Plane + Parallax, Tensors and Factorization

Bill Triggs

INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot, France
Bill.Triggs@inrialpes.fr — http://www.inrialpes.fr/movi/people/ Triggs

Abstract. We study the special form that the general multi-image tensor formalism takes under the plane + parallax decomposition, including matching tensors and constraints, closure and depth recovery relations, and inter-tensor consistency constraints. Plane + parallax alignment greatly simplifies the algebra, and uncovers the underlying geometric content. We relate plane + parallax to the geometry of translating, calibrated cameras, and introduce a new parallax-factorizing projective reconstruction method based on this. Initial plane + parallax alignment reduces the problem to a single rank-one factorization of a matrix of rescaled parallaxes into a vector of projection centres and a vector of projective heights above the reference plane. The method extends to 3D lines represented by via-points and 3D planes represented by homographies.

Keywords: Plane + parallax, matching tensors, projective reconstruction, factorization, structure from motion.

1 Introduction

This paper studies the special forms that matching tensors take under the plane + parallax decomposition, and uses this to develop a new projective reconstruction method based on rank-1 parallax factorization. The main advantage of the plane + parallax analysis is that it greatly simplifies the usually rather opaque matching tensor algebra, and clarifies the way in which the tensors encode the underlying 3D camera geometry. The new plane + parallax factorizing reconstruction method appears to be even stabler than standard projective factorization, especially for near-planar scenes. It is a one-step, closed form, multi-point, multi-image factorization for projective structure, and in this sense improves on existing minimal-configuration and iterative depth recovery plane + parallax SFM methods [19, 4, 23, 22, 45]. As with standard projective factorization [37], it can be extended to handle 3D lines (via points) and planes (homographies) alongside 3D points.

Matching tensors [29, 8, 36] are the image signature of the camera geometry. Given several perspective images of the same scene taken from different viewpoints, the 3D camera geometry is encoded by a set of 3×4 homogeneous camera projection matrices. These depend on the chosen 3D coordinate system, but the dependence can be eliminated algebraically to give four series of multi-image **tensors** (multi-index arrays of components), each interconnecting 2–4 images. The different images of a 3D feature are

To appear in ECCV'2000. This work was supported by Esprit LTR project CUMULI. I would like to thank A. Zisserman and P. Anandan for comments. [03 April 2000 final]

constrained by multilinear **matching relations** with the tensors as coefficients. These relations can be used to estimate the tensors from an initial set of correspondences, and the tensors then constrain the search for further correspondences. The tensors implicitly characterize the relative projective camera geometry, so they are a useful starting point for 3D reconstruction. Unfortunately, they are highly redundant, obeying a series of complicated internal self-consistency constraints whose general form is known but too complex to use easily, except in the simplest cases [36, 5, 17, 6].

On the other hand, a camera is simply a device for recording incoming light in various directions at the camera's optical centre. Any two cameras with the same centre are equivalent in the sense that — modulo field-of-view and resolution constraints which we ignore for now — they see exactly the same set of incoming light rays. So their images can be warped into one another by a 1-1 mapping (for projective cameras, a 2D homography). Anything that can be done using one of the images can equally well be done using the other, if necessary by pre-warping to make them identical.

From this point of view, it is clear that the camera centres are the essence of the *3D* camera geometry. Changing the camera orientations or calibrations while leaving the centres fixed amounts to a 'trivial' change of image coordinates, which can be undone at any time by homographic (un)warping. In particular, the algebraic structure (degeneracy, number of solutions, *etc*.) of the matching constraints, tensors and consistency relations — and *a fortiori* that of any visual reconstruction based on these — is essentially a 3D matter, and hence depends *only* on the camera centres.

It follows that much of the complexity of the matching relations is only apparent. At bottom, the geometry is simply that of a configuration of 3D points (the camera centres). But the inclusion of arbitrary calibration-orientation homographies everywhere in the formulae makes the algebra appear much more complicated than need be. One of the main motivations for this work was to study the matching tensors and relations in a case — that of projective plane + parallax alignment — where most of the arbitrariness due to the homographies has been removed, so that the underlying geometry shows up much more clearly.

The observation that the camera centres lie at the heart of the projective camera geometry is by no means new. It is the basis of Carlsson's 'duality' between 3D points and cameras (*i.e.* centres) [2, 43, 3, 10], and of Heyden & Åström's closely related 'reduced tensor' approach [13–15, 17]. The growing geometry tradition in the plane + parallax literature [19, 23, 22, 4, 45] is also particularly relevant here.

Organization: §2 introduces our plane + parallax representation and shows how it applies to the basic feature types; §3 displays the matching tensors and constraints in the plane + parallax representation; §4 discusses tensor scaling, redundancy and consistency; §5 considers the tensor closure and depth recovery relations under plane + parallax ; §6 introduces the new parallax factorizing projective reconstruction method; §7 shows some initial experimental results; and §8 concludes.

Notation: Bold italic '$\boldsymbol{x}$' denotes 3-vectors, bold sans-serif 'x' 4-vectors, upper case '$\boldsymbol{H}, \mathsf{H}$' matrices, Greek '$\lambda, \mu$' scalars (*e.g.* homogeneous scale factors). We use homogeneous coordinates for 3D points x and image points $\boldsymbol{x}$, but usually inhomogeneous ones $\boldsymbol{c}$ for projection centres $\mathsf{c} = \begin{pmatrix} \boldsymbol{c} \\ 1 \end{pmatrix}$. We use P for 3×4 camera projection matrices, $\boldsymbol{e}$ for epipoles. 3D points $\mathsf{x} = \begin{pmatrix} \boldsymbol{x} \\ w \end{pmatrix}$ are parametrized by a point $\boldsymbol{x}$ on the reference plane and a ‘projective height’ w above it. $\wedge$ denotes cross-product, $[\,-\,]_\times$ the associated 3×3 skew matrix $[\,\boldsymbol{x}\,]_\times\,\boldsymbol{y} = \boldsymbol{x}\wedge\boldsymbol{y}$, and $[\,\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\,]$ the triple product.

2 The Plane + Parallax Representation

Data analysis is often simplified by working in terms of small corrections against a reference model. Image analysis is no exception. In **plane + parallax**, the reference model is a real or virtual **reference plane** whose points are held fixed throughout the image sequence by image warping (see, *e.g.* [24, 33, 19] and their references). The reference is often a perceptually dominant plane in the scene such as the ground plane. Points that lie above the plane are not exactly fixed, but their motion can be expressed as a residual **parallax** with respect to the plane. The parallax is often much smaller than the uncorrected image motion, particularly when the camera motion is mainly rotational. This simplifies feature extraction and matching. For each projection centre, alignment implicitly defines a unique reference orientation and calibration, and in this sense entirely cancels any orientation and calibration variations. Moreover, the residual parallaxes directly encode useful structural information about the size of the camera translation and the distance of the point above the plane. So alignment can be viewed as a way of focusing on the essential *3D* geometry — the camera centres and 3D points — by eliminating the ‘nuisance variables’ associated with orientation and calibration. The ‘purity’ of the parallax signal greatly simplifies many geometric computations. In particular, we will see that it dramatically simplifies the otherwise rather cumbersome algebra of the matching tensors and relations (*c.f.* also [19, 22, 4]).

The rest of this section describes our “plane at infinity + parallax” representation. It is projectively equivalent to the more common “ground plane + parallax” representation (*e.g.* [19, 42]), but has algebraic advantages — simpler formulae for scale factors, and the link to translating cameras — that will be discussed below.

Coordinate frame: We suppose given a 3D **reference plane** with a predefined projective coordinate system, and a 3D **reference point** not on the plane. The plane may be real or virtual, explicit or implicit. The plane coordinates might derive from an image or be defined by features on the plane. The reference point might be a 3D point, a projection centre, or arbitrary. We adopt a projective 3D coordinate system that places the reference point at the 3D origin $(0\;0\;0\;1)^\top$, and the reference plane at infinity in standard position (*i.e.* its reference coordinates coincide with the usual coordinates on the plane at infinity). Examining the possible residual 4×4 homographies shows that this fixes the 3D projective frame up to a single global scale factor. If $\mathsf{H} = \left(\begin{smallmatrix} \boldsymbol{A} & \boldsymbol{t} \\ \boldsymbol{b}^\top & \lambda \end{smallmatrix}\right)$, then the constraint that H fixes each point $\left(\begin{smallmatrix} \boldsymbol{x} \\ 0 \end{smallmatrix}\right)$ on the reference plane implies that $\boldsymbol{A} = \mu\,\boldsymbol{I}$ and $\boldsymbol{b} = \boldsymbol{0}$, and the constraint that H fixes the origin $\left(\begin{smallmatrix} \boldsymbol{0} \\ 1 \end{smallmatrix}\right)$ implies that $\boldsymbol{t} = \boldsymbol{0}$. So $\mathsf{H} = \left(\begin{smallmatrix} \mu\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda \end{smallmatrix}\right)$, which is a global scaling by μ/λ.

