From Reference Frames to Reference Planes: Multi-View Parallax Geometry and Applications⋆

M. Irani[1], P. Anandan[2], and D. Weinshall[3]

[1] Dept. of Applied Math and CS, The Weizmann Inst. of Science, Rehovot, Israel,
`irani@wisdom.weizmann.ac.il`
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA,
`anandan@microsoft.com`,
[3] Institute of Computer Science Hebrew University 91904 Jerusalem, Israel
`daphna@cs.huji.ac.il`

Abstract. This paper presents a new framework for analyzing the geometry of multiple 3D scene points from *multiple* uncalibrated images, based on decomposing the projection of these points on the images into two stages: (i) the projection of the scene points onto a (real or virtual) physical reference planar surface in the scene; this creates a virtual "image" on the reference plane, and (ii) the re-projection of the virtual image onto the actual image plane of the camera. The positions of the virtual image points are directly related to the 3D locations of the scene points and the camera centers relative to the reference plane alone. All dependency on the internal camera calibration parameters and the orientation of the camera are folded into homographies relating each image plane to the reference plane.

Bi-linear and tri-linear constraints involving multiple points and views are given a concrete physical interpretation in terms of geometric relations on the physical reference plane. In particular, the possible dualities in the relations between scene points and camera centers are shown to have simple and symmetric mathematical forms. In contrast to the *plane+parallax* (p+p) representation, which also uses a reference plane, the approach described here removes the dependency on a reference image plane and extends the analysis to multiple views. This leads to simpler geometric relations and complete symmetry in multi-point multi-view duality.

The simple and intuitive expressions derived in the reference-plane based formulation lead to useful applications in 3D scene analysis. In particular, simpler tri-focal constraints are derived that lead to simple methods for New View Synthesis. Moreover, the separation and compact packing of the *unknown* camera calibration and orientation into the 2D projection transformation (a homography) allows also partial reconstruction using partial calibration information.

Keywords: Multi-point multi-view geometry, uncalibrated images, new view synthesis, duality of cameras and scene points, plane+parallax, trilinearity.

1 Introduction

The analysis of 3D scenes from multiple *perspective* images has been a topic of considerable interest in the vision literature. Given two calibrated cameras, their relative orientations can be determined by applying the epipolar constraint to the observed image points, and the 3D structure of the scene can be recovered relative to the coordinate frame of a reference camera (referred to here as the *reference frame*–e.g., see [13, 6]). This is done by using the *epipolar constraint* and recovering the "Essential Matrix" E which depends on the rotation R and translation T between the two cameras. Constraints directly involving the image positions of a point in three calibrated views of a point have also been derived [19].

If the calibration of the cameras is unavailable, then it is known that reconstruction is still possible from two views, but only up to a 3D projective transformation [4]. In this case the *epipolar constraint* still holds, but the Essential Matrix is replaced by the "Fundamental Matrix", which also incorporates the unknown camera calibration information. The 3D scene points, the camera centers and their image positions are represented in 3D and 2D projective spaces (using homogeneous projective coordinates). In this case, the **reference frame** reconstruction may either be a reference camera coordinate frame [8], or as defined by a set of 5 basis points in the 3D world [14]. A complete set of constraints relating the image positions of multiple points in multiple views have been derived [5, 15]. Alternatively, given a projective coordinate system specified by 5 basis points, the set of constraints directly relating the projective coordinates of the camera centers to the image measurements (in 2D projective coordinates) and their dual constraints relating to the projective coordinates of the 3D scene points have also been derived [2, 20].

Alternatively, multiple uncalibrated images can be handled using the "plane + parallax" (P+P) approach, which analyzes the parallax displacements of a point between two views relative to a (real or virtual) physical planar surface Π in the scene [16, 12, 11]. The magnitude of the parallax displacement is called the "*relative-affine structure*" in [16]. [12] shows that this quantity depends both on the "Height" H of P from Π and its depth Z relative to the reference camera. Since the *relative-affine-structure* measure is relative to both the *reference frame* (through Z) and the reference plane (through H), we refer to the P+P framework also as the **reference-frame + reference-plane** formulation. The P+P has the practical advantage that it avoids the inherent ambiguities associated with estimating the relative orientation (rotation + translation) between the cameras; this is because it requires only estimating the homography induced by the reference plane between the two views, which folds together the rotation and translation. Also, when the scene is "flat", the F matrix estimation is unstable, whereas the planar homography can be reliably recovered [18].

In this paper, we remove the dependency on the *reference frame* of the analysis of multi-point multi-view geometry. We break down the projection from 3D to 2D into 2 operations: the projection of the 3D world onto the 2D reference plane Π, followed by a 2D projective transformation (homography) which maps

the reference plane to the image plane. Given the "virtual images" formed by the projection onto the reference plane, we derive algebraic and geometric relations involving the image locations of multiple points in multiple views in these virtual images. The positions of virtual image points are directly related to the 3D locations of the scene points and the camera centers relative to the reference plane alone. All dependency on the internal camera calibration parameters and the orientation of the camera are folded into homographies relating each image plane to the reference plane. We obtain a structure measure that depends *only* on the heights of the scene points relative to the reference plane

In this paper, we derive a complete set dual relationships involving 2 and 3 points in 2 and 3 views. On the reference plane the multi-point multi-view geometry is simple and intuitive. These relations are directly related to physical points on the reference plane such as the *epipole* and the *dual-epipole*[9]. We identify these points, and also two new entities called the *tri-focal line* and the *dual trifocal-line* which are analogous to the epipole and the dual-epipole when considering three-view and three-point geometries on the reference plane. Structures such as the fundamental matrix and the trilinear tensor have a rather simple form and depend only on the epipoles, and nothing else. The symmetry between points and cameras is complete, and they can be simply switched around to get from the epipolar geometry to the dual-epipolar geometry.

