\top \mu)(s^\top Ap) = 0, \end{aligned}$$

and this holds for all choices of the vectors δ, μ. As argued in [13] this entails that the rank of the estimation matrix for the trifocal tensor from measurements arising from a planar surface is at most 21 (instead of 26). In other words, there are 6 degrees of freedom due to the indeterminacy of the epipoles (δ, μ). What is left to show is that all the solutions in the null space are in the form of Eqn. 7. To see that note that A, B can be homographies due to any other plane $\bar{\pi}$ and still satisfy the trilinearity (Eqn. 2) if and only if $\delta = v'$ and $\mu = v''$ are the true epipoles: Let $\bar{A} = \lambda A + v' n^\top$ and $\bar{B} = \lambda B + v'' n^\top$ be the homographies associated with the plane $\bar{\pi}$, then

$$v'^j b_i^k - v''^k a_i^j \cong v'^j (\lambda b_i^k + n_i v''^k) - v''^k (\lambda a_i^j + n_i v'^j)$$

for all choices of λ, n and thus in particular

$$v'^j b_i^k - v''^k a_i^j \cong v'^j \bar{b}_i^k - v''^k \bar{a}_i^j.$$

To conclude, because the epipoles cannot be determined from the trilinearities (Eqn. 2) then all the tensors in the null space are of the form of Eqn. 7 where A, B are the homographies due to the plane π.

We have, therefore, an ambiguity whose source arises from the uncertainty in recovering the epipoles from the image measurements. Thus, recovering projection matrices is not possible. Yet, consider the problem of reprojection:

$$\begin{aligned} p^i s_j \mathcal{T}_i^{jk} &= p^i s_j (\delta^j b_i^k - \mu^k a_i^j) \\ &= (s_j \delta^j) b_i^k p^i - \mu^k (s_j a_i^j p^i) \\ &= (s^\top \delta) Bp - (s^\top Ap)\mu \cong p''. \end{aligned}$$

In other words, for all choices of δ, μ, a matching point and line in views 1,2 uniquely determine the location of the matching point in view 3, *provided* that the matching triplet p, p', p'' arise from the plane π. We can conclude, therefore, that the null space for estimating the trifocal tensor from image measurements arising from a planar surface is *orthogonal* to the reprojection equation $p^i s_j \mathcal{T}_i^{jk}$ where the matching points arise from the same planar surface that was sampled in the process of recovering $\mathcal{T}_i^{jk}$.

In practical terms, given a collection of matching triplets p, p', p'' sampling a certain volume in space, each triplet provides 4 linear equations for the trifocal tensor. The eigenvector associated with the smallest eigenvalue of the estimation matrix is the trifocal tensor. In case the matching triplets came from a 3D scene, the solution is unique whereas in case the the matching triplets came from a planar configuration the solution is not unique (the 6 eigenvectors corresponding to the 6 smallest eigenvalues span the solution space) — but that does not matter, as long as the matching points used for the estimation of the tensor span the scene volume of interest (if the points came from a plane it means that the scene is planar, for example), then the reprojection is valid nevertheless. The following theorem summarizes the findings so far:

Theorem 1. *In case a collection of matching triplets p, p', p'' whose corresponding 3D points sample some volume in space are given, then the eigenvector associated with the smallest eigenvalue of the estimation matrix to the trifocal tensor forms a trifocal tensor that is valid for reprojecting point p, p' onto p'', regardless of whether the volume is a 2D plane or a 3D volume, provided that the corresponding 3D points come from the same volume in space sampled during the estimation process.*