3D points: 3D points are represented as linear combinations of the reference point/origin and a point on the reference plane:

$$\mathsf{x} \;\equiv\; \begin{pmatrix} \boldsymbol{x} \\ w \end{pmatrix} \;=\; \begin{pmatrix} \boldsymbol{x} \\ 0 \end{pmatrix} + w \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} \qquad (1)$$

$\boldsymbol{x}$ is the intersection with the reference plane, of the line through X and the origin. w is called X's **projective height** above the plane. $w = 0$ is the reference plane, $w = \infty$ the origin. w depends on the normalization convention for $\boldsymbol{x}$. If the reference plane is made finite ($z = 0$) by interchanging z and w coordinates, w becomes the vertical height above the plane. But in our projective, plane-at-infinity based frame with affine normalization $\boldsymbol{x}_z = 1$, w is the inverse z-distance (or with spherical normalization $\|\boldsymbol{x}\| = 1$, the inverse "Euclidean" distance) of X from the origin.

Camera matrices: Plane + parallax aligned cameras fix the image of the reference plane, so their leading 3×3 submatrix is the identity. They are parametrized simply by their projection centres:

$$\mathsf{P} \;=\; \begin{pmatrix} u\,\boldsymbol{I}_{3\times3} & -\boldsymbol{c} \end{pmatrix} \qquad \text{with projection centre} \qquad \mathsf{c} \;=\; \begin{pmatrix} \boldsymbol{c} \\ u \end{pmatrix} \tag{2}$$

Hence, any 3D point can be viewed as a plane + parallax aligned camera and vice versa. But, whereas points often lie on or near the reference plane ($w \to 0$), cameras centred on the plane ($u \to 0$) are too singular to be useful — they project the entire 3D scene to their centre point $\boldsymbol{c}$.

We will break the nominal symmetry between points and cameras. Points will be treated projectively, as general homogeneous 4-component quantities with arbitrary height component w. But camera centres $\mathsf{c} = \left(\begin{smallmatrix} \boldsymbol{c} \\ u \end{smallmatrix}\right)$ will be assumed to lie outside the reference plane and scaled affinely ($u \to 1$), so that they and their camera matrices $\mathsf{P} = \begin{pmatrix} u\,\boldsymbol{I} & -\boldsymbol{c} \end{pmatrix}$ are parametrized by their *in*homogeneous centre 3-vector $\boldsymbol{c}$ alone.

This asymmetry is critical to our approach. Our coordinate frame and reconstruction methods are essentially projective and are most naturally expressed in homogeneous coordinates. Conversely, scaling u to 1 freezes the scales of the projection matrices, and everywhere that matching tensors are used, it converts formulae that would be bilinear or worse in the $\boldsymbol{c}$'s and u's, to ones that are merely *linear* in the $\boldsymbol{c}$'s. This greatly simplifies the tensor estimation process compared to the general unaligned case. The representation becomes singular for cameras near the reference plane, but that is not too much of a restriction in practice. In any case it *had* to happen — no minimal linear representation can be globally valid, as the general redundant tensor one is.

Point projection: In image i, the image $\mathsf{P}_i\,\mathsf{X}_p$ of a 3D point $\mathsf{X}_p = \left(\begin{smallmatrix} \boldsymbol{x}_p \\ w_p \end{smallmatrix}\right)$ is displaced linearly from its reference image[1] $\boldsymbol{x}_p$ towards the centre of projection $\boldsymbol{c}_i$, in proportion to its height w_p :

$$\lambda_{ip}\,\boldsymbol{x}_{ip} \;=\; \mathsf{P}_i\,\mathsf{X}_p \;=\; \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{c}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_p \\ w_p \end{pmatrix} \;=\; \boldsymbol{x}_p - w_p\,\boldsymbol{c}_i \tag{3}$$

Here λ_{ip} is a **projective depth** [32, 37] — an initially-unknown projective scale factor that compensates for the loss of the scale information in $\mathsf{P}_i\,\mathsf{X}_p$ when $\boldsymbol{x}_{ip}$ is measured in its image. Although the homogeneous rescaling freedom of X_p makes them individually arbitrary, the combined projective depths of a 3D point — or more precisely its vector

[1] The origin/reference point need not coincide with a physical camera, but can still be viewed as a **reference camera** $\mathsf{P}_0 = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \end{pmatrix}$, projecting 3D points X_p to their reference images $\boldsymbol{x}_p$.

of **rescaled image points** $(\lambda_{ip}\, \boldsymbol{x}_{ip})_{i=1\ldots m}$ — implicitly define its 3D structure. This is similar to the general projective case, except that in plane + parallax the projection matrix scale freedom is already frozen: while the homogeneous scale factors of x_p and $\boldsymbol{x}_{ip}$ are arbitrary, $\boldsymbol{c}_i$ has a fixed scale linked to its camera's position.

Why not a ground plane?: In many applications, the reference plane is nearby. Pushing it out to infinity forces a deeply projective 3D frame. It might seem preferable to use a finite reference plane, as in, *e.g.* [19, 42]. For example, interchanging the z and w coordinates puts the plane at $z = 0$, the origin at the vertical infinity $(0\ 0\ 1\ 0)^\top$, and (modulo Euclidean coordinates on the plane itself) creates an obviously rectilinear 3D coordinate system, where $\boldsymbol{x}$ gives the ground coordinates and w the vertical height above the plane. However, a finite reference plane would hide a valuable insight that is obvious from (2): *The plane + parallax aligned camera geometry is projectively equivalent to translating calibrated cameras*. Any algorithm that works for these works for projective plane + parallax, and (up to a 3D projectivity!) vice versa. Although not new (see, *e.g.* [14]), this analogy deserves to be better known. It provides simple algebra and geometric intuition that were very helpful during this work. It explicitly realizes — albeit in a weak, projectively distorted sense, with the reference plane mapped to infinity — the suggestion that plane + parallax alignment cancels the orientation and calibration, leaving only the translation [22].

3D Lines: Any 3D line L can be parametrized by a homogeneous 6-tuple of Plücker coordinates $(\boldsymbol{l}, \boldsymbol{z})$ where: (*i*) $\boldsymbol{l}$ is a line 3-vector — L's projection from the origin onto the reference plane; (*ii*) $\boldsymbol{z}$ is a point 3-vector — L's intersection with the reference plane; (*iii*) $\boldsymbol{z}$ lies on $\boldsymbol{l}$, $\boldsymbol{l}\cdot\boldsymbol{z} = 0$ (this is the Plücker constraint); (*iv*) the relative scaling of $\boldsymbol{l}$ and $\boldsymbol{z}$ is fixed and gives L's 'steepness': lines on the plane have $\boldsymbol{z} \to \boldsymbol{0}$, while the ray from the origin to $\boldsymbol{z}$ has $\boldsymbol{l} \to \boldsymbol{0}$. This parametrization of L relates to the usual 3D projective Plücker (4×4 skew rank 2 matrix) representations as follows:

$$\mathsf{L}^* = \begin{pmatrix} [\boldsymbol{l}]_\times & \boldsymbol{z}^\top \\ -\boldsymbol{z} & 0 \end{pmatrix} \quad \text{contravariant form} \qquad \mathsf{L}_* = \begin{pmatrix} [\boldsymbol{z}]_\times & \boldsymbol{l}^\top \\ -\boldsymbol{l} & 0 \end{pmatrix} \quad \text{covariant form} \tag{4}$$

The line from $\binom{\boldsymbol{x}}{w}$ to $\binom{\boldsymbol{y}}{v}$ is $(\boldsymbol{l}, \boldsymbol{z}) = (\boldsymbol{x} \wedge \boldsymbol{y}, w\,\boldsymbol{y} - v\,\boldsymbol{x})$. A 3D point $\mathsf{x} = \binom{\boldsymbol{x}}{w}$ lies on L iff $\mathsf{L}_*\,\mathsf{x} = \binom{w\,\boldsymbol{l} + \boldsymbol{z} \wedge \boldsymbol{x}}{\boldsymbol{l}\cdot\boldsymbol{x}} = \mathsf{0}$. In a camera at c_i, L projects to:

$$\mu_i\,\boldsymbol{l}_i = \boldsymbol{l} + \boldsymbol{z} \wedge \boldsymbol{c}_i \tag{5}$$

This vanishes if c_i lies on L.

