Parallax Geometry of Pairs of Points for 3D Scene Analysis[1]

Michal Irani P. Anandan

David Sarnoff Research Center
CN5300, Princeton, NJ 08543-5300, U.S.A.
Email: {michal,anandan}@sarnoff.com

We present a geometric relationship between the image motion of *pairs* of points over multiple frames. This relationship is based on the *parallax* displacements of points with respect to an arbitrary planar surface, and does not involve epipolar geometry. A constraint is derived over two frames for any pair of points, relating their projective structure (with respect to the plane) based only on their image coordinates and their parallax displacements. Similarly, a $3D$-rigidity constraint between pairs of points over multiple frames is derived. This leads to the derivation of an image point which is dual to the epipole. We show applications of these parallax-based constraints to solving three important problems in $3D$ scene analysis: (i) the recovery of $3D$ scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views. Moreover, we show that this approach can handle difficult situations for $3D$ scene analysis, e.g., where there is only a small set of parallax vectors, and in the presence of independently moving objects.

KEYWORDS:

Planar parallax motion, Moving object detection, 3D scene recovery, Rigidity constraints, Multi-frame analysis, Duality of the epipole.

[1]This work was supported by ARPA under contract DAAA15-93-C-0061

1 Introduction

The analysis of three dimensional scenes from image sequences has a number of goals. These include (but are not limited to): (i) the recovery of 3D scene structure, (ii) the detection of moving objects in the presence of camera induced motion, and (iii) the synthesis of new camera views based on a given set of views. The traditional approach to these types of problems has been to first recover the epipolar geometry between pairs of frames and then apply that information to achieve the abovementioned goals. However, this approach is plagued with the difficulties associated with recovering the epipolar geometry [34].

Recent approaches to 3D scene analysis have overcome some of the difficulties in recovering the epipolar geometry by decomposing the motion into a combination of a planar homography and residual parallax [17, 21, 24]. The residual parallax displacements depend on the *projective structure* (see Section 2 for a definition), and on the translation between the camera origins. While these methods remove some of the ambiguities in estimating camera rotation, they still require the explicit estimation of the epipole itself, which can be difficult under many circumstances. In particular, epipole estimation is ill-conditioned when the epipole lies significantly away from the center of the image and the parallax motion vectors are nearly parallel to each other. Also, when there is only a small number of parallax vectors and the scene contains moving objects, these objects will incorrectly influence the estimation of the epipole.

More recently, progress has been made towards deriving constraints directly based on collections of points in multiple views. Examples are the trilinearity constraints [28, 23] which eliminate the scene structure in favor of the camera geometries; the dual-shape tensor [33] which eliminates the camera motion in favor of scene structure; the more general framework of multipoint multiview geometry [11, 5]; and the work on multiple view invariants without requiring the recovery of the epipolar geometry [34]. In general, the treatment of multipoint geometry assumes that the scene is static and relies on the fact that almost all points selected for the shape estimation are known to belong to a single rigid body. In its current form, this class of methods does not address the problem of shape recovery in dynamic scenes, in particular when the amount of image motion due to independent moving object is not negligible.

In this paper we develop geometric relationships between the residual (planar) parallax displacements of *pairs* of points. These geometric relationships address the problem of 3D scene analysis even in *difficult* conditions, i.e., when the epipole estimation is ill-conditioned, when there is a small number of parallax vectors, and in the presence of moving objects. We show how these relationships can be applied to each of the three problems outlined at the beginning of this section. Moreover, the use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and the "$3D$ algorithms" for each of the problems mentioned above.

The geometric relationships presented in this work are expressed in terms of *residual parallax displacements* of points after canceling a *planar homography*[2]. Techniques for estimating the planar homography from pairs of images, even in the presence of moving objects, were previously described in [1, 14, 15], and were employed in the experimental part of this work.

In Section 2 a *parallax-based structure constraint* is derived, which relates the *projective structure* of two points to their image positions and their *parallax displacements* alone. By eliminating the relative projective structure of a pair of points between *three* frames, we arrive at a constraint on the parallax displacements of two points *moving as a rigid object* over those frames. We refer to this as the *multi-frame rigidity constraint.* Similarly a *multi-point rigidity constraint* over two frames is derived.

In Section 3 a geometric interpretation of the the *multi-frame rigidity constraint* is presented. This leads to the derivation of a dual point to the epipole. Further, the *multi-frame rigidity constraint* is shown to be dual to the *epipolar constraint.*

In Section 4 the pairwise parallax-based constraints are applied to solving the three abovementioned problems in $3D$ scene analysis, even in difficult situations (i.e., when the epipole estimation is ill-conditioned, when there is a small number of parallax vectors, and in the presence of moving objects), assuming uncalibrated cameras.

In Section 5 the pairwise parallax constraints are generalized to full image motion by including the planar homography component.

2 Parallax-Based Constraints on Pairs of Points

In this section we derive a constraint on the parallax motion of pairs of points between two frames. We show how this constraint can be used to recover relative $3D$ structure of two points from their parallax vectors alone, without any additional information, and in particular, without requiring the recovery of the camera epipoles. The parallax constraint is then extended to *multiple* frames and to multiple points to obtain *rigidity* constraints on image points based on their parallax displacements alone, without involving any scene or camera geometry.

[2]The decomposition of image motion into a *homography + parallax* contains, as a special case, the decomposition of image motion into *rotation + translation*, and is therefore more general [16, 17, 21, 24]

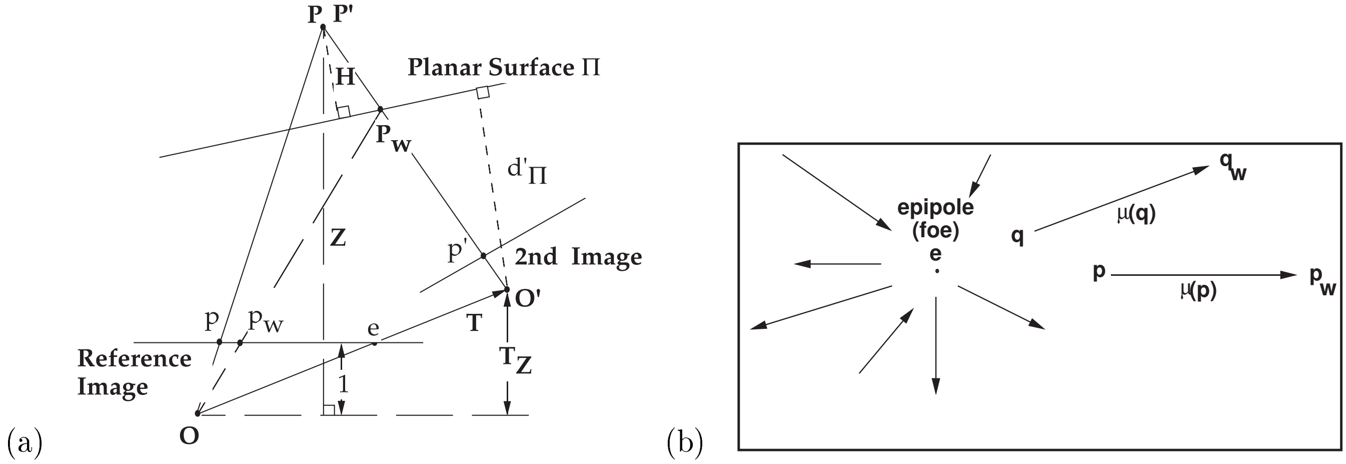

Figure 1: The plane+parallax decomposition. (a) The geometric interpretation of the planar homography + parallax decomposition explained in Equation 1. (b) The epipolar field of the residual parallax displacements.

2.1 The Planar Parallax Decomposition

To derive the parallax constraint, first we describe the decomposition of the image motion into a homography (i.e., the image motion of an arbitrary planar surface) and residual parallax displacements. This decomposition has been previously derived and used [17, 21, 24]. A detailed rederivation of this decomposition can be found in the appendix of this paper.