The simple and intuitive expressions derived in the reference-plane based formulation in this paper lead to useful applications in 3D scene analysis. In particular, simpler tri-focal constraints are derived, and these lead to simple methods for New View Synthesis. Also, the separation and compact packing of the *unknown* camera calibration and orientation into the 2D projection transformation (a homography) that relates the image plane to the reference plane, leads to potentially powerful reconstruction and calibration algorithms. For instance, based on minimal partial domain information, partial calibration and partial reconstruction can be achieved. This is also briefly discussed in this paper.

The remainder of this paper is organized as follows: Section 2 introduces our notations, and describes the two-view geometric and algebraic constraints (bilinearity and parallax) in the reference plane representation. Section 3 describes duality (between scene points and camera centers) on the reference plane. Section 4 examines the relations involving 3 views and the corresponding dual relations. Section 5 discusses applications of this representation and shows initial results for one particular application, namely new-view synthesis.

2   Two View Geometry on the Reference Plane

Figure 1 illustrates the two stage decomposition of the image formation process. Figure 1a shows the projection of one scene point from two camera centers onto the reference plane Π. Figure 1b shows the re-projection from the plane to one of the camera image planes (the "reference frame"). In this and in all subsequent figures in this paper, we adopt the following notations: P denotes scene points in 3D, C denotes camera centers; i, j, k are indices used for scene points (e.g.,

Fig. 1. The reference plane representation: (a) the projection of the points onto the reference plane itself, removing the dependency on the reference image plane. p_{ts} is the epipole, and the red line is the epipolar line. (b) the *re*-projection of the reference plane image onto a reference image frame (camera "t").

P_i) and r, s, t are indices used for camera centers (e.g., C_r). Also, p_{it} denotes the projection of the scene point P_i through camera center C_t. It is the intersection of the ray $P_i C_t$ with the reference plane Π. Similarly p_{is} is the intersection of $P_i C_s$ with the reference plane. We define p_{it} and p_{is} as the "virtual-images" of P_i on the reference plane from cameras C_t and C_s respectively. We define the intersection of $C_t C_s$ with Π as the **epipole** on Π. We use p_{ts} to denote the *epipole*. Note that the location of the epipole p_{ts} on the reference plane Π is is independent of the orientations and the internal calibration parameters of the cameras s and t.

To derive the algebraic constraints involving multiple points in multiple views, we define a coordinate system (x, y, Z) relative to the reference plane Π, where (x, y) are parallel to Π and Z is perpendicular to it. For points on the reference plane, we define $Z = 1$, for other points we define $Z = H + 1$, where H denotes the height (i.e., the perpendicular distance) of the point from the plane Π. Thus, $P_i = (x_i, y_i, Z_i)$, where $Z_i = H_i + 1$, denotes the 3D coordinates of the scene point P_i. Similarly $C_t = (x_t, y_t, Z_t)^T$, where $Z_t = H_t + 1$, and and $C_s = (x_s, y_s, Z_s)^T$, where $Z_s = H_s + 1$. The points p_{it}, p_{is} and p_{ts} on the reference plane are the intersections of the lines $C_t P_i$, $C_s P_i$, and $C_t C_s$ with the reference plane Π:

$$p_{it} = \begin{pmatrix} x_{it} \\ y_{it} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{H_i x_t - H_t x_i}{H_i - H_t} \\ \frac{H_i y_t - H_t y_i}{H_i - H_t} \\ 1 \end{pmatrix} \qquad p_{is} = \begin{pmatrix} x_{it} \\ y_{it} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{H_i x_s - H_s x_i}{H_i - H_s} \\ \frac{H_i y_s - H_s y_i}{H_i - H_s} \\ 1 \end{pmatrix} \quad (1)$$

$$p_{ts} = \begin{pmatrix} x_{ts} \\ y_{ts} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{H_s x_t - H_t x_s}{H_s - H_t} \\ \frac{H_s y_t - H_t y_s}{H_s - H_t} \\ 1 \end{pmatrix} \quad (2)$$

Note that the expressions given above do not involve any of the camera internal calibration parameters or the orientations of the image planes. Also note that

there is only a single epipole, which is unlike the case of the reference-frame based formulation, which involves two epipoles, one on each image frame.

The points p_{it}, p_{is} and p_{ts} on the reference plane are related to their corresponding points on an image plane (e.g., a reference image) via a single 2D projective transformation, which is the homography between that image plane and the plane Π. Figure 1b shows the re-projection onto the reference image t — the points p, p_w, and e are the projections of the image points p_{it}, p_{is}, and the epipole p_{ts} respectively.

There are two basic results concerning two views of a point as observed on the reference plane Π. The first is the expression for the "parallax" on the reference plane, and the second is the bilinear constraint involving the two image locations of the scene point and the epipole. These are described below.

Parallax on the Reference Plane: Given the expressions in Equations 1 and 2 , it can be easily verified that

$$p_{is} - p_{it} = \gamma(p_{is} - p_{ts}), \tag{3}$$

where $\gamma = \frac{H_i(H_t - H_s)}{(H_t - H_i)H_s}$

Note that this expression for parallax (Equation 3) involves only the heights of the scene point and of the camera centers relative to the reference plane Π. It *does not* include any quantities relative to any of the camera coordinate systems (e.g., the *reference frame*) such as Z or T_Z as before. Also, the parallax magnitude γ does not depend on the x, y locations of either the camera centers or the scene point[1].