Displacements and epipoles: Given two cameras with centres $\mathsf{c}_i = \binom{\boldsymbol{c}_i}{1}$ and $\mathsf{c}_j = \binom{\boldsymbol{c}_j}{1}$, the **3D displacement vector** between their two centres is $\boldsymbol{c}_{ij} = \boldsymbol{c}_i - \boldsymbol{c}_j$. The scale of $\boldsymbol{c}_{ij}$ is meaningful, encoding the relative 3D camera position. Forgetting this scale factor gives the **epipole** $\boldsymbol{e}_{ij}$ — the 2D projective point at which the ray from c_j to c_i crosses the reference plane:

$$\boldsymbol{e}_{ij} \simeq \boldsymbol{c}_{ij} \equiv \boldsymbol{c}_i - \boldsymbol{c}_j \tag{6}$$

We will see below that it is really the inter-camera displacements $\boldsymbol{c}_{ij}$ and not the epipoles $\boldsymbol{e}_{ij}$ that appear in tensor formulae. Correct relative scalings are essential for geometric coherence, but precisely because of this they are also straightforward to estimate. Once found, the displacements $\boldsymbol{c}_{ij}$ amount to a reconstruction of the plane + parallax aligned camera geometry. To find a corresponding set of camera centres, simply fix the 3D coordinates of one centre (or function of the centres) arbitrarily, and the rest follow immediately by adding displacement vectors.

Parallax: Subtracting two point or line projection equations (3, 5) gives the following important **parallax equations**:

$$\lambda_i\,\boldsymbol{x}_i - \lambda_j\,\boldsymbol{x}_j \;=\; -w\,\boldsymbol{c}_{ij} \tag{7}$$

$$\mu_i\,\boldsymbol{l}_i - \mu_j\,\boldsymbol{l}_j \;=\; \boldsymbol{z}\wedge\boldsymbol{c}_{ij} \tag{8}$$

Given the correct projective depths λ, μ, the relative parallax caused by a camera displacement is proportional to the displacement vector. The RHS of (7) already suggests the possibility of factoring a multi-image, multi-point matrix of rescaled parallaxes into (w) and $(\boldsymbol{c}_{ij})$ matrices. Results equivalent to (7) appear in [19, 22], albeit with more complicated scale factors owing to the use of different projective frames.

Equation (7) has a trivial interpretation in terms of 3D displacements. For a point $\mathsf{x} = \left(\begin{smallmatrix}\boldsymbol{x}\\ w\end{smallmatrix}\right)$ above the reference plane, scaling to $w = 1$ gives projection equations $\lambda_i\,\boldsymbol{x}_i \;=\; \mathsf{P}_i\,\mathsf{x} \;=\; \boldsymbol{x} - \boldsymbol{c}_i$, so $\lambda_i\,\boldsymbol{x}_i$ is the 3D displacement vector from c_i to x. (7) just says that the sum of displacements around the 3D triangle $\overline{\mathsf{c}_i\,\mathsf{c}_j\,\mathsf{x}}$ vanishes. On the reference plane, this entails the alignment of the 2D points $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ and $\boldsymbol{e}_{ij}$ (along the line of intersection of the 3D plane of $\overline{\mathsf{c}_i\,\mathsf{c}_j\,\mathsf{x}}$ with the reference plane — see fig. 1), and hence the vanishing of the triple product $[\,\boldsymbol{x}_i, \boldsymbol{e}_{ij}, \boldsymbol{x}_j\,] = 0$. However the 3D information in the relative scale factors is more explicit in (7).

3D Planes: The 3D plane $\mathsf{p} \;=\; (\boldsymbol{n}^\top\;d)$ has equation $\mathsf{p}\cdot\mathsf{x} \;=\; \boldsymbol{n}\cdot\boldsymbol{x} + d\,w = 0$. It intersects the reference plane in the line $\boldsymbol{n}\cdot\boldsymbol{x} \;=\; 0$. The relative scaling of $\boldsymbol{n}$ and d gives the 'steepness' of the 3D plane: $\boldsymbol{n} = \boldsymbol{0}$ for the reference plane, $d = 0$ for planes through the origin. A point $\boldsymbol{x}_j$ in image j back-projects to the 3D point $\mathsf{B}_j\,\boldsymbol{x}_j$ on p, which induces an image j to image i homography $\boldsymbol{H}_{ij}$, where:

$$\mathsf{B}_j(\mathsf{p}) \;\equiv\; \begin{pmatrix} \boldsymbol{I} - \boldsymbol{c}_j\,\boldsymbol{n}^\top/(\boldsymbol{n}\cdot\boldsymbol{c}_j + d) \\ -\boldsymbol{n}^\top/(\boldsymbol{n}\cdot\boldsymbol{c}_j + d) \end{pmatrix} \qquad \boldsymbol{H}_{ij}(\mathsf{p}) \;=\; \mathsf{P}_i\,\mathsf{B}_j \;=\; \boldsymbol{I} + \frac{\boldsymbol{c}_{ij}\,\boldsymbol{n}^\top}{\boldsymbol{n}\cdot\boldsymbol{c}_j + d} \tag{9}$$

For any i, j and any p, this fixes the epipole $\boldsymbol{e}_{ij}$ and each point on the intersection line $\boldsymbol{n}\cdot\boldsymbol{x} \;=\; 0$. $\boldsymbol{H}_{ij}$ is actually a **planar homology** [30, 11] — it has a double eigenvalue corresponding to the points on the fixed line.

Any chosen plane $\mathsf{p} \;=\; (\boldsymbol{n}^\top\;d)$ can be made the reference plane by applying a 3D homography H and compensating image homographies $\boldsymbol{H}_i$:

$$\mathsf{H} \;=\; \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{n}^\top/d & 1 \end{pmatrix} \;=\; \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{n}^\top/d & 1 \end{pmatrix}^{-1} \qquad \boldsymbol{H}_i \;=\; \boldsymbol{I} - \frac{\boldsymbol{c}_i\,\boldsymbol{n}^\top}{\boldsymbol{n}\cdot\boldsymbol{c}_i + d} \;=\; \left(\boldsymbol{I} + \frac{\boldsymbol{c}_i\,\boldsymbol{n}^\top}{d}\right)^{-1}$$

Reference positions $\boldsymbol{x}$ are unchanged, projective heights are warped by an affinity $w \rightarrow w + \boldsymbol{n} \cdot \boldsymbol{x}/d$, and camera centres are rescaled $\boldsymbol{c} \rightarrow \frac{d}{\boldsymbol{n} \cdot \boldsymbol{c} + d}\,\boldsymbol{c}$ (infinitely, if they lie on the plane p):

$$\mathsf{x} = \begin{pmatrix} \boldsymbol{x} \\ w \end{pmatrix} \longrightarrow \mathsf{H}\,\mathsf{x} = \begin{pmatrix} \boldsymbol{x} \\ w + \boldsymbol{n} \cdot \boldsymbol{x}/d \end{pmatrix} \tag{10}$$

$$\mathsf{p} = \begin{pmatrix} \boldsymbol{n}^\top & d \end{pmatrix} \longrightarrow \mathsf{p}\,\mathsf{H}^{-1} = \begin{pmatrix} \boldsymbol{0} & d \end{pmatrix} \tag{11}$$

$$\mathsf{P}_i = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{c}_i \end{pmatrix} \longrightarrow \boldsymbol{H}_i\,\mathsf{P}_i\,\mathsf{H}^{-1} = \begin{pmatrix} \boldsymbol{I} & \frac{-d\,\boldsymbol{c}_i}{\boldsymbol{n} \cdot \boldsymbol{c}_i + d} \end{pmatrix} \tag{12}$$

When p is the true plane at infinity, the 3D frame becomes affine and the aligned camera motion becomes truly translational.