There are several motivations for using the *homography + parallax* decomposition of image motion: This decomposition is more general than the traditional decomposition of image motion into rotational and translational components, as the rotation is a special type of a homography (of a plane at infinity). Moreover, there are robust and efficient algorithms for obtaining this decomposition using $2D$ techniques. These techniques [1, 14, 15] can estimate the planar homography from pairs of image frames even in the presence of moving objects. These techniques are briefly described in Section 4.1.

Figure 1 provides a geometric interpretation of the planar parallax. Let $\vec{P} = (X, Y, Z)^T$ and $\vec{P'} = (X', Y', Z')^T$ denote the Cartesian coordinates of a scene point with respect to two different camera views, respectively. Let the 3×3 matrix R and the 3×1 vector T denote the rotation and translation between the two camera systems respectively.

Let $\vec{p} = (x, y, 1)^T = \frac{1}{Z} K \vec{P}$ and $\vec{p'} = (x', y', 1)^T = \frac{1}{Z'} K' \vec{P'}$ denote the images of the scene point P in the two camera views as expressed in homogeneous coordinates. K and K' are 3×3 matrices representing the internal calibration parameters of the two cameras (see appendix). Also, define

$\vec{t} = (t_x, t_y, t_z)^T = K\vec{T}$. Note that $(K\vec{P})_z = Z$, $(K'\vec{P'})_z = Z'$, and $t_z = T_Z$.

Let Π be an arbitrary planar surface and A' denote the homography that aligns the planar surface Π between the second and first frame (i.e., for all points $\vec{P} \in S$, $\vec{P} = A'\vec{P'}$).

It can be shown (see appendix, as well as [17, 21, 24]) that when $T_z \neq 0$ the image motion can be written (in homogeneous coordinates) as:

$$\vec{p} = \vec{p_w} + \gamma \frac{T_z}{d'_\pi}(\vec{p_w} - \vec{e}), \tag{1}$$

where $\vec{p_w}$ is the image point in the first frame which results from warping the corresponding point $\vec{p'}$ in the second frame by the homography A', i.e., $\vec{p_w} = \frac{A'\vec{p'}}{{a'}_3^T\vec{p'}}$ (${a'}_3$ being the third row of A'). When $\vec{P}$ is on the $3D$ planar surface Π, then $\vec{p_w} = \vec{p}$. Otherwise, the remaining displacement between $\vec{p_w}$ and $\vec{p}$ is proportional to the $3D$ *projective structure* (γ) of $\vec{P}$ with respect to the planar surface Π: $\gamma = \frac{H}{Z}$, where H is the distance of $\vec{P}$ from the plane Π. d'_π is the distance to the plane Π from the second camera center (see Figure 1). $\vec{e} = \frac{1}{T_Z}\vec{t}$ is the epipole (or the *focus-of-expansion*, FOE) in the first frame.

When $T_Z = 0$, Eq. (1) can be rewritten in terms of $\vec{p_w}$ as:

$$\vec{p} = \vec{p_w} - \frac{\gamma}{d'_\pi}\vec{t} \tag{2}$$

.

Rewriting Eq. (1) in the form of image *displacements* yields (in homogeneous coordinates):

$$\vec{p'} - \vec{p} = (\vec{p'} - \vec{p_w}) - \gamma \frac{T_Z}{d'_\pi}(\vec{p_w} - \vec{e}). \tag{3}$$

Define $\vec{u} = \vec{p'} - \vec{p} = (u_x, u_y, 0)^T$, where $(u_x, u_y)^T$ is the measureable 2D image displacement vector of the image point $\vec{p}$ between the two frames. Similarly, define $\vec{u_\pi} = \vec{p'} - \vec{p_w} = (u_{\pi x}, u_{\pi y}, 0)^T$ and $\vec{\mu} = -\gamma \frac{T_Z}{d'_\pi}(\vec{p_w} - \vec{e}) = (\mu_x, \mu_y, 0)^T$. Hence,

$$\vec{u} = \vec{u_\pi} + \vec{\mu}, \tag{4}$$

$\vec{u_\pi}$ denotes the planar part of the $2D$ image displacement (i.e., the homography due to Π), and $\vec{\mu}$ denotes the residual *parallax* $2D$ displacement.

When $T_z = 0$ (Eq. (2)): $\vec{\mu} = \frac{\gamma}{d'_\pi}\vec{t}$.

Eq. (4) provides the form of parallax notation that will be used in the remaining of the paper. Although so far we destinguished between the cases of $T_z = 0$ and $T_z \neq 0$, they will be unified and treated as a single case in the following sections.

2.2 The Parallax Based Structure Constraint

.

Let $\vec{p_1} = (x_1, y_1, 1)$ and $\vec{p_2} = (x_2, y_2, 1)$ be two image points that belong to the static background. Let $\vec{\mu_1}$ and $\vec{\mu_2}$ be the planar-parallax displacement vectors of those two points.

$$\vec{\mu_1} = \gamma_1 \frac{T_z}{d'_\pi}(\vec{e} - \vec{p_{w_1}}) \tag{5}$$

$$\vec{\mu_2} = \gamma_1 \frac{T_z}{d'_\pi}(\vec{e} - \vec{p_{w_2}}) \tag{6}$$

Therefore,

$$\vec{\mu_1}\gamma_2 - \vec{\mu_2}\gamma_1 = \gamma_1\gamma_2 \frac{T_Z}{d'}(\vec{p_{w_2}} - \vec{p_{w_1}}) \tag{7}$$

This last step eliminated the epipole $\vec{e}$. Eq. (7) entails that the vectors on both sides of the equation are parallel. Since $\gamma_1\gamma_2\frac{T_Z}{d'}$ is a scalar, we get:

$$(\vec{\mu_1}\gamma_2 - \vec{\mu_2}\gamma_1) \parallel \Delta\vec{p}_w,$$

where $\Delta\vec{p}_w = (\vec{p_{w_2}} - \vec{p_{w_1}})$. This leads to the *pairwise parallax constraint*:

$$(\vec{\mu_1}\gamma_2 - \vec{\mu_2}\gamma_1)^T (\Delta\vec{p}_w)_\perp = 0, \tag{8}$$

where $\vec{v}_\perp$ signifies a vector perpendicular to $\vec{v}$.

When $T_Z = 0$, a constraint stronger than Eq. (8) can be derived:

$$(\vec{\mu_1}\gamma_2 - \vec{\mu_2}\gamma_1) = 0.$$

However, Eq. (8), still holds. This is important, as we do not have a-priori knowledge of T_Z to distinguish between the two cases.

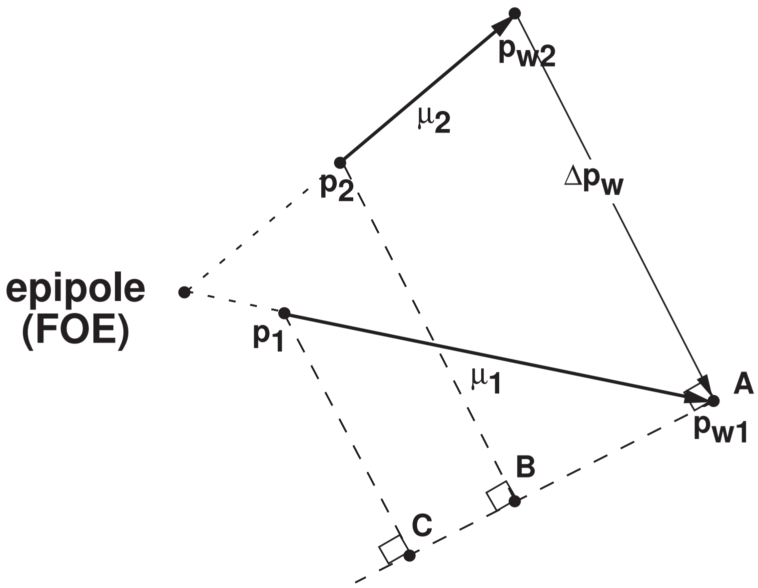

Figure 2: *The relative structure constraint.*
This figure geometrically illustrates the relative structure constraint (Eq. 9): $\frac{\gamma_2}{\gamma_1} = \frac{\vec{\mu_2}^T(\Delta\vec{p_w})_\perp}{\vec{\mu_1}^T(\Delta\vec{p_w})_\perp} = \frac{AB}{AC}$.