This state of matters is not characteristic solely to the trifocal tensor, it is a general geometric property. Consider performing reprojection using pairwise fundamental matrices, for example. Let F_{13} be the fundamental matrix satisfying $p''^\top F_{13} p = 0$ for all matching pairs p, p'', and let F_{23} be the fundamental matrix satisfying $p''^\top F_{23} p' = 0$ for all matching pairs p', p''. The reprojection equation is an intersection of epipolar lines:

$$p'' \cong F_{13} p \times F_{23} p'.$$

Generally it is not a good idea to rely on epipolar intersection as it becomes degenerate when the three camera centers are collinear, but nevertheless this provides an alternative to the reprojection equation using the trifocal tensor. When the triplet p, p', p'' arise from a planar configuration, then $F_{13} = [\delta]_x B$ and $F_{23} = [\mu]_x B A^{-1}$ satisfy the bilinear constraints $p''^\top F_{13} p = 0$ and $p''^\top F_{23} p' = 0$, for all choices of the vectors δ, μ. Thus, the rank of the estimation matrix for the fundamental matrix becomes 6 (instead of 8). Reprojection, however, is unaffected by the choice of δ, μ *provided* that the pairs p, p' to be reprojected arise from the same planar surface that was sampled in the process of recovering F_{13} and F_{13}:

$$\begin{aligned} F_{13} p \times F_{23} p' &= ([\delta]_x B p) \times ([\mu]_x B A^{-1} p') \\ &= (\delta \times p'') \times (\mu \times p'') \cong p'' \end{aligned}$$

Note that unlike the trifocal tensor estimation that requires a triplet p, p', p'' of matching points in the estimation process, here the requirement is pairs of matching pairs p, p'' and p', p'' that do not necessarily arise from the same point in 3D. This raises the possibility, for example, that F_{23} is estimated from a 3D scene, yet F_{13} is estimated from a planar scene. The process of reprojection would remain valid nevertheless *provided that the points* p, p' *used for reprojection arise from a surface whose dimensionality is lesser or equal to the dimensionalities of the surfaces used for estimation of* F_{13} *and* F_{23}.

4 Sensitivity Analysis of "thin" Volumes

We have seen in the previous section that the ambiguity of the tensor estimation in the presence of a planar configuration of points does not affect the reprojection process of points coming from the planar surface. In this section we wish to investigate the reprojection process for "thin" volumes — the point configuration does not form a 2D plane but almost does so (shallow surface, aerial photograph, for example). Strictly speaking, a point configuration can be either 2D (plane) or 3D (non-coplanar), there is no in-between. But, in practice it is important to investigate the (numerical) sensitivity of the reprojection process in order to be convinced that the transition between planar and 3D is a continuous one. In other words, we would like to establish the fact that the estimation of the trifocal tensor, from a point configuration that spans any volume in 3D space, will produce a valid reprojection of that volume.

We wish to show that all tensors that can be recovered from a "thin" volume are equal to the first order. To do so, think of a "thin" volume as two planes infinitesimally separated (to be defined later). We will show that any form of indeterminacy of the epipoles (whether complete or partial) leads to at most a second order error in the infinitesimal variables — hence can be neglected. In other words, we will employ infinitesimal calculus (see [2]) of the first order in our investigation, such that if ϵ is an infinitesimal variable in a calculation, then $\epsilon^2 = 0$ (and higher orders).

We will first consider the estimation of the trifocal tensor from a point configuration arising from two distinct planes $\pi, \bar{\pi}$. Let A, B be the homography matrices due to π from views 1 to 2 and from views 1 to 3, respectively, and let $\bar{A}, \bar{B}$ be the homographies due to $\bar{\pi}$. Then, there exist λ, n that satisfy:

$$\bar{A} = \lambda A + v' n^\top$$
$$\bar{B} = \lambda B + v'' n^\top,$$

where v', v'' are the epipoles in views 2,3 respectively (projection of the first camera center onto views 2,3). The vector n is the projection on view 1 of the intersecting line between $\pi, \bar{\pi}$ and $(n^\top, \lambda)$ is the plane passing through the first camera center and the line n in view 1. Let the space of solutions to the trifocal tensor arising from matching triplets corresponding to π be

$$\delta^j b_i^k - \mu^k a_i^j$$

where δ, μ are free vectors, and let the space of solutions arising from $\bar{\pi}$ be

$$\bar{\delta}^j \bar{b}_i^k - \bar{\mu}^k \bar{a}_i^j$$

where $\bar{\delta}, \bar{\mu}$ are free vectors. The space of solutions arising from measurements corresponding to both $\pi, \bar{\pi}$ is the intersection of the null spaces, i.e., we wish to find $\delta, \mu, \bar{\delta}, \bar{\mu}$ that satisfy

$$\delta^j b_i^k - \mu^k a_i^j = \bar{\delta}^j \bar{b}_i^k - \bar{\mu}^k \bar{a}_i^j$$