The Bilinear Constraint: Equation 3 implies that $p_{it}, p_{is},$ *and* p_{ts} *are collinear.* Similar to the definition of the epipolar line on the image plane, the line containing these three points on Π is the intersection of the epipolar plane containing P_i, C_t, and C_s with Π. Thus, this is the *epipolar* line as observed on the reference-plane. The collinearity of these three points can be expressed as $p_{it}{}^T F p_{is} = 0$ where $F = \begin{bmatrix} 0 & 1 & -y_{ts} \\ -1 & 0 & x_{ts} \\ y_{ts} & -x_{ts} & 0 \end{bmatrix}$ is the "Fundamental Matrix". As opposed to the reference frame based formulation, where the fundamental matrix depends on the camera rotations and the internal calibration parameters of the camera, here it depends only on the epipole. Moreover, the epipole is *explicit* in the F matrix here, whereas, it is *implicit* in the standard formulation.

---

[1] The expression for $\gamma = \frac{HT_Z}{Zd_\pi}$ in the P+P case can be related to the current expression as follows: Consider a *virtual* camera centered at C_t, whose image plane is the plane Π, and its optical axis coincides with the H direction. Then $H = H_i$, $Z = H_t - H_i$, $T_Z = H_t - H_s$ and $d_\pi = H_s$.

What happens when the epipole goes to ∞? In Equation 2, it can be seen that when $H_s = H_t$, the epipole p_{ts} goes to ∞. In this case,

$$p_{ts} = \begin{pmatrix} x_{ts} \\ y_{ts} \\ 0 \end{pmatrix} = \begin{pmatrix} x_t - x_s \\ y_t - y_s \\ 0 \end{pmatrix},$$

and the expression for parallax can be rewritten as: $(p_{is} - p_{it}) = \frac{H_i}{(H_t - H_i)} p_{ts}$. In other words, all the parallax vectors are parallel to each other (i.e., meet at ∞). The Fundamental Matrix $F = \begin{bmatrix} 0 & 0 & -y_{ts} \\ 0 & 0 & x_{ts} \\ y_{ts} & -x_{ts} & 0 \end{bmatrix}$. We can, of course, unify the finite and the infinite case by using 2D projective notations. However, in this paper we choose to use 2D Euclidean coordinate representations, in order to emphasize the *physical meaning* of the various observed and derived quantities. Moreover, the parallax expression in Equation 3, which involves metric relations is meaningless in a projective coordinate representation.

Also, when $H_t = H_i$ or $H_s = H_i$, then p_{it} or p_{is} go to ∞ respectively. This occurs when, fortuitously, the plane Π is chosen to be parallel to the optic ray from the scene point to one of the cameras. In this case, the corresponding image point cannot be observed on the reference plane, and our analysis does not apply.

3   Duality on the Reference Plane

In this section, we derive a set of dual relations on the reference-plane by switching the roles of camera centers and scene points as was previously done in [2, 20].

Consider two points P_i and P_j and one camera center C_t. Consider the intersection of the rays $P_i P_j$, $P_i C_t$ and $P_j C_t$ with the reference plane Π (see Figure 2a). These occur respectively at p_{ij}, p_{it} and p_{jt}. In a manner analogous to the "epipolar plane" (defined by 2 camera centers and a scene point), we define the plane containing P_i, P_j and C_t (2 scene points and a camera center) as the "*dual* epipolar plane". By the same analogy, we define its intersection with Π (i.e., the line connecting p_{it}, p_{jt} and p_{ij}) as the "*dual* epipolar line", and p_{ij} as the "*dual* epipole". Note that the dual-epipole, the dual-epipolar lines, and the dual-epipolar planes relate to a pair of scene points over multiple views, in the same way the epipole, the epipolar lines, and the epipolar planes relate to multiple scene points over a pair of views.

By applying the duality of scene points and camera centers, we can derive the dual of the bilinear constraint and the parallax expressions in algebraic form. They are:

Dual Parallax: $\qquad\qquad p_{it} - p_{jt} = \gamma_d (p_{it} - p_{ij})$, where $\gamma_d = \frac{H_t (H_j - H_i)}{(H_j - H_t) H_i}$,

and

Dual Bilinearity Constraint: $\qquad\qquad p_{it}^T F_d p_{jt} = 0,$
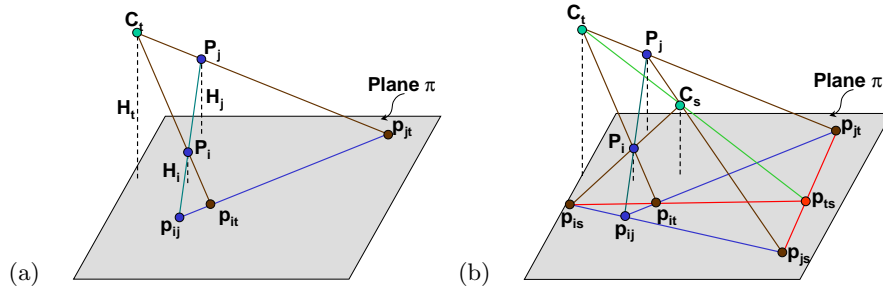
Fig. 2. Duality on the reference plane: (a) the dual-epipolar geometry associated with two points in one view. p_{ij} is the *dual*-epipole, and the blue line going through p_{ij} is the *dual*-epipolar line. (b) both sets of epipolar lines (shown in red) and dual-epipolar lines (shown in blue) that arise when considering two points in two views.

where $F_d = \begin{bmatrix} 0 & 1 & -y_{ij} \\ -1 & 0 & x_{ij} \\ y_{ij} & -x_{ij} & 0 \end{bmatrix}$ is (defined as) the "*Dual* Fundamental Matrix".

The duality of the bilinear constraint has been previously explored - e.g., Carlsson[2] and Weinshall, *et al.*[20] derive dual bilinear and trilinear relations in terms of the projective coordinate representations of the scene points, camera centers, and image points. Here, however, we derive these relations in the context of the reference plane images, and provide physical meaning to the dual relations. Also, Irani and Anandan [9] pointed out the dual epipole in the context of the plane+parallax representation. In that case, since the projection on a camera image plane ("reference frame") is included in the formulation, there exists an asymmetry in the various constraints and their dual constraints. Here, *complete* symmetry is achieved by projecting all the observed quantities onto the reference plane itself.