From Eq. (8) we can easily derive:

$$\frac{\gamma_2}{\gamma_1} = \frac{\vec{\mu_2}^T(\Delta\vec{p_w})_\perp}{\vec{\mu_1}^T(\Delta\vec{p_w})_\perp}. \tag{9}$$

Figure 2 displays the constraint geometrically.

The fact that relative structure of one point with respect to another can be obtained using only the two parallax vectors is not surprising: *In principle*, one could use the two parallax vectors to recover the epipole (the intersection point of the two vectors), and then use the distances of the points from the computed epipole and the magnitudes of their parallax displacements to estimate their relative projective structure. The benefit of the constraint (9) is that it provides this information *directly* from the positions and parallax vectors of the two points, without the need to go through the computation of the epipole, using as much information as one point can give on another. Fig. 3 graphically shows an example of a configuration in which estimating the epipole is very unreliable, whereas estimating the relative structure *directly* from Eq. (9) *is* reliable.

2.3 The Parallax-Based Rigidity Constraint

In this section we extend the parallax-based structure constraint to *multiple frames* and to *multiple points*, to obtain *rigidity* constraints that are based only on parallax displacements of the image

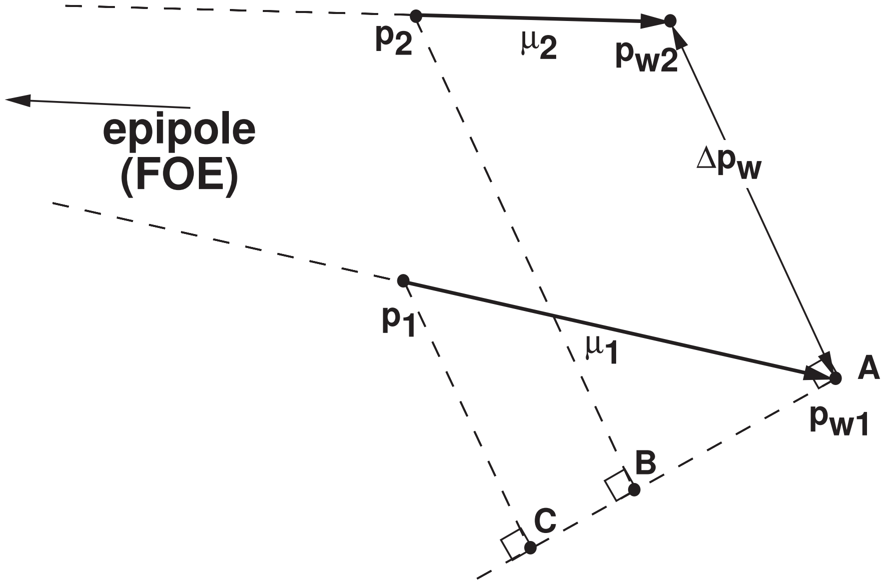

Figure 3: *Reliable structure recovery despite ambiguous epipole.*
When the parallax vectors are nearly parallel, the epipole estimation is unreliable. However, the relative structure constraint $\frac{\gamma_2}{\gamma_1} = \frac{AB}{AC}$ (Eq. 9) can be reliably used to recover relative structure in this case.

points, and involve neither *structure* parameters nor *camera geometry*.

Rigidity Over Multiple Frames:

Let $\vec{p_1} = (x_1, y_1, 1)$ and $\vec{p_2} = (x_2, y_2, 1)$ be two image points in the first (reference) frame. Let $\vec{\mu_i}^j$ denote the $2D$ planar parallax displacement of $\vec{p_i}$ $(i = 1, 2)$ from the first frame to frame j.

Using the structure invariance constraint (9), for any two frames j and k we get:

$$\frac{\gamma_2}{\gamma_1} = \frac{\vec{\mu_2}^{j^T} (\Delta\vec{p_w})^j_\perp}{\vec{\mu_1}^{j^T} (\Delta\vec{p_w})^j_\perp} = \frac{\vec{\mu_2}^{k^T} (\Delta\vec{p_w})^k_\perp}{\vec{\mu_1}^{k^T} (\Delta\vec{p_w})^k_\perp}. \tag{10}$$

Multiplying by the denominators yields the rigidity constraint of the two points over three frames (the reference frame, frame j, and frame k):

$$(\vec{\mu_1}^{k^T} (\Delta\vec{p_w})^k_\perp)(\vec{\mu_2}^{j^T} (\Delta\vec{p_w})^j_\perp) - (\vec{\mu_1}^{j^T} (\Delta\vec{p_w})^j_\perp)(\vec{\mu_2}^{k^T} (\Delta\vec{p_w})^k_\perp) = 0. \tag{11}$$

Based on the planar parallax motion trajectory of a single image point (e.g., $\vec{p_1}$) over several frames, the rigidity constraint (11) states a constraint on the planar parallax motion trajectory of any other point (e.g., $\vec{p_2}$). The rigidity constraint (11) can therefore be applied to detect inconsistencies in the $3D$ motion of two image points (i.e., say whether the two image points are projections of $3D$ points belonging to a same or different $3D$ moving objects) based on their *parallax* motion among three (or more) frames alone, without the need to estimate either *camera geometry* or *structure* parameters.

In contrast to previous approaches (e.g., the trilinear tensor [23]), when planar parallax motion is available, Eq. (11) provides certain advantages: (i) it is based on the parallax motion of a *single* image point, (ii) it does not require any numerical estimation (e.g., unlike [23], it does not require estimation of tensor parameters), (iii) it does not involve, *explicitly or implicitly*, any shape or camera geometry information other than that already implicit in the planar parallax motion itself.

Rigidity Over Multiple Points:

Instead of considering pairs of points over multiple frames, we can consider multiple points over pairs of frames, to come up with a different form of the rigidity constraint.

Let $\vec{p_1} = (x_1, y_1, 1)$, $\vec{p_2} = (x_2, y_2, 1)$, and $\vec{p_3} = (x_3, y_3, 1)$ be three image points in the first (reference) frame. Let $\vec{\mu_i}$ denote the $2D$ planar parallax motion of $\vec{p_i}$ $(i = 1, 2, 3)$ from the first frame to any other frame. Also let $\Delta \vec{p_{w_{j,i}}} = \vec{p_{w_j}} - \vec{p_{w_i}}$.

Using the shape invariance constraint (9):

$$\frac{\gamma_2}{\gamma_1} = \frac{\vec{\mu_2}^T (\Delta \vec{p_{w_{2,1}}})_\perp}{\vec{\mu_1}^T (\Delta \vec{p_{w_{2,1}}})_\perp},$$

$$\frac{\gamma_3}{\gamma_2} = \frac{\vec{\mu_3}^T (\Delta \vec{p_{w_{3,2}}})_\perp}{\vec{\mu_2}^T (\Delta \vec{p_{w_{3,2}}})_\perp},$$

$$\frac{\gamma_3}{\gamma_1} = \frac{\vec{\mu_3}^T (\Delta \vec{p_{w_{3,1}}})_\perp}{\vec{\mu_1}^T (\Delta \vec{p_{w_{3,1}}})_\perp}.$$

Using the equality

$$\frac{\gamma_3}{\gamma_1} = \frac{\gamma_3}{\gamma_2}\frac{\gamma_2}{\gamma_1}$$

and multiplying by the denominators, we get the rigidity constraint for three points over a pair of frames:

$$(\vec{\mu_3}^T (\Delta \vec{p_{w_{3,2}}})_\perp)(\vec{\mu_2}^T (\Delta \vec{p_{w_{2,1}}})_\perp)(\vec{\mu_1}^T (\Delta \vec{p_{w_{3,1}}})_\perp) - (\vec{\mu_2}^T (\Delta \vec{p_{w_{3,2}}})_\perp)(\vec{\mu_1}^T (\Delta \vec{p_{w_{2,1}}})_\perp)(\vec{\mu_3}^T (\Delta \vec{p_{w_{3,1}}})_\perp) = 0 \tag{12}$$

The fact that three points in two frames form a rigidity constraint is not surprising: *In principle*, one could use two of the three parallax vectors to obtain the epipole (the intersection point of the two vectors). $3D$ rigidity will constrain the parallax vector of the third point to lie on the epipolar line emerging from the computed epipole through the third point. The benefit of the rigidity constraint (12) is in the fact that it provides this information directly from the positions and parallax vectors of the three points, without the need to go through the unstable computation of the epipole, using as much information as two point can give on the third.