After rearranging terms:

$$(\delta^j - \lambda\bar{\delta}^j) b_i^k - (\mu^k - \lambda\bar{\mu}^k) a_i^j = n_i(\bar{\delta}^j v''^k - \bar{\mu}^k v'^j).$$

Since the left-hand side is at least a rank-4 tensor (A, B cannot be lower than rank-2) and the right-hand side is a rank-2 tensor, equality can hold only if $\delta = \lambda\bar{\delta}$ and $\mu = \lambda\bar{\mu}$. Thus, $\bar{\delta}, \bar{\mu}$ must satisfy

$$\bar{\delta}^j v''^k - \bar{\mu}^k v'^j = 0,$$

which could happen if and only if $\bar{\delta} = \alpha v'$ and $\bar{\mu} = \alpha v''$ for all α. Taken together, the intersection of the null spaces is a unique tensor:

$$v'^j b_i^k - v''^k a_i^j.$$

The derivation above is simply another route for proving the existence and form of the trifocal tensor from image measurements arising from points matches corresponding to a 3D set of points. However, it is shown that two planes are sufficient for a unique determination (two distinct planes and the camera center of view 1 forms a simplex). Analogously, the fundamental matrix between views 1,2 is known to be uniquely determined from the relationship: $A^\top F + F^\top A = 0$ and $\bar{A}^\top F + F^\top \bar{A} = 0$ and the proof follows the same lines as above.

We will use this line of derivation of the trifocal tensor to consider next the situation where the two planes $\pi, \bar{\pi}$ are infinitesimally separated. This is defined by letting $\bar{A}, \bar{B}$ be defined as:

$$\bar{A} = \lambda A + dA$$
$$\bar{B} = \lambda B + dB$$

where dA, dB are matrices whose entries are infinitesimal to the first order, i.e., higher orders of these variables can be neglected. Because dA, dB may be arbitrary (i.e., v', v'' are completely masked out in the presence of noise) the null spaces may not have a common intersection. But, instead of an intersection we are looking for $\delta, \mu, \bar{\delta}, \bar{\mu}$ such that the null spaces have a common infinitesimal locus, i.e., a locus that is defined by second (or higher) order terms of dA, dB. In other words, let $\mathcal{T}(\delta, \mu)$ be the space of tensors (null space) of the form $\delta^j b_i^k - \mu^k a_i^j$ and let $\bar{\mathcal{T}}(\bar{\delta}, \bar{\mu})$ be the space of tensors $\bar{\delta}^j \bar{b}_i^k - \bar{\mu}^k \bar{a}_i^j$, then we are looking for $\delta, \mu, \bar{\delta}, \bar{\mu}$ such that $\mathcal{T}(\delta, \mu) - \bar{\mathcal{T}}(\bar{\delta}, \bar{\mu}) =_{inf} 0$ where the symbol $=_{inf}$ denotes equality up to second order terms of infinitesimal variables.

Let $\delta = \lambda\bar{\delta}$ and $\mu = \lambda\bar{\mu}$, then $\bar{\delta}, \bar{\mu}$ must satisfy

$$\bar{\delta}^j db_i^k - \bar{\mu}^k da_i^j =_{inf} 0.$$

Let $\bar{\delta}$ take some linear combination of the rows of dA and let μ be equal to some linear combination of the rows of dB (see Appendix for proof that such a choice corresponds to an L_2 norm minimization of the expression above). Then, the expression above involves bilinear products of infinitesimal variables — thus the equality of the first order is achieved, i.e. $\mathcal{T}(\delta, \mu) - \bar{\mathcal{T}}(\bar{\delta}, \bar{\mu}) =_{inf} 0$ for the choices we made. The theorem below summarizes the findings above:

Theorem 2. *In the case where the trifocal tensor is estimated from point matches coming from an infinitesimally thin volume in space, then in the worst case condition (measurement noise completely masks out the location of the epipoles v', v''), the solutions in the null space are valid for reprojection of points of the sampled volume — upto a measure zero of infinitesimal variation.*