Figure 2b completes the picture by considering two points (P_i, P_j) in two views (C_t, C_s). This configuration gives rise to one set of epipolar lines (corresponding to each scene point) going through the epipole p_{ts}, and one set of *dual*-epipolar lines (corresponding to each camera) going through the *dual*-epipole p_{ij}.

4 Three View Geometry on the Reference Plane

In this section we extend our treatment to three views. [5] shows that there are no additional *independent* constraints that can be derived in more than three views. In this section we present a geometric interpretation of the three-view constraints in terms of physical quantities on the reference plane Π. We derive the algebraic three-view constraints and show that they have a very simple mathematical form. We will also show that the tensor-representation of these constraints in

the reference-plane has a very simple mathematical form when compared to the tensors in the standard formulations[15, 7, 5].

4.1 Geometric Observations

Figure 3 shows three views p_{is}, p_{it}, and p_{ir} of a point P_i as projected onto the reference plane Π. The new camera center is labeled as C_r, and the two new epipoles as p_{rt} and p_{sr}[2].
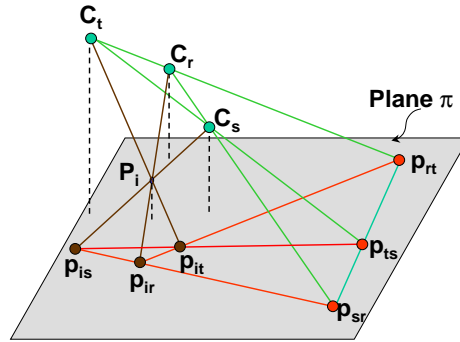

Fig. 3. Geometry of three views on the reference plane – a 3D view. The 3 red lines are the epipolar lines of pairs of views, and the turquoise line the <u>trifocal-line</u>. The 6 points on Π lie on 4 lines forming a "complete quadrilateral".

Taken pairwise at a time, the three views give rise to three epipolar constraints:

$$p_{it}, p_{is}, p_{ts} \text{ are collinear.} \qquad p_{is}, p_{ir}, p_{sr} \text{ are collinear.}$$
$$p_{ir}, p_{it}, p_{rt} \text{ are collinear.}$$

There is, however, a fourth collinearity constraint, namely:

$$\text{The epipoles } p_{ts}, p_{sr}, p_{rt} \text{ are collinear.}$$

This line is simply the intersection of the plane containing the three camera centers C_t, C_s and C_r with Π (see Figure 3). This plane is referred to as the *tri-focal plane*. Based on this definition we define the line connecting the three epipoles as the "tri-focal line".

The fact that the six points lie on four lines is fundamental to the projection of three views of a point onto a reference plane Π. Note that this figure on the

---

[2] Note that geometrically this figure is identical to Figure 2b, but the labeling of the point is different. The scene point P_j in Figure 2b has been replaced by a camera center C_r. In fact, this is because of the complete symmetry between scene points and camera centers in our representation.

plane (Figure 4a) is known as the "complete quadrilateral" and plays a central role in plane projective geometry [3].

Given the three cameras, every point in the scene forms a triangle (e.g., with vertices p_{it}, p_{is} and p_{ir}. Different points (e.g., indexed i, j, etc.) will form different triangles, all of which share the same tri-focal line (see Figure 4b). In other words, all these triangles are perspective from the tri-focal line[3].

4.2 The Trifocal Ratio

Each pair of views from the three views provides an expression for parallax similar to Equation 3. For example, consider:

$$p_{is} - p_{it} = \frac{H_i(H_t - H_s)}{(H_t - H_i)H_s}(p_{is} - p_{ts})$$
$$p_{ir} - p_{it} = \frac{H_i(H_t - H_r)}{(H_t - H_i)H_r}(p_{ir} - p_{rt})$$

$$(4)$$

From these equations we can eliminate $H_i/(H_t - H_i)$ to obtain:

$$\frac{||p_{is} - p_{it}||}{||p_{is} - p_{ts}||} = \lambda_{rst}\frac{||p_{ir} - p_{it}||}{||p_{ir} - p_{rt}||} \tag{5}$$

where $\lambda_{rst} = \frac{(H_t - H_s)}{H_s}\frac{H_r}{(H_t - H_r)}$. The above equation is true upto a sign change. Note that λ_{rst} does not depend on the point i. In other words, for every scene point, the locations of its image from the three views on the reference plane is related by the same Equation 5.

This constraint is further explored in Section 4.3.

Given two "images" of the point P_i on Π, e.g., p_{is} and p_{ir}, and the corresponding epipoles, p_{rt} and p_{ts}, we can determine the location of the third "image" p_{it} by intersecting the two epipolar lines $p_{is}p_{ts}$ and $p_{ir}p_{rt}$ (see Figure 4a).

There are, however, two cases in which the three epipolar lines collapse into a single line (and hence, their intersection is not unique). These are the same situations noted in [5, 15], but here we examine it in the context of the reference-plane images. The first is the case when the three camera centers are collinear (see Figure 5) - in this case the three epipoles collapse into a single point (denoted as e in Figure 5). The three epipolar lines also collapse into a single line, and therefore p_{it} cannot be determined by the intersection of the epipolar lines. However, given the common epipole e and λ_{rst}, p_{it} can be recovered from p_{is} and p_{ir} using Equation 5. In fact, in this case, λ_{rst} is the *cross ratio* of these four points (the three image points and the epipole).