3 The Dual of the Epipole

In this section, we present a different way of deriving the *parallax-based rigidity constraint* (see Section 2), this time geometrically rather than algebraically. This leads to a simple and intuitive geometric interpretation of the multiframe rigidity constraint, and to the derivation of a *dual point* to the epipole. Although we point out this distinct image point (the epipole dual), the rigidity constraint itself does *not* require the estimation of the dual epipole, just as it does not require the estimation of the epipole itself.

3.1 Multi-Frame Rigidity - A Geometric Interpretation

The 3D geometric structure associated with the planar parallax of pairs of points between two frames is illustrated in Figure 4. In this figure, Π is the planar surface, and P and Q are the two scene points. As in the case of Figure 1, P_w and Q_w are the intersections of rays O'P and O'Q with the plane Π. Similarly, the points $\vec{p_w}$ and $\vec{q_w}$ on the reference image are the projections of P_w and Q_w, and are therefore the points to which the planar homography transforms $\vec{p'}$ and $\vec{q'}$ respectively. In other words, $\vec{p_w} = \frac{A'\vec{p'}}{{a'}_3^T\vec{p'}}$ and $\vec{q_w} = \frac{A'\vec{q'}}{{a'}_3^T\vec{q'}}$. Below, we refer to $\vec{p_w}$ and $\vec{q_w}$ as "warped points".

Let R be the intersection of the line connecting the scene points P and Q with the plane Π. Note that the points P, Q, R, P_w and Q_w are co-planar. Hence, P_w, Q_w and R are colinear. Of course, P, Q and R are colinear by construction.

Let $\vec{r}$ be the projection of R on the reference image plane. Since $R \in \Pi$, $\vec{r} = \vec{r_w}$. Since $\vec{p}, \vec{q}, \vec{r}$ and $\vec{p_w}, \vec{q_w}, \vec{r}$ are the image projection of the colinear world points P, Q, R, we can infer the following: $\vec{p_w}, \vec{q_w}$ and $\vec{r}$ are colinear and $\vec{p}, \vec{q}$ and $\vec{r}$ are colinear.

In other words, the line connecting $\vec{p_w}$ and $\vec{q_w}$ and the line connecting $\vec{p}$ and $\vec{q}$ intersect at $\vec{r}$, the image of the point R. (See also [34] for the same observation.)

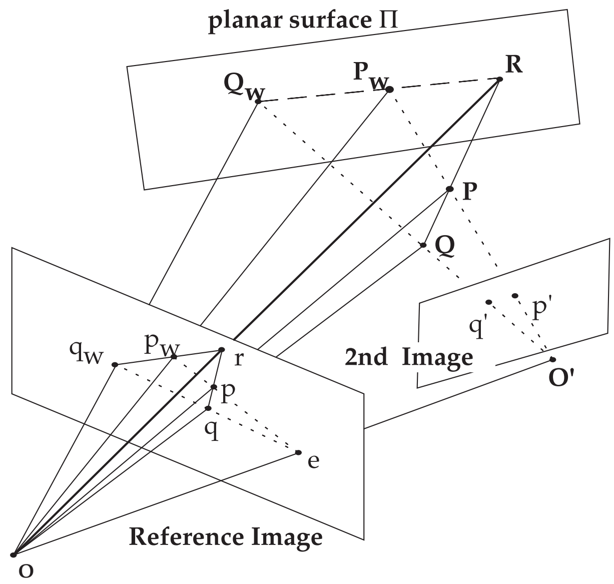

Figure 4: Parallax geometry.
The line connecting points p, q in the reference image intersects the line connecting the warped points p_w, q_w at r. The point r is the image of the point R which is the intersection of the line PQ with the planar surface Π.

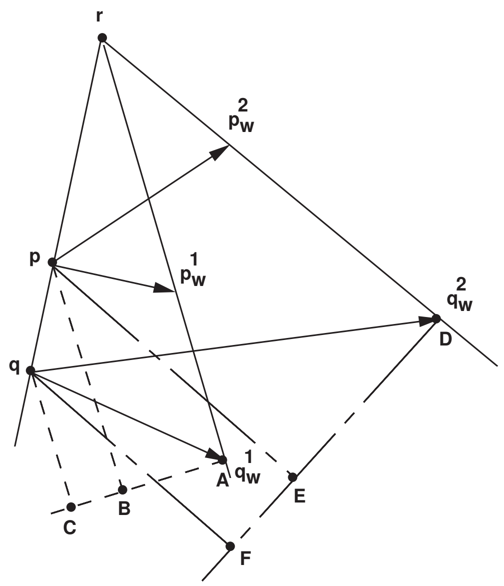

Figure 5: The dual of the epipole.
The line connecting points p, q in the reference image and the lines connecting the corresponding warped points p_w^1, q_w^1 and p_w^2, q_w^2 from other frames all intersect at r, the dual of the epipole.

Note that the point $\vec{r}$ does not depend on the second camera view. Therefore, if multiple views are considered, then the lines connecting the warped points $\vec{p_w^j} = \frac{A^j \vec{p^j}}{{a^j}_3^T \vec{p^j}}$ and $\vec{q_w^j} = \frac{A^j \vec{q^j}}{{a^j}_3^T \vec{q^j}}$ (for any frame j), meet at $\vec{r}$ for all such views.

The convergence of the lines is illustrated in Figure 5. Referring to that figure, since the lines qC, pB and rA are parallel to each other and intersect the lines qpr and CBA:

$$\frac{qr}{pr} = \frac{CA}{BA} = \frac{\gamma_q}{\gamma_p}.$$

Similarly,

$$\frac{qr}{pr} = \frac{FD}{ED} = \frac{\gamma_q}{\gamma_p}.$$

Hence

$$\frac{qr}{pr} = \frac{CA}{BA} = \frac{FD}{ED}.$$

This is the same as the multi-frame rigidity constraint derived in Section 2 (Eq. (10)). Note, however, that the rigidity constraint itself does not require the estimation of the point of convergence $\vec{r}$, just as it does not require the estimation of the epipole $\vec{e}$.

3.2 The Epipole Dual

The point $\vec{r}$ shown in Figures 4 and 5 has dual properties with respect to the epipole: It relates to a *pair of points* over *multiple frames* the same way the epipole relates to *multiple points* over a *pair of frames.* Moreover, just like the epipole is invariant with respect to scene structure, the point of convergence $\vec{r}$ is invariant with respect to camera motion. However, *unlike* the epipole, the point of convergence $\vec{r}$ depends on the choice of the planar surface Π, since the point of intersection R depends on that surface. To further understand the geometric meaning of the point $\vec{r}$, we examine two special cases of Π, with the corresponding physical interpretations of $\vec{r}$:

Case 1: Π = the Plane at Infinity

When Π coincides with the plane at infinity, the homography induced by Π reduces to the image motion due to camera rotation, i.e., the $homography+parallax$ decomposition reduces to the traditional $rotation+translation$ decomposition. In this case, the point of convergence $\vec{r}$ is the *vanishing point* of the line passing through the scene points P and Q (namely, the image of the intersection of the line with the plane at infinity).