Given multiple planes p_k and images i, and choosing some fixed base image 0, the 3 columns of each $\boldsymbol{H}_{i0}(\mathsf{p}_k)$ can be viewed as three point vectors and incorporated into the rank-one factorization method below to reconstruct the $\boldsymbol{c}_{i0}$ and $\boldsymbol{n}_k^\top / (\boldsymbol{n}_k \cdot \boldsymbol{c}_0 + d_k)$. Consistent normalizations for the different $\boldsymbol{H}_{i0}$ are required. If $\boldsymbol{e}_{i0}$ is known, the correct normalization can be recovered from $[\,\boldsymbol{e}_{i0}\,]_\times\, \boldsymbol{H}_{i0} = [\,\boldsymbol{e}_{i0}\,]_\times$. This amounts to the point depth recovery equation (19) below applied to the columns of $\boldsymbol{H}_{i0}$ and $\boldsymbol{H}_{00} = \boldsymbol{I}$. Alternatively, $\boldsymbol{H}_{i0} = \boldsymbol{I} + \ldots$ has two repeated unit eigenvalues, and the right (left) eigenvectors of the remaining eigenvalue are $\boldsymbol{e}_{i0}$ ($\boldsymbol{n}^\top$). This allows the normalization, epipole and plane normal to be recovered from an estimated $\boldsymbol{H}_{i0}$. Less compact rank 4 factorization methods also exist, based on writing $\boldsymbol{H}_{i0}$ as a 9-vector, linear in the components of $\boldsymbol{I}$ and either $\boldsymbol{c}_{i0}$ or $\boldsymbol{n}_k$ [28, 44, 45].

Carlsson duality: Above we gave the plane + parallax correspondence between 3D points and (the projection centres of aligned) cameras [19, 22]:

$$\mathsf{x} = \begin{pmatrix} \boldsymbol{x} \\ w \end{pmatrix} \quad \leftrightarrows \quad \mathsf{P} = \begin{pmatrix} w\boldsymbol{I} & -\boldsymbol{x} \end{pmatrix}$$

Carlsson [2, 3] (see also [13, 14, 43, 10, 42]) defined a related but more 'twisted' duality mapping based on the alignment of a projective basis rather than a plane:

$$\mathsf{x} = \begin{pmatrix} \boldsymbol{x} \\ w \end{pmatrix} \quad \leftrightarrows \quad \mathsf{P} = \begin{pmatrix} 1/x & & & -1/w \\ & 1/y & & -1/w \\ & & 1/z & -1/w \end{pmatrix} \simeq \begin{pmatrix} x & & \\ & y & \\ & & z \end{pmatrix}^{-1} \begin{pmatrix} w\boldsymbol{I} & -\boldsymbol{x} \end{pmatrix}$$

Provided that x, y, z are non-zero, the two mappings differ only by an image homography. Plane + parallax aligns a 3D plane pointwise, thus forcing the image $-\boldsymbol{x}$ of the origin to depend on the projection centre. Carlsson aligns a 3D projective basis, fixing the image of the origin and just 3 points on the plane (and incidentally introducing potentially troublesome singularities for projection centres on the x, y and z coordinate planes, as well as on the $w = 0$ one). In either case the point-camera "duality" (isomorphism would be a better description) allows some or all points to be treated as cameras and vice versa. This has been a fruitful approach for generating new algorithms [2, 43, 3, 10, 42, 19, 22, 4]. All of the below formulae can be dualized, with the proviso that camera centres should avoid the reference plane and be affinely normalized, while points need not and must be treated homogeneously.

3 Matching Tensors and Constraints in Plane + Parallax

Matching Tensors: The matching tensors for aligned projections are very simple functions of the scaled epipoles / projection centre displacements. From a tensorial point of view[2], the simplest way to derive them is to take the homography-epipole decompositions of the generic matching tensors [29, 8, 36], and substitute identity matrices for the homographies:

$$\begin{aligned}
\boldsymbol{c}_{12} &= \boldsymbol{c}_1 - \boldsymbol{c}_2 && \text{displacement from } \mathsf{c}_2 \text{ to } \mathsf{c}_1 \\
\boldsymbol{F}_{12} &= [\,\boldsymbol{c}_{12}\,]_\times \;=\; [\,\boldsymbol{c}_1 - \boldsymbol{c}_2\,]_\times && \text{image 1-2 fundamental matrix} \\
\boldsymbol{T}_1^{23} &= \boldsymbol{I}_1^2 \otimes \boldsymbol{c}_{13} - \boldsymbol{c}_{12} \otimes \boldsymbol{I}_1^3 && \text{image 1-2-3 trifocal tensor} \\
\boldsymbol{Q}^{A_1A_2A_3A_4} &= \textstyle\sum_{i=1}^3 (-1)^{i-1} \boldsymbol{\epsilon}^{A_1\ldots\hat{A}_i\ldots A_4} \cdot \boldsymbol{c}_{i4}^{A_i} && \text{image 1-2-3-4 quadrifocal tensor}
\end{aligned}$$

The plane + parallax fundamental matrix and trifocal tensor have also been studied in [22, 4]. The use of affine scaling $u_i \to 1$ for the centres $\mathsf{c}_i = \binom{\boldsymbol{c}_i}{u_i}$ is essential here, otherwise $\boldsymbol{T}$ is bilinear and $\boldsymbol{Q}$ quadrilinear in $\boldsymbol{c}, u$.

Modulo scaling, $\boldsymbol{c}_{12}$ is the epipole $\boldsymbol{e}_{12}$ — the intersection of the ray from c_2 to c_1 with the reference plane. Coherent relative scaling of the terms of the trifocal and quadrifocal tensor sums is indispensable here, as in most other multi-term tensor relations. But for this very reason, the correct scales can be found using these relations. As discussed above, the correctly scaled $\boldsymbol{c}_{ij}$'s characterize the relative 3D camera geometry very explicitly, as a network of 3D displacement vectors. It is actually rather misleading to think in terms of epipolar points on the reference plane: the $\boldsymbol{c}_{ij}$ are neither estimated (*e.g.* from the trifocal tensor) nor used (*e.g.* for reconstruction) like that, and treating their scale factors as arbitrary only confuses the issue.

Matching constraints: The first few matching relations simplify as follows:

$$[\,\boldsymbol{x}_1, \boldsymbol{c}_{12}, \boldsymbol{x}_2\,] \;=\; 0 \qquad \text{epipolar point} \tag{13}$$

$$(\boldsymbol{x}_1 \wedge \boldsymbol{x}_2)\,(\boldsymbol{c}_{13} \wedge \boldsymbol{x}_3)^\top - (\boldsymbol{c}_{12} \wedge \boldsymbol{x}_2)\,(\boldsymbol{x}_1 \wedge \boldsymbol{x}_3)^\top \;=\; \boldsymbol{0} \qquad \text{trifocal point} \tag{14}$$

$$(\boldsymbol{l}_1 \wedge \boldsymbol{l}_2)\,(\boldsymbol{l}_3 \cdot \boldsymbol{c}_{13}) - (\boldsymbol{l}_2 \cdot \boldsymbol{c}_{12})\,(\boldsymbol{l}_1 \wedge \boldsymbol{l}_3) \;=\; \boldsymbol{0} \qquad \text{trifocal line} \tag{15}$$

$$(\boldsymbol{l}_2 \cdot \boldsymbol{x}_1)\,(\boldsymbol{l}_3 \cdot \boldsymbol{c}_{13}) - (\boldsymbol{l}_2 \cdot \boldsymbol{c}_{12})\,(\boldsymbol{l}_3 \cdot \boldsymbol{x}_1) \;=\; 0 \qquad \text{trifocal point-line} \tag{16}$$

$$\begin{aligned}
(\boldsymbol{l}_2 \wedge \boldsymbol{l}_3)\,(\boldsymbol{l}_1 \cdot \boldsymbol{c}_{14}) + (\boldsymbol{l}_3 \wedge \boldsymbol{l}_1)\,(\boldsymbol{l}_2 \cdot \boldsymbol{c}_{24}) & \\
+ (\boldsymbol{l}_1 \wedge \boldsymbol{l}_2)\,(\boldsymbol{l}_3 \cdot \boldsymbol{c}_{34}) &\;=\; \boldsymbol{0} \qquad \text{quadrifocal 3-line}
\end{aligned} \tag{17}$$

Equation (16) is the primitive trifocal constraint. Given three images $\boldsymbol{x}_{i|i=1\ldots3}$ of a 3D point x, and arbitrary image lines $\boldsymbol{l}_2, \boldsymbol{l}_3$ through $\boldsymbol{x}_2, \boldsymbol{x}_3$, (16) asserts that the 3D optical ray of $\boldsymbol{x}_1$ meets the 3D optical planes of $\boldsymbol{l}_2, \boldsymbol{l}_3$ in a common 3D point (x). The tri- and quadrifocal 3-line constraints (16,17) both require that the optical planes of $\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3$

[2] There is no space here to display the general projective tensor analogues of the plane + parallax expressions given here and below — see [35].