---

[3] It is known in plane-projective geometry that if two triangles are perspective from a line they are also perspective from a point [3] – this is the converse of the Desargues' Theorem. Given the two triangles corresponding to i and j as in Figure 4b, then the point of perspectivity is in fact the *dual-epipole* p_{ij}.
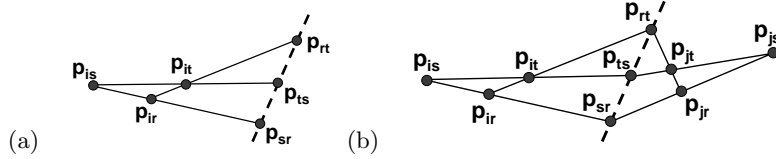
Fig. 4. Three view geometry on the reference plane: (a) the complete quadrilateral formed by the image of a single point in three views and the trifocal-line (shown as a dashed line) containing the three epipoles. (b) different triangles due to different scene points share the same trifocal-line.

Another interesting case is when the scene point P_i lies on the "tri-focal plane" of the three cameras. In this case the three image points p_{it}, p_{is} and p_{ir} all lie on the tri-focal line itself, i.e., once again the three epipolar lines collapse onto the tri-focal line. Hence we cannot use the intersection of epipolar lines to determine p_{it}. In this case too, p_{it} can be determined by Equation 5, using λ_{rst}.

The ratio λ_{rst} has a special significance. If we consider the tri-focal line, we can show (by replacing P_i with C_r in Equation 3)that:

$$p_{sr} - p_{rt} = \lambda_{rst}(p_{sr} - p_{ts}) \tag{6}$$

(Hence, the name "*trifocal-ratio*".) In other words, in the general case:

$$
\begin{aligned}
\lambda_{rst} &= \frac{||p_{sr} - p_{rt}||}{||p_{sr} - p_{ts}||} \\
&= \frac{||p_{is} - p_{it}||}{||p_{is} - p_{ts}||} \frac{||p_{ir} - p_{rt}||}{||p_{ir} - p_{it}||}
\end{aligned}
\tag{7}
$$

Note that in the singular case, when the epipoles collapse, the ratio of the distances between the epipoles (the top equation) is undefined, but the bottom equation is still valid and can be used.

4.3 The Trifocal Tensor

Returning to Equation 5, we can write down component equalities as follows:

$$\frac{x_{is} - x_{it}}{x_{is} - x_{ts}} = \frac{y_{is} - y_{it}}{y_{is} - y_{ts}} = \lambda_{rst}\frac{x_{ir} - x_{it}}{x_{ir} - x_{rt}} = \lambda_{rst}\frac{y_{ir} - y_{it}}{y_{ir} - y_{rt}} \tag{8}$$

By taking two of these equalities at a time and cross-multiplying by the denominators we can get six linear constraints. Of these two are the same as the bilinear (epipolar) constraints involving only two views at a time. The other four, which involve all three views are:

$$
\begin{aligned}
(x_{is} - x_{it})(x_{ir} - x_{rt}) &= \lambda_{rst}(x_{ir} - x_{it})(x_{is} - x_{ts}) \\
(x_{is} - x_{it})(y_{ir} - y_{rt}) &= \lambda_{rst}(y_{ir} - y_{it})(x_{is} - x_{ts}) \\
(y_{is} - y_{it})(x_{ir} - x_{rt}) &= \lambda_{rst}(x_{ir} - x_{it})(y_{is} - y_{ts}) \\
(y_{is} - y_{it})(y_{ir} - y_{rt}) &= \lambda_{rst}(y_{ir} - y_{it})(y_{is} - y_{ts})
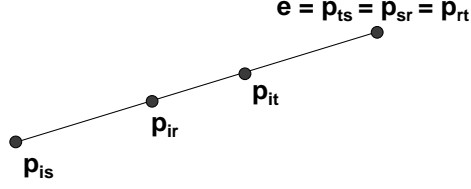\end{aligned}
\tag{9}
$$

Fig. 5. The Epipolar Lines Collapse

Note that these three view constraints are actually only *bilinear* in the image locations of the scene point (as opposed to the trilinear constraints in [15])). This is because by considering the projection of the points on the reference plane itself, we eliminate the homographies induced between the views (which appear in [16]).

The trilinear forms given in Equation 9 can be unified into a single tensor equation in a manner analogous to [17]:

$$(s^\alpha p_{it})(r^\beta p_{rt}) - \lambda_{rst}(r^\beta p_{it})(s^\alpha p_{ts}) = 0 \tag{10}$$

where

$$s = \begin{bmatrix} -1 & 0 & x_{is} \\ 0 & -1 & y_{is} \end{bmatrix}, \quad r = \begin{bmatrix} -1 & 0 & x_{ir} \\ 0 & -1 & y_{ir} \end{bmatrix}$$

and $\alpha, \beta = 1, 2$ indicate the row indices of s and r (e.g., $s^1 = [-1 \ 0 \ x_{is}])^4$.

Based on further algebraic manipulation, Equation 10 can be rewritten as:

$$\forall \alpha, \beta = 1, 2 \quad 0 = \sum_{a=1}^{3}\sum_{b=1}^{3}\sum_{c=1}^{3} (p_{it})_a (r^\beta)_b (s^\alpha)_c \left((p_{rt})_b \delta_{ac} - \lambda_{rst}(p_{ts})_c \delta_{ab} \right)$$

$$= \sum_{a=1}^{3}\sum_{b=1}^{3}\sum_{c=1}^{3} (p_{it})_a (r^\beta)_b (s^\alpha)_c (T^{rst})_{abc} \tag{11}$$

where δ follows the standard definition: $\delta_{pq} = 1$ if $p = q$ and 0 otherwise. T^{rst} is $3 \times 3 \times 3$ tensor

$$(T^{rst})_{abc} = \left((p_{rt})_b \delta_{ac} - \lambda_{rst}(p_{ts})_c \delta_{ab} \right)$$

In the above equations, $(p_{it})_1, (p_{it})_2, (p_{it})_3$, etc. denote the first (i.e., x), the second (i.e., y), and the third (i.e, 1) components of p_{it}, etc. Similarly $(T^{rst})_{abc}$ denotes the entry indexed by a, b, c in the Tensor.