Case 2: Π = the Image Plane

When Π coincides with the image plane, the point R is the intersection of the line connecting the scene points P and Q with the image plane. In this case, R and $\vec{r}$ coincide, and the point $\vec{r}$ is an *exact dual to the epipole.* The term "duality" is used here to refer to the symmetry between the geometry of multiple scene points and multiple camera views [33, 5]. The key properties of the duality are described below:

Epipole $\vec{e}$:	Dual Epipole $\vec{r}$:
1. The *epipole* is the point of intersection of the line connecting *two camera centers* with the image plane. *Every pair of camera centers defines an epipole in the image plane.*	1. The *dual epipole* is the point of intersection of the line connecting *two scene points* with the image plane. *Every pair of scene points defines a dual epipole in the image plane.*
2. The pencil of planes connecting the two camera centers with each scene point forms the set of *epipolar planes.*	2. The pencil of planes connecting the two scene points with each camera center forms the set of *dual epipolar planes.*
3. The intersection of the *pencil of epipolar planes* with the image plane forms the pencil of *epipolar lines.* These lines form a radial field emerging from the epipole $\vec{e}$.	3. The intersection of the *pencil of dual epipolar planes* with the image plane forms the pencil of *dual epipolar lines.* These lines form a radial field emerging from the dual epipole $\vec{r}$.

Note that the pencil of planes formed by the dual epipolar planes is the same pencil of planes observed in [5].

4 Applications of Pairwise Parallax Geometry

In this section we show how pairwise parallax geometry in its various forms, which was introduced in the previous sections, provides an approach to handling some well-known problems in $3D$ scene analysis, in particular: moving object detection, shape recovery, and new view generation.

An extensive literature exists on methods for solving the above mentioned problems. They can be roughly classified into two main categories:

- $\underline{2D \text{ methods}}$ (e.g.,[14, 15, 2, 3, 4, 6, 7, 26, 27, 20]): These methods assume that the image motion of the scene can be described using a $2D$ parametric transformation. They handle *dynamic* scenarios, but are limited to planar scenes or to very small camera translations. These fail in the presence of parallax motion.

- $\underline{3D \text{ methods}}$ (e.g., [12, 17, 21, 24, 10, 13, 33, 11, 5]): These methods handle general $3D$ scenes, but are (in their current form) limited to static scenarios or to scenes where the parallax is both dense and of significant magnitude in order to overcome "noise" due to moving objects.

The use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and "$3D$ algorithms". The need for bridging this gap exists in real image sequences, because it is not possible to predict in advance which situation will occur. Moreover, both types of scenarios can occur within the same sequence, with gradual transitions between them.

4.1 Estimating Planar Parallax Motion

The estimation of the planar parallax motion used for performing the experiments presented in this section was done using two successive computational steps: (i) $2D$ image alignment to compensate for a detected planar motion (i.e., the homography) in the form of a $2D$ parametric transformation, and, (ii) estimation of residual image displacements between the aligned images (i.e, parallax or independent motion).

We use previously developed methods [1, 14, 15] in order to compute the $2D$ *parametric* image motion of a $3D$ planar surface in the scene. These techniques lock onto a "dominant" planar motion in an image pair, even in the presence of other differently moving objects in the field of view. They do not require prior knowledge of the regions of support of the moving objects, nor that of the planar surface. This computation provides only the $2D$ motion parameters of the planar surface, but no explicit $3D$ shape or motion information.

Once such a $2D$ *parametric* image motion has been computed, it can be applied to the *entire* image plane to geometrically warp the second image frame towards the first (the reference frame). This registration aligns the image regions that correspond to scene parts that lie on the planar surface. The *residual* image displacements (misalignments) between the two *planar-registered* image frames are due to only one of the following: (i) *parallax* motion of static scene points that are not located on the planar surface (see Fig. 1 and Eq. (4)), or, (ii) motion of independent moving objects in the scene. The *residual* image displacements are estimated using the optical flow estimation technique described in [1].

This technique for computing parallax motion assumes the existence of an approximate planar surface in the scene (although the scene itself need *not* piecewise planar). Most indoor scenes have a planar surface (e.g., walls, floor, pictures, windows, etc.), and in outdoor scenes the ground or any large enough *distant* object can serve as a planar surface. Other methods have been suggested for computing the $2D$ image motion of a *virtual* planar surface [18, 25]. These do not requiring the existence of a *physical* planar surface in the scene.

4.2 Moving Object Detection

A number of techniques exist to handle multiple motions analysis in the simpler $2D$ case, where motions of independent moving objects can be modeled by $2D$ parametric transformation [14, 15, 2, 3, 4, 6, 7, 26, 27, 20]. These methods, however, detect static scene points with planar parallax motion as moving objects, since they have a different $2D$ image motion than the planar part of the background scene.

In the general $3D$ case, the moving object detection problem is much more complex, since it requires detecting $3D$ motion inconsistencies. Typically, this is done by recovering the epipolar geometry. However, trying to estimate epipolar geometry (i.e., camera motion) in the presence of multiple moving objects, with no prior segmentation, is a difficult problem. The problem becomes even more acute when there exists only sparse parallax information. A careful treatment of the issues and problems associated with moving object detection in 3D scenes is given in [29].

Methods have been proposed for recovering camera geometry in the presence of moving objects [19, 31] in cases when the available parallax was *dense* enough and the independent motion was *sparse* enough to be treated as *noise*.

Fig. 6.a graphically displays an example of a configuration in which estimating the epipole in presence of multiple moving objects can be very erroneous, even when using clustering techniques in the epipole domain as suggested by [19, 30]. Relying on the epipole computation to detect inconsistencies in $3D$ motion fails in such cases.

The parallax rigidity constraint(Equation (11)) can be applied to detect inconsistencies in the $3D$ motion of one image point relative to another directly from their "parallax" displacements over multiple (three or more) frames, without the need to estimate either *camera geometry* or *shape* parameters. This provides a useful mechanism for clustering (or segmenting) the "parallax" vectors (i.e., the residual motion after planar registration) into consistent groups belonging to consistently $3D$ moving objects, even in cases such as in Fig. 6.a, where the parallax information is minimal, and the independent motion is not negligible. Fig. 6.b graphically explains how the rigidity constraint (11) detects the $3D$ inconsistency of Fig. 6.a over three frames.

Fig. 7 shows an example of using the rigidity constraint (11) to detect $3D$ inconsistencies in a real image sequence. In this sequence the camera is in motion (translating from left to right), inducing parallax motion of different magnitudes on the house, road, and road-sign. The car moves independently from left to right. The detected $2D$ planar motion was that of the house. The planar parallax motion was computed after $2D$ registration of the three images with respect to

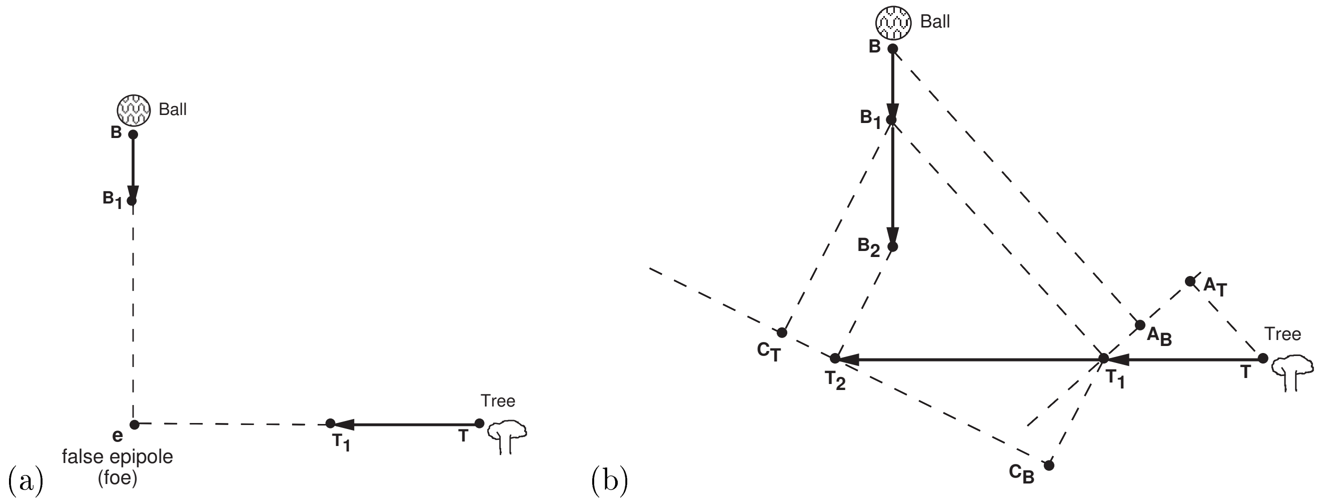

Figure 6: Reliable detection of 3D motion inconsistency with sparse parallax information.
(a) Camera is translating to the right. The only static object with pure parallax motion is the tree. Ball is falling independently. The epipole may be incorrectly be computed as e. The false epipole e is consistent with both motions.
(b) The rigidity constraint applied to this scenario detects 3D inconsistency over three frames, since $\frac{T_1 A_B}{T_1 A_T} \neq \frac{T_2 C_B}{-T_2 C_T}$. In this case, even the signs do not match.