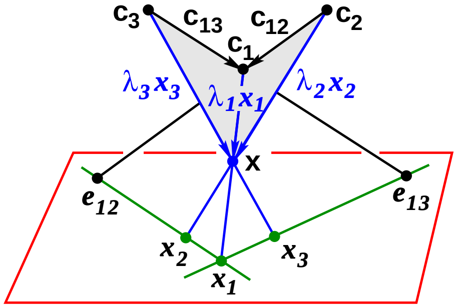

Fig. 1. The geometry of the trifocal constraint.

intersect in a common 3D line. The quadrifocal 4-point constraint is straightforward but too long to give here.

The trifocal point constraint contains [29, 35, 22, 4] two epipolar constraints in the form $\boldsymbol{x} \wedge \boldsymbol{x}' \simeq \boldsymbol{c} \wedge \boldsymbol{x}'$, plus a proportionality-of-scale relation for these parallel 3-vectors:

$$(\boldsymbol{x}_1 \wedge \boldsymbol{x}_2) \;:\; (\boldsymbol{c}_{12} \wedge \boldsymbol{x}_2) \;=\; (\boldsymbol{x}_1 \wedge \boldsymbol{x}_3) \;:\; (\boldsymbol{c}_{13} \wedge \boldsymbol{x}_3) \tag{18}$$

The homogeneous scale factors of the $\boldsymbol{x}$'s cancel. This equation essentially says that $\boldsymbol{x}_3$ must progress from $\boldsymbol{e}_{13}$ to $\boldsymbol{x}_1$ in step with $\boldsymbol{x}_2$ as it progresses from $\boldsymbol{e}_{12}$ to $\boldsymbol{x}_1$ (and both in step with x as it progresses from c_1 to $\boldsymbol{x}_1$ on the plane — see fig. 1). In terms of 3D displacement vectors $\boldsymbol{c}$ and $\lambda\,\boldsymbol{x}$ (or if the figure is projected generically into another image), the ratio on the LHS of (18) is 1, being the ratio of two different methods of calculating the area of the triangle $\overline{\mathsf{c}_1\,\mathsf{c}_2\,\mathsf{x}}$. Similarly for the RHS with $\overline{\mathsf{c}_1\,\mathsf{c}_3\,\mathsf{x}}$. Both sides involve x, hence the lock-step.

Replacing the lines in the line constraints (15,16,17) with corresponding tangents to iso-intensity contours gives ***tensor brightness constraints*** on the normal flow at a point. The Hanna-Okamoto-Stein-Shashua brightness constraint (16) predominates for small, mostly-translational image displacements like residual parallaxes [7, 31]. But for more general displacements, the 3 line constraints give additional information.

4 Redundancy, Scaling and Consistency

A major advantage of homography-epipole parametrizations is the extent to which they eliminate the redundancy that often makes the general tensor representation rather cumbersome. With plane + parallax against a fixed reference plane, the redundancy can be entirely eliminated. The aligned m camera geometry has $3m - 4$ d.o.f.: the positions of the centres modulo an arbitrary choice of origin and a global scaling. These degrees of freedom are explicitly parametrized by, *e.g.*, the displacements $\boldsymbol{c}_{i1\,|\,i=2\ldots m}$, again modulo global rescaling. The remaining displacements can be found from $\boldsymbol{c}_{ij} = \boldsymbol{c}_i - \boldsymbol{c}_j = \boldsymbol{c}_{i1} - \boldsymbol{c}_{j1}$, and all of the matching tensors are simple linear functions of these. Conversely, the matching constraints are linear in the tensors and hence in the basic displacements $\boldsymbol{c}_{i1}$, so the complete vector of basic displacements *with the correct*

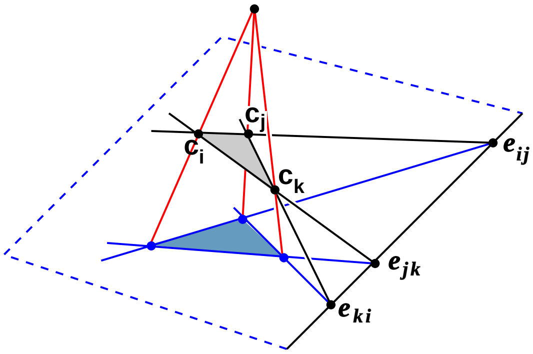

Fig. 2. The various image projections of each triplet of 3D points and/or camera centres are in Desargues correspondence [22, 4].

relative scaling can be estimated linearly from image correspondences. These properties clearly simplify reconstruction. They are possible only because plane + parallax is a *local* representation — unlike the general, redundant tensor framework, it becomes singular whenever a camera approaches the reference plane. However, the domain of validity is large enough for most real applications.

Consistency relations: As above, if they are parametrized by an independent set of inter-centre displacements, individual matching tensors in plane + parallax have no remaining internal consistency constraints and can be estimated linearly. The *inter*-tensor consistency constraints reduce to various more or less involved ways of enforcing the coincidence of versions of the same inter-camera displacement vector $\boldsymbol{c}_{ij}$ derived from different tensors, and the vanishing of cyclic sums of displacements:

$$\boldsymbol{c}_{ji} \wedge \boldsymbol{c}_{ij} \;=\; \boldsymbol{0} \qquad\qquad \boldsymbol{c}_{ij} \wedge (\boldsymbol{c}_{ij})' \;=\; \boldsymbol{0} \qquad\qquad \boldsymbol{c}_{ij} + \boldsymbol{c}_{jk} + \boldsymbol{c}_{kl} + \ldots + \boldsymbol{c}_{mi} \;=\; \boldsymbol{0}$$

In particular, each **cyclic triplet** of non-coincident epipoles is not only *aligned*, but has a *unique* consistent relative scaling $\boldsymbol{c}_{ij} \;\equiv\; \lambda_{ij}\, \boldsymbol{e}_{ij}$:

$$[\, \boldsymbol{e}_{ij}, \boldsymbol{e}_{jk}, \boldsymbol{e}_{ki} \,] \;=\; 0 \qquad \Longleftrightarrow \qquad \boldsymbol{c}_{ij} + \boldsymbol{c}_{jk} + \boldsymbol{c}_{ki} \;=\; \boldsymbol{0}$$

This and similar cyclic sums can be used to linearly recover the missing displacement scales. However, this fails if the 3D camera centres are aligned: the three epipoles coincide, so the vanishing of their cyclic sum still leaves 1 d.o.f. of relative scaling freedom. This corresponds to the well-known singularity of many fundamental matrix based reconstruction and transfer methods for aligned centres [40]. Trifocal or observation (depth recovery) based methods [32, 37] must be used to recover the missing scale factors in this case.

The cyclic triplet relations essentially encode the coplanarity of triplets of optical centres. All three epipoles lie on the line of intersection of this plane with the reference plane. Also, the three images of any fourth point or camera centre form a Desargues theorem configuration with the three epipoles (see fig. 2). A multi-camera geometry induces multiple, intricately interlocking Desargues configurations — the reference plane 'signature' of its coherent 3D geometry.

5 Depth Recovery and Closure Relations

Closure relations: In the general projective case, the **closure relations** are the bilinear constraints between the (correctly scaled) matching tensors and the projection matrices, that express the fact that the former are functions of the latter [35, 40]. **Closure based reconstruction** [38, 40] uses this to recover the projection matrices linearly from the matching tensors. In plane + parallax, the closure relations trivialize to identities of the form $\boldsymbol{c}_{ij} \wedge (\boldsymbol{c}_i - \boldsymbol{c}_j) = \boldsymbol{0}$ (since $\boldsymbol{c}_{ij} = \boldsymbol{c}_i - \boldsymbol{c}_j$). Closure based reconstruction just reads off a consistent set of $\boldsymbol{c}_i$'s from these linear constraints, with an arbitrary choice of origin and global scaling. $\boldsymbol{c}_i \equiv \boldsymbol{c}_{i1}$ is one such solution.