---

[4] Note that as in [17], s^1 is the vertical line on Π passing through p_{is} and s^2 is the horizontal line on Π passing through p_{is}. Similarly r^1 and r^2 are the vertical and horizontal lines on Π passing through p_{ir}. Also, as in [17] the relationships in Equation 10 are valid for any line passing through p_{is} and any other line passing through p_{ir}. In other words, Equation 10 captures the same point-line-line relationship described in [17] and [5].

Note that the elements of T^{rst} depend on the two epipoles p_{rt} and p_{ts} and λ_{rst}. This is in contrast to the general form of the trifocal tensor – for example, the trilinear tensor in [15] also depends on the homographies due to the plane Π between the different cameras and the tensor described in [5] which depends on the camera projection matrices. As in the case of the Fundamental Matrix F in our formulation, the epipoles are explicit within the Tensor T, whereas in the general formulation, the tensor is implicitly related to the epipole. Given the Tensor T^{rst} we can recover the two epipoles p_{rt} and p_{ts} and the trifocal-ration λ_{rst}; using Equation 6 we can recover the third epipole p_{sr}.

4.4 Duals of the Three View Constraints

3 Scene Points + 1 Camera: As in the case of two-view analysis, the duality between scene points and camera centers also applies to three-view analysis. By switching the roles of scene points and camera centers in Figure 3 (i.e., $P_i \to C_t, C_t \to P_i, C_s \to P_j, C_r \to P_k$) we can derive new constraints involving one camera center and three points. The resulting geometric configuration is also a *complete quadrilateral*, but with a different labeling of the points. Figure 6a indicates the labeling corresponding to *one view of three points*. In this case the *dual*-trifocal-line contains the *dual-epipoles* p_{ij}, p_{jk}, and p_{ki}. The three-view constraint given in Equation 5 is replaced by

$$\frac{||p_{jt} - p_{it}||}{||p_{jt} - p_{ij}||} = \lambda_{ijk} \frac{||p_{kt} - p_{it}||}{||p_{kt} - p_{ki}||} \tag{12}$$

where $\lambda_{ijk} = \frac{(H_j - H_i)}{H_j} \frac{H_k}{(H_k - H_i)}$, is the *dual* to the *trifocal-ratio* λ_{rst}. Dual to the other forms of the three-view constraints, (e.g., Equation 9) can also be obtained by the same substitution of indices (i.e, $i \to t, t \to i, s \to j, r \to k$), leading to the *dual*-tensor form:

$$(T^{ijk})_{abc} = ((p_{ki})_b \delta_{ac} - \lambda_{ijk}(p_{ij})_c \delta_{ab}))$$

and the corresponding constraint set:

$$\forall \alpha, \beta = 1, 2 \quad 0 = \sum_{a=1}^{3} \sum_{b=1}^{3} \sum_{c=1}^{3} (p_{it})_a (k^\beta)_b (j^\alpha)_c (T^{ijk})_{abc} \tag{13}$$

where the definitions of the 2×3 matrices k and s are analogous to the definitions of r and s given earlier. Note that the the three-view constraints and their dual, the three-point constraints are completely symmetric. The dual-tensor depends on the *dual-epipoles* and the *dual* to the *trifocal-ratio*,

Other Combinations of 3+1 Points: The complete symmetry between scene points and camera centers implies that we can arbitrarily choose the label (either as a scene point or as a camera center) for each of the four 3D points in Figure 3. So far, we have considered two choices: 3 camera center + 1 scene point,
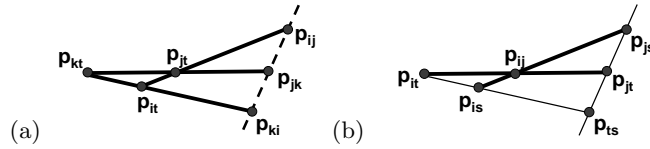
Fig. 6. Duals to the Three View Geometry: (a) the complete quadrilateral formed by 3 points + 1 camera center. (b) the quadrilateral formed by 2 points + 2 cameras. Note that the epipolar-lines (thin lines) intersect at an *epipole*, the dual-epipolar lines (thick lines) intersect at a *dual-epipole*, and an epipolar line intersects a dual-epipolar line at an *image-point*.

and 3 scene points + 1 camera center. The basic structure is that four points are divided into a group of 3 and a single point. We can obtain other duals by choosing the four points to consist of 2 camera centers + 2 scene points and grouping them as 2 camera centers and a scene point + 1 scene point or as 2 scene points and a camera center + 1 camera center.

In Figure 6b we show the resulting quadrilateral corresponding to the first of these groupings. Since the configuration shown in this figure is based on 2 camera centers and 2 scene points, the six points on the quadrilateral consist of four image points, one epipole, and one dual-epipole. Note that the two epipolar lines intersect at an *epipole*, the two dual-epipolar lines intersect at a *dual-epipole*, and each epipolar line intersects each dual-epipolar line at an image point.

Unlike the 3D world, where there are two types of points, camera centers and scene points, on the reference-plane, there are three-types of points – epipoles, dual-epipoles, and image points. Each of these form the center of a radial field of lines that go through that point, all three have completely dual-roles on Π.

5 Applications

The simple and intuitive expressions derived in the reference-plane based formulation in this paper lead to useful applications in 3D scene analysis. In particular, the simpler (bilinear) tri-focal constraints with the identified tri-focal ratio lead to a simple method for New View Synthesis. Initial experimental results are shown in this section. Moreover, the separation and compact packing of the *unknown* camera calibration and orientation into the 2D projection transformation (a homography) that relates the image plane to the reference plane, leads to potentially powerful reconstruction and calibration algorithms. These are briefly discussed in this section.