the house (see Fig. 7.d). A single point on the road-sign was selected as a point of reference (see Fig. 7.e). Fig. 7.f displays the measure of *in*consistency of each point in the image with respect to the selected road-sign point. Bright regions indicate large values when applying constraint (11), i.e., large violations in $3D$ rigidity detected over three frames with respect to the road-sign point. The region which was detected as moving $3D$-*in*consistently with respect to the road-sign point corresponds to the car. Regions close to the image boundary were ignored. All other regions of the image were detected as moving $3D$-consistently with the road-sign point. Therefore, assuming an *uncalibrated* camera, this method provides a mechanism for segmenting all non-zero residual motion vectors (after $2D$ planar stabilization) into groups moving *consistently* (in the $3D$ sense).

Fig. 8 shows another example of using the rigidity constraint (11) to detect $3D$ inconsistencies. In this sequence the camera is mounted on a helicopter flying from left to right, inducing some parallax motion (of different magnitudes) on the house-roof and trees (bottom of the image), and on the electricity poles (by the road). Three cars move independently on the road. The detected $2D$ planar motion was that of the ground surface (see Fig. 8.d). A single point was selected on a tree as a point of reference (see Fig. 8.e). Fig. 8.f displays the measure of *in*consistency of each point in the image with respect to the selected reference point. Bright regions indicate $3D$-inconsistency detected over three frames. The three cars were detected as moving *in*consistently with the selected tree point. Regions close to the image boundary were ignored. All other image regions were detected

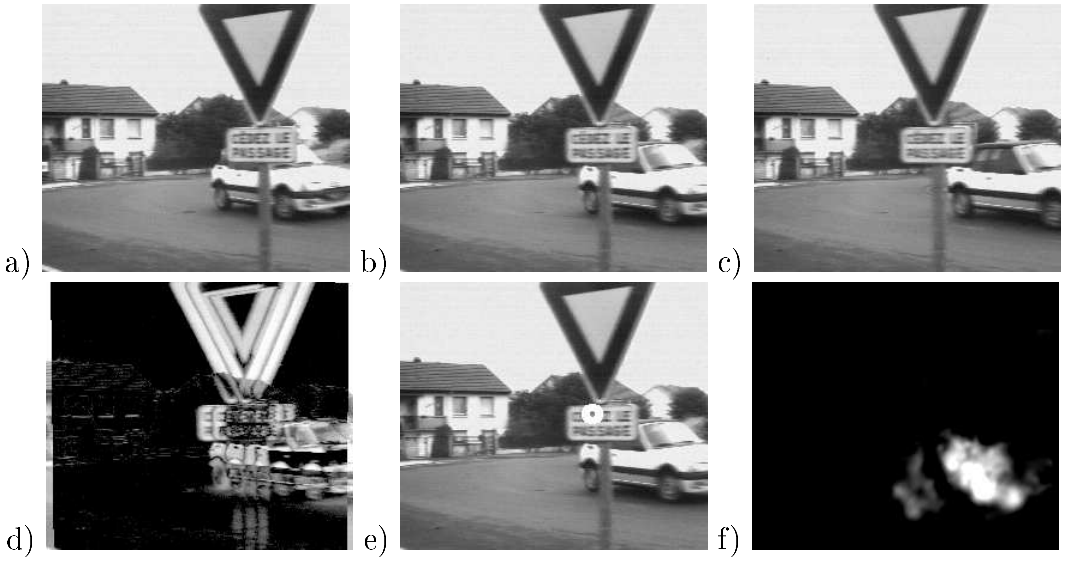

Figure 7: Moving object detection relying on a single reference point.
(a,b,c) Three image frames from a sequence obtained by a camera translating from left to right, inducing parallax motion of different magnitudes on the house, road, and road-sign. The car moves independently from left to right. The middle frame (Fig. 7.b) was chosen as the frame of reference.
(d) Differences taken after $2D$ image registration. The detected $2D$ planar motion was that of the house, and is canceled by the $2D$ registration. All other scene parts that have different $2D$ motions (i.e., *parallax* motion or *independent* motion) are misregistered.
(e) The selected point of reference (a point on the road-sign) highlighted by a white circle.
(f) The measure of $3D$-*in*consistency of all points in the image with respect to the selected road-sign point. Bright regions indicate violations in $3D$ rigidity detected over three frames with respect to that point. These regions correspond to the car. Regions close to the image boundary were ignored. All other regions of the image appear to move $3D$-consistently with the road-sign point.

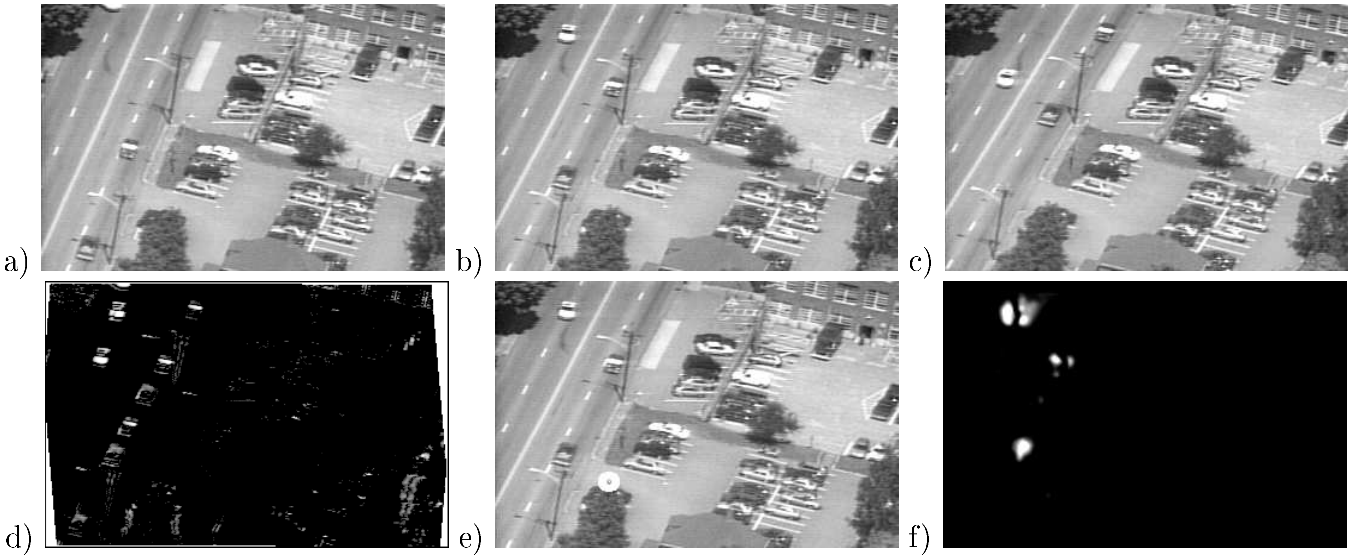

Figure 8: Moving object detection relying on a single reference point.
(a,b,c) Three image frames from a sequence obtained by a camera mounted on a helicopter (flying from left to right while turning), inducing some parallax motion (of different magnitudes) on the house-roof and trees (bottom of the image), and on the electricity poles (by the road). Three cars move independently on the road. The middle frame (Fig. 8.b) was chosen as the frame of reference.
(d) Differences taken after $2D$ image registration. The detected $2D$ planar motion was that of the ground surface, and is canceled by the $2D$ registration. All other scene parts that have different $2D$ motions (i.e., *parallax* motion or *independent* motion) are misregistered.
(e) The selected point of reference (a point on a tree at the bottom left of the image) highlighted by a white circle.
(f) The measure of $3D$-*in*consistency of each point in the image with respect to the selected tree point. Bright regions indicate violations in $3D$ rigidity detected over three frames with respect to that point. These regions correspond to the three cars (in the reference image 8.b). Regions close to the image boundary were ignored. All other regions of the image appear to move $3D$-consistently with the tree point.

as moving consistently with the selected tree point.