Depth recovery relations: Attaching the projection matrices in the closure relations to a 3D point gives **depth recovery relations** linking the matching tensors to correctly scaled image points [35, 32, 40]. These are used, *e.g.* for projective depth (scale factor) recovery in factorization based projective reconstruction [32, 37]. For plane + parallax registered points and lines with unknown relative scales, the first few depth recovery relations reduce to:

$$\boldsymbol{c}_{ij} \wedge (\lambda_i \boldsymbol{x}_i - \lambda_j \boldsymbol{x}_j) = \boldsymbol{0} \qquad \text{epipolar} \qquad (19)$$

$$\boldsymbol{c}_{ij}\, (\lambda_k\, \boldsymbol{x}_k - \lambda_i\, \boldsymbol{x}_i)^\top - (\lambda_j\, \boldsymbol{x}_j - \lambda_i\, \boldsymbol{x}_i)\, (\boldsymbol{c}_{ik})^\top = \boldsymbol{0} \qquad \text{trifocal} \qquad (20)$$

$$(\mu_i\, \boldsymbol{l}_i - \mu_j\, \boldsymbol{l}_j) \cdot \boldsymbol{c}_{ij} = 0 \qquad \text{line} \qquad (21)$$

These follow immediately from the parallax equations (7,8). As before, the trifocal point relations contain two epipolar ones, plus an additional relative vector scaling proportionality: $(\lambda_i\, \boldsymbol{x}_i - \lambda_j\, \boldsymbol{x}_j) : \boldsymbol{c}_{ij} = (\lambda_i\, \boldsymbol{x}_i - \lambda_k\, \boldsymbol{x}_k) : \boldsymbol{c}_{ik}$. See fig. 1.

6 Reconstruction by Parallax Factorization

Now consider factorization based projective reconstruction under plane + parallax. Recall the general projective factorization reconstruction method [32, 37]: m cameras with 3×4 camera matrices $\mathsf{P}_{i \,|\, i=1\ldots m}$ view n 3D points $\mathsf{x}_{p \,|\, p=1\ldots n}$ to produce mn image points $\lambda_{ip}\, \boldsymbol{x}_{ip} = \mathsf{P}_i\, \mathsf{x}_p$. These projection equations can be gathered into a $3m \times n$ matrix:

$$\begin{pmatrix} \lambda_{11}\, \boldsymbol{x}_{11} & \ldots & \lambda_{1n}\, \boldsymbol{x}_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{m1}\, \boldsymbol{x}_{m1} & \ldots & \lambda_{mn}\, \boldsymbol{x}_{mn} \end{pmatrix} = \begin{pmatrix} \mathsf{P}_1 \\ \vdots \\ \mathsf{P}_m \end{pmatrix} \begin{pmatrix} \mathsf{x}_1 \ldots \mathsf{x}_n \end{pmatrix} \qquad (22)$$

So the $(\lambda\, \boldsymbol{x})$ matrix factorizes into rank 4 factors. Any such factorization amounts to a projective reconstruction: the freedom is exactly a 4×4 projective change of coordinates H, with $\mathsf{x}_p \rightarrow \mathsf{H}\,\mathsf{x}_p$ and $\mathsf{P}_i \rightarrow \mathsf{P}_i\, \mathsf{H}^{-1}$. With noisy data the factorization is not exact, but we can use a numerical method such as truncated SVD to combine the measurements and estimate an approximate factorization and structure. To implement this with image measurements, we need to recover the unknown projective depths (scale factors) λ_{ip}. For this we use matching tensor based depth recovery relations such as

$\boldsymbol{F}_{ij}\,(\lambda_{jp}\,\boldsymbol{x}_{jp}) \;=\; \boldsymbol{e}_{j\,i} \wedge (\lambda_{ip}\,\boldsymbol{x}_{ip})$ [35, 32, 37]. Rescaling the image points amounts to an implicit projective reconstruction, which the factorization consolidates and concretizes. For other factorization based SFM methods, see (among others) [34, 27, 18, 25, 26].

Plane + parallax point factorization: The general rank 4 method continues to work under plane + parallax with aligned points $\boldsymbol{x}_{ip}$, but in this case a more efficient rank 1 method exists, that exploits the special form of the aligned projection matrices:

1. Align the mn image points to the reference plane and (as for the general-case factorization) estimate their scale factors λ_{ip} by chaining together a network of plane + parallax depth recovery relations (19) or (20).
2. Choose a set of arbitrary weights ρ_i with $\sum_{i=1}^{m}\,\rho_i \,=\, 1$. We will work in a 3D frame based at the weighted average of the projection centres: *i.e.* $\bar{\boldsymbol{c}} \;=\; \sum_{i=1}^{m}\rho_i\,\boldsymbol{c}_i$ will be set to $\boldsymbol{0}$. For the experiments we work in an average-of-centres frame $\rho_i \,=\, \frac{1}{m}$. Alternatively, we could choose some image j as a base image, $\rho_i \;=\; \delta_{ij}$.
3. Calculate the weighted mean of the rescaled images of each 3D point, and their residual parallaxes relative to this in each image. The theoretical values are given for reference, based on (3) and our choice of frame $\bar{\boldsymbol{c}} \to \boldsymbol{0}$:

$$\bar{\boldsymbol{x}}_p \;\equiv\; \textstyle\sum_{i=1}^{m}\rho_i\;(\lambda_{ip}\,\boldsymbol{x}_{ip}) \qquad \approx \qquad \boldsymbol{x}_p - w_p\,\bar{\boldsymbol{c}} \qquad \longrightarrow \qquad \boldsymbol{x}_p \tag{23}$$

$$\boldsymbol{\delta x}_{ip} \;\equiv\; \lambda_{ip}\,\boldsymbol{x}_{ip} - \bar{\boldsymbol{x}}_p \qquad \approx \qquad -\,(\boldsymbol{c}_i - \bar{\boldsymbol{c}})\;w_p \qquad \longrightarrow \qquad -\,\boldsymbol{c}_i\,w_p \tag{24}$$

4. Factorize the combined residual parallax matrix to rank 1, to give the projection centres $\boldsymbol{c}_i$ and point depths w_p, with their correct relative scales:

$$\begin{pmatrix} \boldsymbol{\delta x}_{11} & \dots & \boldsymbol{\delta x}_{1n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\delta x}_{m1} & \dots & \boldsymbol{\delta x}_{mn} \end{pmatrix} \approx \begin{pmatrix} -\boldsymbol{c}_1 \\ \vdots \\ -\boldsymbol{c}_m \end{pmatrix} \begin{pmatrix} w_1 & \dots & w_n \end{pmatrix} \tag{25}$$

   The ambiguity in the factorization is a single global scaling $\boldsymbol{c}_i \to \mu\,\boldsymbol{c}_i$, $w_p \to w_p/\mu$ (the length scale of the scene).
5. The final reconstructions are $\mathsf{P}_i \;=\; \begin{pmatrix}\boldsymbol{I} & -\boldsymbol{c}_i\end{pmatrix}$ and $\mathsf{x}_p \;=\; \begin{pmatrix}\bar{\boldsymbol{x}}_p \\ w_p\end{pmatrix}$.

This process requires the initial plane + parallax alignment, and estimates of the epipoles for projective depth recovery. It returns the 3D structure and camera centres in a projective frame that places the reference plane at infinity and the origin at the weighted average of camera centres.

With affine coordinates on the reference plane, the heights w_p reduce to inverse depths $1/z_p$ (w.r.t. the projectively distorted frame). Several existing factorization based SFM methods try to cancel the camera rotation and then factor the resulting translational motion into something like (inverse depth)·(translation), *e.g.* [12, 25, 21]. Owing to perspective effects, this is usually only achieved approximately, which leads to an iterative method. Here we require additional knowledge — a known, alignable reference plane and known epipoles for depth recovery — and we recover only projective structure, but this allows us to achieve exact results from perspective images with a single non-iterative rank 1 factorization. It would be interesting to investigate the relationships between our method and [25, 26, 21], but we have not yet done so.