5 Experiments

We show results on three real image sequences. In all cases we use a progressive scan Canon ELURA DV cam-corder that produces RGB images of size 720×480 pixels. We use the KLT package [1] to automatically detect and track a list of interest points throughout the sequence (about 100 points on average). We then estimate the trifocal tensor on successive triplets of frames, reprojecting points from the first two frames in the triplet to the third. The trifocal tensor is estimated using the method described in this paper with the usual LMeDS (Least Median of Squares) [7,15]. Specifically, the algorithm proceeds as follows. Sets of seven points are sampled randomly from the set of all matching points. The estimation matrix is constructed and the tensor is taken to be the eigenvector that corresponds to the smallest eigenvalue. Then we measure the reprojection error of the recovered tensor for the rest of the points and take the median of the reprojection error as the score of this tensor. The process is repeated for 50 times. The tensor with the lowest score is the winner. We then recompute the tensor, using the same method, but now with all the points whose reprojection error is lower than the score of this tensor.

Experiment 1 We move the camera from a "volumetric" scene to a very shallow scene gathering 36 images as we move. We compute the trifocal tensor of successive triplets of images and reprojected the points in the first two images to the third. Figure 1 shows some of the 36 images, with the tracked and reprojected points super-imposed. The average reprojection error is about 0.5 pixels. More interestingly, we plotted the average reprojection error across the 36 images and did not find a clear correlation between the reprojection error and the "thickness" of the 3D scene. Recall, the camera is moving from a full 3D scene to a very shallow scene.

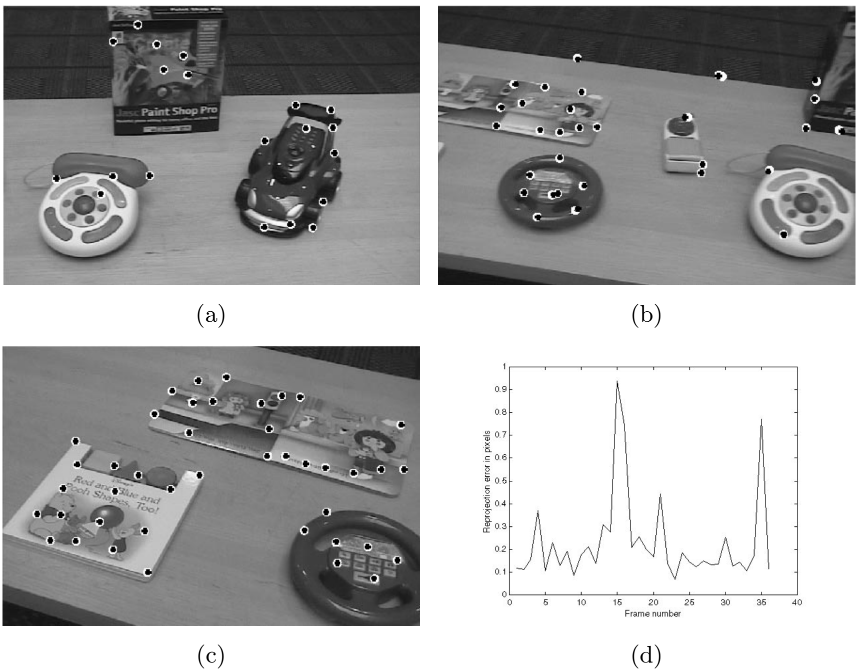
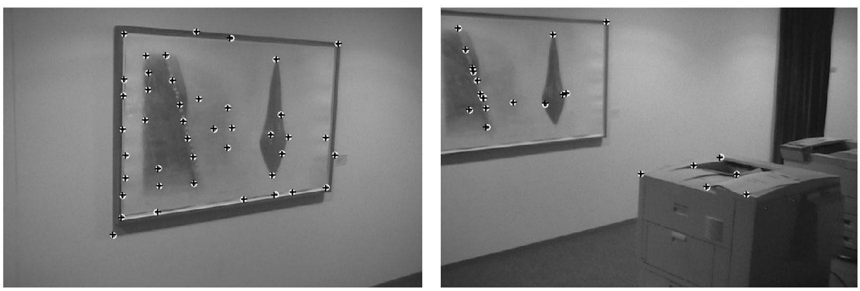