In [23] a rigidity constraint between three frames in the form of the trilinear tensor has been presented using regular image displacements. However, it requires having a set of image points which are known *a priori* to belong to the single $3D$ moving object. Selecting an inconsistent set of points will lead to an erroneous tensor. This issue was addressed in [32], where the trilinear constraint was used to segment and group multiple moving objects using robust techniques.

The ability of the parallax rigidity constraint to detect $3D$-inconsistencies with respect to a *single* point, provides a natural way to *bridge* the gap between $2D$ algorithms (which assume that any $2D$ motion different than the planar motion is an independently moving object), and $3D$ algorithms (which rely on having prior knowledge of a consistent set of points, or alternatively, dense parallax data).

4.3 Shape Recovery

Methods for recovering 3D depth from multiple views of calibrated cameras have been suggested. More recently, methods have been developed for recovering *projective structure* or *affine structure* from *uncalibrated* cameras [10, 13, 22]. Under this category, methods for recovering structure using *planar parallax* motion have been proposed [17, 21, 24]. Those methods rely on prior estimation of partial camera geometry, in particular, on estimating the camera epipole. Some methods for recovering camera or scene geometry in the presence of sparse independent motion have also been suggested [19, 30].

Recently, a complete theory has been presented [33, 5] for estimating the shape *directly* from image displacements over multiple frames, *without* the need to recover the epipolar geometry. They assume that the scene is static, i.e., that all points selected for the shape estimation are known to belong to a single rigid body. They do not address the problem of shape recovery in dynamic scenes, where the amount of image motion due to independent moving object is not negligible.

The parallax-based structure constraint (Eq. (9)) can be used to recover a relative 3D structure between pairs of points directly from their parallax displacements. This implies that the structure of the entire scene can be recovered relative to a *single* reference image point (of non-zero parallax). Singularities occur when the denominator of constraint (9) tends to zero, i.e., at points that lie along the line passing through the reference point in the direction of its parallax vector.

Fig. 9 shows an example of recovering structure of an entire scene relative to a *single* reference point. Three views obtained by a hand-held camera of a rug covered with toy cars and boxes (whose heights were measured), were used as the source data. The detected $2D$ planar motion was that of the rug (Fig. 9.d). A *single* point with non-zero planar parallax was selected as a reference point for estimating relative shape (Fig. 9.e). Fig. 9.f shows the recovered relative structure of the entire scene from *two* frames (Figs. 9.b and 9.c). Regions close to the image boundary were ignored. The obtained results were quite accurate except along the singular line in the direction of the parallax displacement of the reference point. The singular line is evident in Fig. 9.f.

The singularities can be removed and the quality of the computed structure can be improved either by using multiple frames or by using multiple reference points:

- Multiple frames: Singularities are removed by using multiple frames if their epipoles are non-colinear. Non-colinearity of epipoles can be detected through change in the direction of the

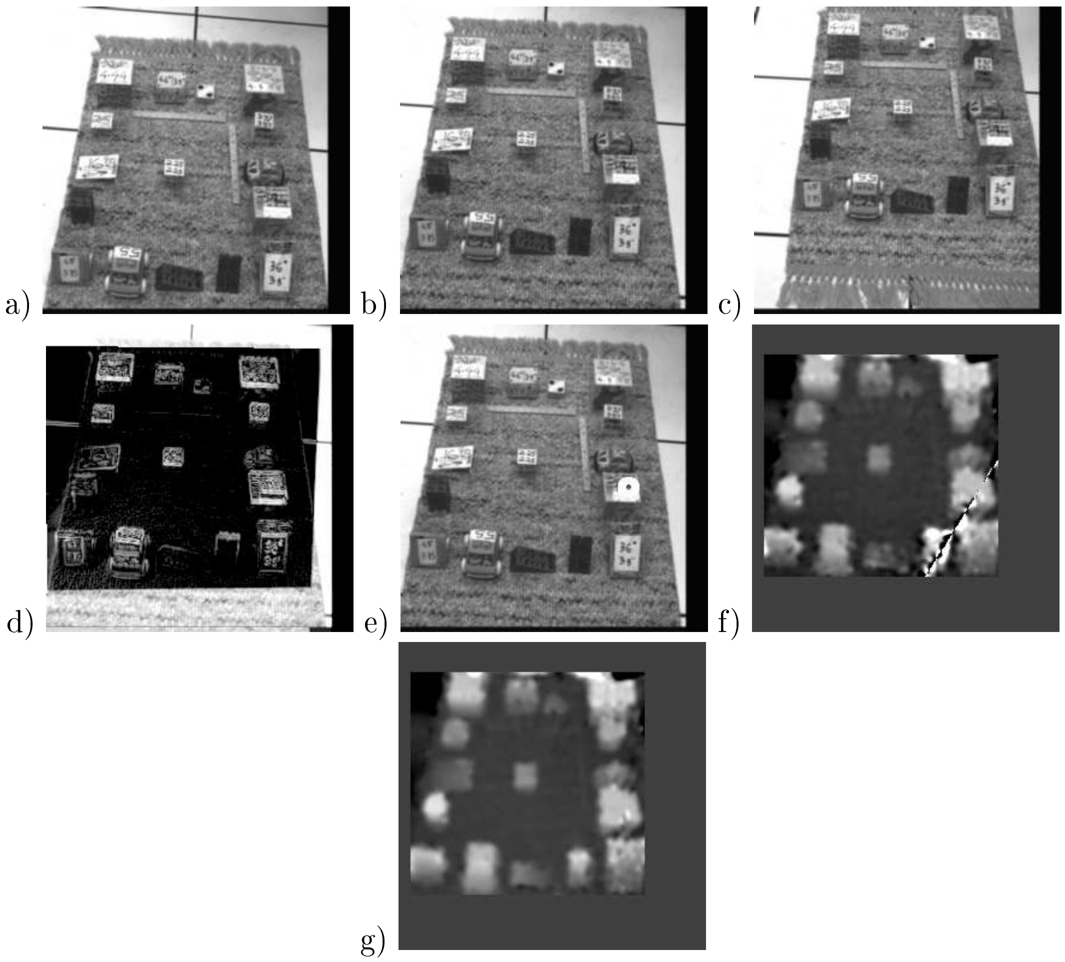

Figure 9: Shape recovery with respect to a single reference point.
(a,b,c) Three image frames from a sequence obtained by a hand-held camera of a rug covered with toy cars and boxes. The middle frame (Fig. 9.b) was chosen as the frame of reference.
(d) Differences taken after $2D$ image registration. The detected $2D$ planar motion was that of the rug, and is canceled by the $2D$ registration. All other scene parts (i.e., toys and boxes) are misregistered.
(e) The selected point of reference (a point on one of the boxes in the bottom right) highlighted by a white circle.
(f) The recovered relative structure of the entire scene from *two* frames (Figs. 9.b and 9.c) relative to the selected point of reference. Regions close to the image boundary were ignored. The interpreted relative heights were quite accurate except along the singular line in the direction of the parallax displacement of the reference point.
(g) The recovered relative structure of the entire scene using all three frames with respect to the selected point of reference. Regions close to the image boundary were ignored. The singular line has disappeared, providing more accurate shape.

parallax displacement of the reference image point.