5.1 New View Generation Using the Three-View Constraints

In this section we show that the reference-plane based formulation provides a simple and intuitive way to generate new views from a given set of views.

We first show some results, followed by an explanation how they were obtained.

Figure 7a and 7b display two images taken by a hand-held camera. The scene contained toys which were placed on a rug on the floor. The camera translated and rotated between the two views. The 3D parallax effects due to the camera translation are apparent in the change of the 2D distance (on the image) between the clown's hat and the upper-right corner of the rug.

Figure 7c is a new *synthesized* view of the scene, as if obtained from a *virtual* camera positioned farther to the left of the scene relative to the two original views (and rotated, to compensate for the translation). Note the smaller distance between the clown's hat and the corner of the rug. For comparison and verification, Figure 7d shows an *actual* view obtained from the same viewing direction and orientation. Also, note the differences between the actual and synthesized view. There are image distortions where the flow was inaccurate (at depth discontinuities, e.g., on the rug around the clowns head, and near the ears of the smaller doll). Also, the synthesized view is missing the left part of the rug, as this portion of the rug was not viewed in any of the 2 input images.
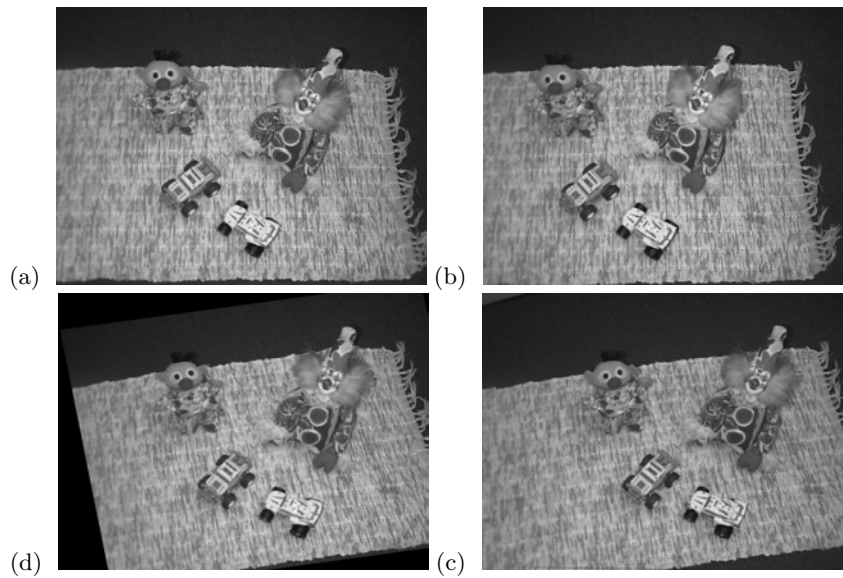

(a)   (b)

(d)   (c)

Fig. 7. *New View Synthesis.*
(a) and (b) show two images taken by a hand-held camera. The camera translated and rotated between the two views. The 3D parallax effects due to the camera translation are apparent by the change in the 2D distance (on the image) between the clown's hat and the upper-right corner of the rug. (c) A new synthesized *view of the scene, Note the smaller distance between the clown's hat and the corner of the rug. (d) an* actual *view obtained from the same viewing direction and orientation. Note the differences between the actual and synthesized view: There are image distortions where the flow was inaccurate (e.g., on the rug around the clowns head, and near the ears of the smaller doll). Also, the synthesized view is missing the left part of the rug, as this portion of the rug was not viewed in any of the 2 input images, (a and b).*

Below is a brief description of how the synthesized view was generated. To work directly with quantities on the reference plane Π would require partial calibration information about the input views. But as explained below, new view synthesis is possible even without such information.

Step I: One of the two input images (camera "s") is first *warped* towards the other input image (camera "t"; the *reference image*) via a 2D projective transformation to align the *images* of the plane Π in the two input image s and t. (Π is the plane of the rug, in our case). The corresponding 2D projective transformation is computed *automatically*, without any prior or additional information, using a 2D registration technique described in [10]. This method locks onto a *dominant* 2D parametric transformation between a pair of images, even in the presence of moving objects or other outliers (such as the toys, in our case). For more details see [10].

Note that after such 2D warping, the two plane-stabilized images are in full alignment in all image regions which correspond to the rug, and are misaligned in all other (i.e., out-of-plane) image points (i.e., the toys). The farther a scene point is from the planar surface (rug), the larger its residual misalignment. We refer to these as planar-parallax displacements (see [11, 12, 9]).

Note that the plane-stabilized sequence is in fact a 2D re-projection of the corresponding "virtual images" on the reference plane Π onto the reference image plane, t (See Figure 1.b). Therefore, a "quadrilateral" on Π will project to a "quadrilateral" on the image plane; different triangles on Π corresponding to different scene points and sharing a common tri-focal line will preserve this relation on the reference image plane t. It can be shown that for any quadrilateral, there exists some λ'_{rst} such that Equation (7) holds. In fact, it can be shown that

$$\lambda'_{rst} = \frac{T_Z{}^{ts}}{H_s} \frac{H_r}{T_Z{}^{tr}}, \qquad (14)$$

where $T_Z{}^{ts}$ is the component of the translation between cameras t and s along the optical (Z) axis of the reference camera t. Similarly $T_Z{}^{tr}$ for the third camera r. H_s and H_r are as before (i.e., heights relative to Π).