Line Factorization: As in the general projective case, lines can be integrated into the point factorization method using via points. Each line is parametrized by choosing two arbitrary (but well-spaced) points on it in one image. The corresponding points on other images of the line are found by epipolar or trifocal point transfer, and the 3D via points are reconstructed using factorization. It turns out that the transfer process automatically gives the correct scale factor (depth) for the via points:

$$\mathbf{x}_i \;\equiv\; -\frac{\mathbf{l}_i \wedge (\mathbf{F}_{ij}\,\mathbf{x}_j)}{\mathbf{l}_i\cdot\mathbf{e}_{ji}} \quad \begin{array}{c}\text{general}\\ \text{case}\end{array} \qquad\qquad \mathbf{x}_j \;\equiv\; \mathbf{x}_i + \frac{\mathbf{l}_j\cdot\mathbf{x}_i}{\mathbf{l}_j\cdot\mathbf{e}_{ij}}\,\mathbf{e}_{ij} \quad \begin{array}{c}\text{plane + parallax}\\ \text{case}\end{array} \tag{26}$$

Under plane + parallax, all images $\mathbf{z}\wedge\mathbf{c}_i + \mathbf{l}$ of a line $(\mathbf{l},\mathbf{z})$ intersect in a common point $\mathbf{z}$. If we estimate this first, only one additional via point is needed for the line.

Plane factorization: As mentioned in §2, inter-image homographies $\mathbf{H}_{i0}$ induced by 3D planes against a fixed base image 0 can also be incorporated in the above factorization, simply by treating their three columns as three separate point 3-vectors. Under plane + parallax, once they are scaled correctly as in §2, the homographies take the form (9). Averaging over i as above gives an $\bar{\mathbf{H}}_{i0}$ of the same form, with $\mathbf{c}_{i0}$ replaced by $\bar{\mathbf{c}} - \mathbf{c}_0 \;\rightarrow\; -\mathbf{c}_0$. So the corresponding "homography parallaxes" $\boldsymbol{\delta}\mathbf{H}_{i0} \;=\; \frac{\mathbf{c}_i\,\mathbf{n}^\top}{\mathbf{n}\cdot\mathbf{c}_0+d}$ factor as for points, with $\frac{\mathbf{n}^\top}{\mathbf{n}\cdot\mathbf{c}_0+d}$ in place of w_p. Alternatively, if $\mathbf{c}_0$ is taken as origin and the $\boldsymbol{\delta}$'s are measured against image 0, $\mathbf{I}$ rather than $\bar{\mathbf{H}}_{i0}$ is subtracted.

Optimality properties: Ideally, we would like our structure and motion estimates to be optimal in some sense. For point estimators like maximum likelihood or MAP, this amounts to globally minimizing a measure of the (robustified, covariance-weighted) total squared image error, perhaps with overfitting penalties, *etc*. Unfortunately — as with all general closed-form projective SFM methods that we are aware of, and notwithstanding its excellent performance in practice — plane + parallax factorization uses an algebraically simple but statistically suboptimal error model. Little can be done about this, beyond using the method to initialize an iterative nonlinear refinement procedure (*e.g.* bundle adjustment). As in other estimation problems, it is safest to refine the results after each stage of the process, to ensure that the input to the next stage is as accurate and as outlier-free as possible. But even if the aligning homographies are refined in this way before being used (*c.f.* [9, 1, 11]), the projective centering and factorization steps are usually suboptimal because the projective rescaling $\lambda_{ip} \neq 1$ skews the statistical weighting of the input points. In more detail, by pre-weighting the image data matrix before factorization, affine factorization [34] can be generalized to give optimal results under an image error model as general as a per-image covariance times a per-3D-point weight[3]. But this is no longer optimal in projective factorization: even if the input er-

[3] *I.e.* image point $\mathbf{x}_{ip}$ has covariance $\rho_p\,\mathbf{C}_i$, where $\mathbf{C}_i$ is a fixed covariance matrix for image i and ρ_p a fixed weight for 3D point p. Under this error model, factoring the weighted data matrix $(\rho_p^{-1/2}\,\mathbf{C}_i^{-1/2}\,\mathbf{x}_{ip})$ into weighted camera matrices $\mathbf{C}^{-1/2}\,\mathsf{P}_i$ and 3D point vectors $\rho_p^{-1/2}\,\mathsf{X}_p$ gives statistically optimal results. *Side note:* For typical images at least 90–95% of the image energy is in edge-like rather than corner-like structures ("the aperture problem"). So assuming that the (residual) camera rotations are small, an error model that permitted each 3D point to have its own highly anisotropic covariance matrix would usually be more appropriate than a per-image covariance. Irani & Anandan [20] go some way towards this by introducing an initial reduction based on a higher rank factorization of transposed weighted point vectors.

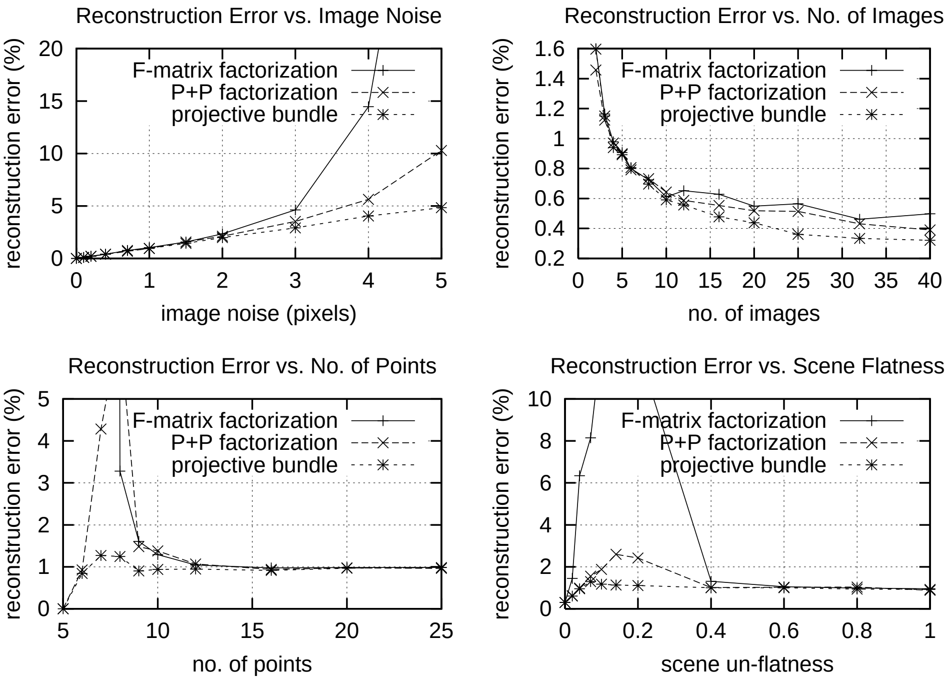

Fig. 3. A comparison of 3D reconstruction errors for plane + parallax SFM factorization, fundamental matrix based projective factorization [32, 37], and projective bundle adjustment.

rors are uniform, rescaling by the non-constant factors λ_{ip} distorts the underlying error model. In the plane + parallax case, the image rectification step further distorts the error model whenever there is non-negligible camera rotation. In spite of this, our experiments suggest that plane + parallax factorization gives near-optimal results in practice.

7 Experiments

Figure 3 compares the performance of the plane + parallax point factorization method described above, with conventional projective factorization using fundamental matrix depth recovery [32, 37], and also with projective bundle adjustment initialized from the plane + parallax solution. Cameras about 5 radii from the centre look inwards at a synthetic spherical point cloud cut by a reference plane. Half the points (but at least 4) lie on the plane, the rest are uniformly distributed in the sphere. The image size is 512×512, the focal length 1000 pixels. The cameras are uniformly spaced around a $90°$ arc centred on the origin. The default number of views is 4, points 20, Gaussian image noise 1 pixel. In the scene flatness experiment, the point cloud is progressively flattened onto the plane. The geometry is strong except under strong flattening and for small numbers of points.

The main conclusions are that plane + parallax factorization is somewhat more accurate than the standard fundamental matrix method, particularly for near planar scenes and linear fundamental matrix estimates, and often not far from optimal. In principle this was to be expected given that plane + parallax applies additional scene constraints (known coplanarity of some of the observed points). However, additional processing steps are involved (plane alignment, point centring), so it was not clear *a priori* how effectively the coplanarity constraints could be used. In fact, the two factorizations have very similar average reprojection errors in all the experiments reported here, which suggests that the additional processing introduces very little bias. The plane + parallax method's greater stability is confirmed by the fact that its factorization matrix is consistently a little better conditioned than that of the fundamental matrix method (*i.e.* the ratio of the smallest structure to the largest noise singular value is larger).