Fig. 1. ((a),(b),(c)) First, middle and last images in a 36 long image sequence. White circles represent the tracked points. Black crosses represent the reprojected points. Average reprojection error is 0.5 pixels. (d) shows reprojection error, in pixels, across the 36 images of sequence. There is no clear correlation between the reprojection error and the volume of the 3D scene. Note that the camera is moving from a full 3D scene to a very shallow 3D scene.

Experiment 2 This experiment is similar to the previous one, only this time the camera moves from a planar scene to a full volume 3D scene, gathering 26 images as it moves. Again, we compute the trifocal tensor of successive triplets of images and reprojected the points in the first two images to the third. Figure 3 shows some of the 26 images, with the tracked and reprojected points superimposed. The average reprojection error is about 0.5 pixels. Again, we plotted the average reprojection error across the 26 images and did not find a clear correlation between the reprojection error and the "thickness" of the 3D scene. Recall, this time the camera is moving from a planar scene to a full 3D scene.



(a) (b)

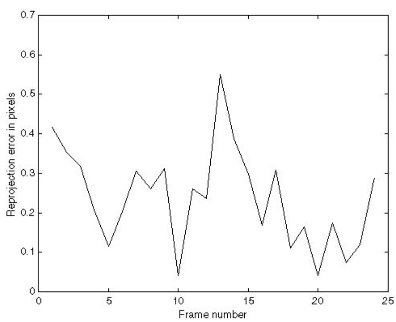


Fig. 2. ((a),(b)) First and last in a sequence of 26 images. White circles represent the tracked points. Black crosses represent the reprojected points. Average reprojection error is 0.5 pixels. (c) shows reprojection error, in pixels, across the 26 images of sequence. There is no clear correlation between the reprojection error and the volume of the 3D scene. Note that the camera is moving from a planar scene to a full 3D scene.

Experiment 3 In this experiment we demonstrate the reprojection power of our method given the same camera configuration but using different sections of the 3D scene. We repeated the experiment twice on the same triplet of images, once using all the points in the scene and once using only points on a plane. The results are shown in Figure 3. White circles represent the tracked points.

Black crosses represent the reprojected points. In the first case the scene has a large volume and our method has no problem reprojecting the points with an average error of 0.5 pixels. In the second case we manually deleted all the points outside a specific plane and ran the algorithm again. The reprojection now was 0.2 pixels.