Step II: Dense flow is estimated between the two plane-stabilized images (using the method described in [12]). Note that after plane stabilization, the flow field between the two images reduces to a *radial epipolar field* centered at the epipole (see Equation (3); see also [11, 12, 9]). The cancellation of the plane homography removes all effects of camera rotation and changes in calibration. This allows to compute the flow field between a plane-stabilized image pair more reliably than general flow, as it is constrained to satisfy a global epipolar constraint.

Step III: We estimate the epipole (p_{ts}) from the radial flow field between the two input plane-stabilized images.

We then specify: (i) the *virtual* epipole (e.g., p_{rt}) between the reference image and the *virtual* "plane-stabilized" image, (ii) a *virtual* tri-focal ratio λ'_{rst} in the

reference frame. Given the virtual tri-focal ratio λ'_{rst}, the virtual epipole p_{rt}, the actual epipole p_{ts}, and the dense flow field between the two plane-stabilized images (between p_{it}'s and corresponding p_{is}'s), we can estimate all image points in the virtual (plane-stabilized) image (namely, all p_{ir}'s) using Equation 8.

The virtual tri-focal ratio λ'_{rst} and the virtual epipole p_{rt} can either be specified directly (e.g., via Equation (14)), or else by specifying the location of two or more image points in the virtual (plane-stabilized) view, and estimate them accordingly.

Step IV: Note that the synthesized *plane-stabilized* image is the same for *any* camera centered at C_r. In other words, it is independent of the internal parameters and the orientation of that camera. By specifying a homography that relates the image plane of the virtual camera to the stabilized image from the reference view, we have the complete flexibility to generate an image obtained by *any* camera situated at C_r. This is done by *unwarping* the synthesized plane-stabilized image via the corresponding 2D projective transformation.

5.2   3D Reconstruction and Camera Calibration

Given uncalibrated images, any approach for obtaining Euclidean (or Affine) reconstruction requires some type of calibration. One of the benefits of our approach is that this process is factored into two separate stages, each of which has a simple and intuitive solution. First, given the input images, the "virtual-images" on Π must be determined. This can be done by taking advantage of the P+P method[12]–(i) determine the planar homography for Π *between* an arbitrarily chosen reference image and each other image, and (ii) determine the homography between the reference image and Π. Note that the parallax Equation 3 is valid even if the image locations on Π are known only upto a 2D *affine* transformation. This means that just by indicating two sets of parallel lines (that are in different orientations) Π, the 3D Heights relative to the reference plane can be recovered. From this information, the parallax magnitude γ (in Equation 3) can be determined. (Note that by specifying the 2D coordinates of four points on the reference plane, the homography can fully determined, leading to Euclidean reconstruction.)

Given γ and the height of one 3D scene point relative to Π the Heights of all other points can be determined upto a global scale factor. Both these calibration steps are simple and intuitive and require minimal specification of information. The resulting reconstruction is with respect to the reference plane Π and does not involve the camera reference frames.

The foregoing outline for a reconstruction method assumes that the correspondences of each point across the multiple-views can be estimated. This involves computing the parallax flow-field(s), and the epipole(s)– these can be done in the same manner as described in [11, 12]. It is worth noting, however, the removal of the planar homography allows the parallax computation to be more robust and accurate [11, 12].

References

1. S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1034–1040, San-Juan, June 1997.
2. S. Carlsson. Duality of reconstruction and positioning from projective views. In *Workshop on Representations of Visual Scenes*, 1995.
3. H.S.M Coxeter, editor. *Projective Geometry*. Springer Verlag, 1987.
4. O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *European Conference on Computer Vision*, pages 563–578, Santa Margarita Ligure, May 1992.
5. O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *International Conference on Computer Vision*, pages 951–956, Cambridge, MA, June 1995.
6. Olivier Faugeras. *Three-Dimensional Computer Vision – A Geometric Viewpoint*. MIT Press, Cambridge, MA, 1996.
7. Richard Hartley. Lines and poins in three views – a unified approach. In *DARPA Image Understanding Workshop Proceedings*, 1994.
8. Richard Hartley. Euclidean Reconstruction from Uncalibrated Views. In *Applications of Invariance in Computer Vision*, J.L. Mundy, D. Forsyth, and A. Zisserman (Eds.), Springer-Verlag, 1993.
9. M. Irani and P. Anandan. Parallax geometry of pairs of points for 3d scene analysis. In *European Conference on Computer Vision*, Cambridge, UK, April 1996.
10. M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, January 1994.
11. M. Irani, B. Rousso, and P. peleg. Recovery of ego-motion using region alignment. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(3):268–272, March 1997.
12. R. Kumar, P. Anandan, and K. Hanna. Direct recovery of shape from multiple views: a parallax based approach. In *Proc 12th ICPR*, 1994.
13. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
14. R. Mohr. Accurate Projective Reconstruction In *Applications of Invariance in Computer Vision*, J.L. Mundy, D. Forsyth, and A. Zisserman, (Eds.), Springer-Verlag, 1993.
15. A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:779–789, 1995.
16. A. Shashua and N. Navab. Relative affine structure: Theory and application to 3d reconstruction from perspective views. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 483–489, Seattle, Wa., June 1994.
17. A. Shashua and P. Ananadan Trilinear Constraints revisited: generalized trilinear constraints and the tensor brightness constraint. IUW, Feb. 1996.
18. P.H.S. Torr. Motion Segmentation and Outlier Detection. *PhD Thesis: Report No. OUEL 1987/93*, Univ. of Oxford, UK, 1993.
19. M. Spetsakis and J. Aloimonos. A unified theory of structure from motion. *DARPA Image Understanding Workshop*, pp.271-283, Pittsburgh, PA, 1990 .
20. D. Weinshall, M.Werman, and A. Shashua. Shape descriptors: Bilinear, trilinear and quadlinear relations for multi-point geometry, and linear projective reconstruction algorithms. In *Workshop on Representations of Visual Scenes*, 1995.