8 Summary

Plane + parallax alignment greatly simplifies multi-image projective geometry, reducing matching tensors and constraints, closure, depth recovery and inter-tensor consistency relations to fairly simple functions of the (correctly scaled!) epipoles. Choosing projective plane + parallax coordinates with the reference plane at infinity helps this process by providing a (weak, projective) sense in which reference plane alignment cancels out precisely the camera rotation and calibration changes. This suggests a fruitful analogy with the case of translating calibrated cameras and a simple interpretation of plane + parallax geometry in terms of 3D displacement vectors.

The simplified parallax formula allows exact projective reconstruction by a simple rank-one (centre of projection)·(height) factorization. Like the general projective factorization method [32, 37], an initial scale recovery step based on estimated epipoles is needed. When the required reference plane is available, the new method appears to perform at least as well as the general method, and significantly better in the case of near-planar scenes. Lines and homography matrices can be integrated into the point-based method, as in the general case.

Future work: We are still testing the plane + parallax factorization and refinements are possible. It would be interesting to relate it theoretically to affine factorization [34], and also to Oliensis's family of bias-corrected rotation-cancelling multiframe factorization methods [25, 26]. Bias correction might be useful here too, although our centred data is probably less biased than the key frames of [25, 26].

The analogy with translating cameras is open for exploration, and more generally, the idea of using a projective choice of 3D and image frames to get closer to a situation with a simple, special-case calibrated method, thus giving a simplified *projective* one. *E.g.* we find that suitable projective rectification of the images often makes affine factorization [34] much more accurate as a *projective* reconstruction method.

One can also consider autocalibration in the plane + parallax framework. It is easy to derive analogues of [41] (if only structure on the reference plane is used), or [16, 39] (if the off-plane parallaxes are used as well). But so far this has not lead to any valuable simplifications or insights. Reference plane alignment distorts the camera calibrations, so the aligning homographies can not (immediately) be eliminated from the problem.

References

[1] A. Capel, D. and Zisserman. Automated mosaicing with super-resolution zoom. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 885–891, June 1998.

[2] S. Carlsson. Duality of reconstruction and positioning from projective views. In P. Anandan, editor, *IEEE Workshop on Representation of Visual Scenes*. IEEE Press, 1995.

[3] S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from multiple images. *Int.J. Computer Vision*, 27(3):227–241, May 1998.

[4] A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *European Conf. Computer Vision*, pages 846–861. Springer-Verlag, 1998.

[5] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Int. Conf. Computer Vision*, pages 951–6, 1995.

[6] O. Faugeras and T. Papadopoulo. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Transactions of the Royal society A*, 1998.

[7] K. Hanna and N. Okamoto. Combining stereo and motion analysis for direct estimation of scene structure. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 357–65, 1993.

[8] R. Hartley. Lines and points in three views and the trifocal tensor. *Int.J. Computer Vision*, 22(2):125–140, 1997.

[9] R. Hartley. Self calibration of stationary cameras. *Int.J. Computer Vision*, 22(1):5–23, 1997.

[10] R. Hartley and G. Debunne. Dualizing scene reconstruction algorithms. In R. Koch and L. Van Gool, editors, *Workshop on 3D Structure from Multiple Images of Large-scale Environments SMILE'98*, pages 14–31. Springer-Verlag, 1998.

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[12] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *Int.J. Computer Vision*, 7:95–117, 1992.

[13] A. Heyden. Reconstruction from image sequences by means of relative depths. In E. Grimson, editor, *Int. Conf. Computer Vision*, pages 1058–63, Cambridge, MA, June 1995.

[14] A. Heyden and K. Åström. A canonical framework for sequences of images. In *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA, June 1995.

[15] A. Heyden and K. Åström. Algebraic varieties in multiple view geometry. In *European Conf. Computer Vision*, pages 671–682. Springer-Verlag, 1996.

[16] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Int. Conf. Pattern Recognition*, pages 339–43, Vienna, 1996.

[17] A. Heyden and K. Åström. Algebraic properties of multilinear constraints. *Mathematical Methods in the Applied Sciences*, 20:1135–1162, 1997.

[18] A. Heyden, R. Berthilsson, and G. Sparr. An iterative factorization method for projective structure and motion from image sequences. *Image & Vision Computing*, 17(5), 1999.

[19] M. Irani and P. Anadan. Parallax geometry of pairs of points for 3d scene analysis. In *European Conf. Computer Vision*, pages 17–30. Springer-Verlag, 1996.

[20] M. Irani and P. Anadan. Factorization with uncertainty. In *European Conf. Computer Vision*. Springer-Verlag, 2000.

[21] M. Irani, P. Anadan, and M. Cohen. Direct recovery of planar-parallax from multiple frames. In *Vision Algorithms: Theory and Practice*. Springer-Verlag, 1999.

[22] M. Irani, P. Anadan, and D. Weinshall. From reference frames to reference planes: Multiview parallax geometry and applications. In *European Conf. Computer Vision*, pages 829–845. Springer-Verlag, 1998.

[23] M. Irani and P. Anandan. A unified approach to moving object detection in 2d and 3d scenes. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 20(6):577–589, June 1998.

[24] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna. Representation of scenes from collections of images. In *IEEE Workshop on Representations of Visual Scenes*, pages 10–17, June 1995.

[25] J. Oliensis. Multiframe structure from motion in perspective. In *IEEE Workshop on Representation of Visual Scenes*, pages 77–84, June 1995.

[26] J. Oliensis and Y. Genc. Fast algorithms for projective multi-frame structure from motion. In *Int. Conf. Computer Vision*, pages 536–542, Corfu, Greece, 1999.

[27] C.J. Poelman and T. Kanade. A parapersective factorization method for shape and motion recovery. In *European Conf. Computer Vision*, pages 97–108, Stockholm, 1994. Springer-Verlag.

[28] A. Shashua and S. Avidan. The rank 4 constraint in multiple (≥ 3) view geometry. In *European Conf. Computer Vision*, pages 196–206, Cambridge, 1996.

[29] A. Shashua and M. Werman. On the trilinear tensor of three perspective views and its underlying geometry. In *Int. Conf. Computer Vision*, Boston, MA, June 1995.

[30] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[31] G. Stein and A. Shashua. Model-based brightness constraints: On direct estimation of structure and motion. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 400–406, 1997.

[32] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *European Conf. Computer Vision*, pages 709–20, Cambridge, U.K., 1996. Springer-Verlag.

[33] R. Szeliski and S-B. Kang. Direct methods for visual scene reconstruction. In *IEEE Workshop on Representation of Visual Scenes*, pages 26–33, June 1995.

[34] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int.J. Computer Vision*, 9(2):137–54, 1992.

[35] B. Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. Submitted to *Int.J. Computer Vision*.

[36] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *Int. Conf. Computer Vision*, pages 338–43, Cambridge, MA, June 1995.

[37] B. Triggs. Factorization methods for projective structure and motion. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 845–51, San Francisco, CA, 1996.

[38] B. Triggs. Linear projective reconstruction from matching tensors. In *British Machine Vision Conference*, pages 665–74, Edinburgh, September 1996.

[39] B. Triggs. Autocalibration and the absolute quadric. In *Int. Conf. Computer Vision & Pattern Recognition*, Puerto Rico, 1997.

[40] B. Triggs. Linear projective reconstruction from matching tensors. *Image & Vision Computing*, 15(8):617–26, August 1997.

[41] B. Triggs. Autocalibration from planar scenes. In *European Conf. Computer Vision*, pages I 89–105, Freiburg, June 1998.

[42] D. Weinshall, P. Anandan, and M. Irani. From ordinal to euclidean reconstruction with partial scene calibration. In R. Koch and L. Van Gool, editors, *3D Structure from Multiple Images of Large-scale Environments SMILE'98*, pages 208–223. Springer-Verlag, 1998.

[43] D. Weinshall, M. Werman, and A. Shashua. Shape tensors for efficient and learnable indexing. In *IEEE Workshop on Representation of Visual Scenes*, pages 58–65, June 1995.

[44] L. Zelnik-Manor and M. Irani. Multi-frame alignment of planes. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 151–156, 1999.

[45] L. Zelnik-Manor and M. Irani. Multi-view subspace constraints on homographies. In *Int. Conf. Computer Vision*, pages 710–715, 1999.