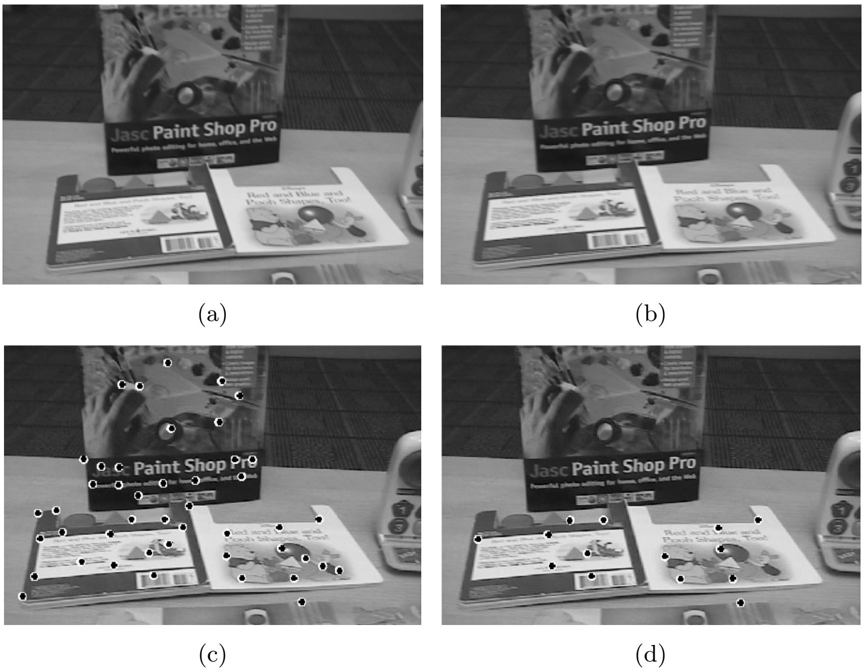


Fig. 3. Original images ((a),(b)) are reprojected to the third image using all the points in the scene (c) or only the points on the plane (d) (Images (c) and (d) show the third image with the different point configurations). The reprojection error is 0.5 pixels for for image (c) and 0.2 pixels for image (d).

6 Summary

We have shown, in this paper, that the ambiguity in recovering multi-linear constraints from planar scenes is *orthogonal* to tasks such as reprojection. Thus, it is not necessary to choose a different model for different scenes (Homography for 2D scenes or trifocal tensor/fundamental matrix for 3D scenes) as the ambiguity in the recovered parameters does not affect our ability to perform reprojection. Moreover, in the case of a "thin" volume which is not 2D nor 3D, our method will generate a tensor that is provably correct for reprojecting all the points within this volume. We thus have a unified method for reprojecting planar, "thin" and full volume scenes. Finally, while the results we have shown are relevant to the process of reprojection, we believe that they can be used in some reconstruction situations as well, but leave it for future research.

A Appendix

Consider the expression

$$| xb^\top - ay^\top |_{L_2} + | xc^\top - dy^\top |_{L_2} + | xe^\top - fy^\top |_{L_2}$$

where a, b, c, d, e, f are vectors, and $||_{L_2}$ stands for the L_2 norm of a matrix defined by the sum of squares of the matrix entries. The vectors x, y that bring the expression to minimum are described by $x = \alpha_1 a + \alpha_2 d + \alpha_3 f$ and $y = \beta_1 b + \beta_2 c + \beta_3 e$ for some coefficients $\alpha_i, \beta_i,\ i = 1, 2, 3$. The derivation is as follows.

Since $| A |_{L_2} = trace(A^\top A) = trace(AA^\top)$, then the trace of the expression above is

$$\begin{aligned}&(b^\top b + c^\top c + e^\top e)(x^\top x) - 2(b^\top y)(x^\top a) - 2(c^\top y)(x^\top d)\\ &\quad -2(e^\top y)(x^\top f) + (a^\top a + d^\top d + f^\top f)(y^\top y) \qquad .\end{aligned}$$

The partial derivatives with respect to x and y are therefore

$$\begin{aligned}\frac{\partial}{\partial x} &= (b^\top b + c^\top c + e^\top e)x - (b^\top y)a - (c^\top y)d - (e^\top y)f = 0\\ \frac{\partial}{\partial y} &= (a^\top a + d^\top d + f^\top f)y - (x^\top a)b - (x^\top d)c - (x^\top f)e = 0.\end{aligned}$$

References

1. S. Birchfield An Implementation of the Kanade-Lucas-Tomasi Feature Tracker. Download from: http://vision.stanford.edu/ birch/klt
2. R. Courant and D. Hilbert. *Methods of mathematical physics.* Interscience Publishers Inc., 1953.
3. 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.
4. R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.
5. 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.
6. 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.
7. 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.

8. P. Pritchett and A. Zisserman. Matching and Reconstruction from Widely Separated Views. In *3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506*, Springer-Verlag, 1998.
9. A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.
10. A. Shashua. Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. In G. Sommer and J.J. Koenderink, editors, *Algebraic Frames For The Perception Action Cycle*, number 1315 in Lecture Notes in Computer Science. Springer, 1997. Proceedings of the workshop held in Kiel, Germany, Sep. 1997.
11. 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.
12. M.E. Spetsakis and J. Aloimonos. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4(3):171–183, 1990.
13. 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.
14. P.H.S Torr and A. Zisserman Concerning bayesian motion segmentation, model averaging, matching and the trifocal tensor. In *Proceedings of the European Conference on Computer Vision*, 1998.
15. P.H.S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–607, 1997.
16. B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995.