- Multiple points: Singularities can be removed by using additional reference image points. An additional reference point should be chosen so that: (i) it does *not* lie on the singular line, i.e., in the direction of the parallax vector of the first reference point (it should preferably be chosen on the line *perpendicular* to that), and (ii) the additional reference point should be first verified to move consistently with the first reference point through the rigidity constraint (11) over a few frames.

Of course, combinations of multiple reference points over multiple frames can also be used. Fig. 9.g shows an example of recovering structure of an entire scene from *three* frames relative to the same *single* reference point as in Fig. 9.f. The singular line in Fig. 9.f has disappeared.

The ability of obtain relatively good structure information even with respect to a *single* point has several important virtues:

- Like [33, 5], it does not require the estimation of the epipole, and therefore does not require dense parallax information.
- Unlike previous methods for recovering structure, (including [33, 5]), it provides the capability to handle dynamic scenes, as it does not require having a collection of image points which is known *a priori* to belong to the single $3D$ rigid object.
- Since it relies on a single parallax vector, it provides a natural continuous way to bridge the gap between $2D$ cases, that assume only planar motion exists (i.e., no parallax data), and $3D$ cases that rely on having dense parallax data.

4.4 New View Generation

In this section we describe an approach based on the parallax rigidity constraint (11) for generating novel views using a set of "model" views. This work is still in early stages, and therefore no experimental results are provided.

Methods for generating new views based on recovering epipolar geometry (e.g., [9]) are likely to be more noise sensitive than methods that generate the new view based on $2D$ information alone [23] i.e., without going from $2D$ through a $3D$ medium in order to reproject the information once again onto a new $2D$ image plane (the virtual view). The approach we describe in this section to new view generation does not require any epipolar geometry or shape estimation.

Given two "model" frames, planar parallax motion can be computed for all image points between the first (reference) frame and the second frame. An image point with non-zero parallax is selected, and a "virtual" parallax vector is defined for that point from the reference frame to the "virtual" frame to be generated. The rigidity constraint (Equation 11) then specifies a single constraint on the virtual parallax motion of all other points between the reference frame to the *virtual* frame. Since each $2D$ parallax vector has two components (i.e., two unknowns), at least two "virtual" parallax vectors need to be specified in order to solve for all other virtual parallax vectors. Once the virtual parallax vectors are computed, the new virtual view can be created by warping the reference image twice: First, warping each image point by its computed virtual parallax. Then, globally warping the entire frame with a $2D$ virtual planar motion of the virtual homography.

Note that two virtual parallax vectors may not provide sufficient constraints for some image points. This is due to unfavorable locations of those points in the image plane with respect to the two selected reference points and their parallax vectors. However, other image points, for whom the constraint was robust and sufficient to produce reliable virtual parallax, can be used (once their virtual parallax has been computed) as additional points to reliably constrain the virtual parallax of the singular points.

5 The Generalized Parallax Constraint

The pairwise-parallax constraint (Eq. (9)) can be extended to handle full image motion (as opposed to *parallax* motion), even when the homography is unknown. This is useful for handling scenes that do not contain a physical planar surface. We will present a form of a *generalized parallax constraint* between two frames in terms of the *unknown* homography parameters and the relative projective structure of pairs of points.

Eqs. (1) and (2) can be unified into a single form as:

$$\begin{aligned} \vec{p} &= \vec{p_w} + \frac{\gamma}{d'_\pi}(T_Z\vec{p_w} - \vec{t}) \\ &= \frac{A'\vec{p'}}{{a'}_3^T\vec{p'}} + \frac{\gamma}{d'_\pi}(T_Z\frac{A'\vec{p'}}{{a'}_3^T\vec{p'}} - \vec{t}). \end{aligned} \tag{13}$$

(See notations in Section 2). A' is an *unknown* homography from frame2 to the first frame. It could relate to *any* planar surface in the scene, in particular a *virtual* plane.

In [23], a similar form to Eq. (13) was used to factor out the relative affine structure over three frames. This lead to a constraint on the image coordinates in terms of the two homographies (between three frames) and the two epipoles. In our case, *before* we factor out the structure, we *first* factor out the epipole, in a similar manner performed in Section 2.

Let $\vec{P_1}$ and $\vec{P_2}$ be two scene points with homogeneous coordinates $\vec{p_1}$ and $\vec{p_2}$. Then from Eq. (13):

$$\frac{d'_\pi}{\gamma_1}(\vec{p_1} - \frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1}) = T_Z\frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1} - \vec{t}$$

$$\frac{d'_\pi}{\gamma_2}(\vec{p_2} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2}) = T_Z\frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2} - \vec{t}.$$

Subtracting the two equations eliminates $\vec{t}$, yielding:

$$\frac{d'_\pi}{\gamma_1}(\vec{p_1} - \frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1}) - \frac{d'_\pi}{\gamma_2}(\vec{p_2} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2}) = T_Z(\frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2}). \tag{14}$$

The equality (14) entails that the vectors on both sides of the equation are parallel, which leads to:

$$(\frac{d'_\pi}{\gamma_1}(\vec{p_1} - \frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1}) - \frac{d'_\pi}{\gamma_2}(\vec{p_2} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2}))^T(\frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2})_\perp = 0.$$

Multiplying both sides by $\frac{\gamma_2}{d'_\pi}$ gives the *generalized parallax constraint* in terms the relative projective structure $\frac{\gamma_2}{\gamma_1}$:

$$(\frac{\gamma_2}{\gamma_1}(\vec{p_1} - \frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1}) - (\vec{p_2} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2}))^T(\frac{A'\vec{p'}_1}{a'^T_3\vec{p'}_1} - \frac{A'\vec{p'}_2}{a'^T_3\vec{p'}_2})_\perp = 0. \tag{15}$$

The *generalized parallax constraint* (15) is expressed in terms of the homography A', the image coordinates of a pair of points in two frames, and the relative projective structure of the two points. The generalized constraint does not involve the epipoles.

The generalized parallax constraint suggests a new implicit representation of *general* $2D$ image motion: Rather than looking at the representation of $2D$ image motion in terms of: *homography + epipole + projective structure* [17, 21, 24] it suggests an implicit representation of $2D$ image motion in terms of: *homography + relative projective structure* of pairs of points. Since this representation does not contain the epipole, it can be easily extended to multiple frames.

Using Eq. (15), the *invariant* relative projective structure $\frac{\gamma_2}{\gamma_1}$ can be expressed explicitly in terms of the plane homography and the image positions of the two points in the two frames:

$$\frac{\gamma_2}{\gamma_1} = \frac{(\vec{p_2} - \frac{A'\vec{p'}_2}{a'^T_3 \vec{p'}_2})^T (\frac{A'\vec{p'}_1}{a'^T_3 \vec{p'}_1} - \frac{A'\vec{p'}_2}{a'^T_3 \vec{p'}_2})_\perp}{(\vec{p_1} - \frac{A'\vec{p'}_1}{a'^T_3 \vec{p'}_1})^T (\frac{A'\vec{p'}_1}{a'^T_3 \vec{p'}_1} - \frac{A'\vec{p'}_2}{a'^T_3 \vec{p'}_2})_\perp} \tag{16}$$

Since $\frac{\gamma_2}{\gamma_1}$ is invariant to new camera positions, it can be factored out over multiple frames (j and k):

$$\frac{(\vec{p_2} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})^T (\frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})_\perp}{(\vec{p_1} - \frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1})^T (\frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})_\perp} = \frac{(\vec{p_2} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})^T (\frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})_\perp}{(\vec{p_1} - \frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1})^T (\frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})_\perp},$$

which leads to a rigidity constraint of the form:

$$\begin{aligned}(\vec{p_2} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})^T (\frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})_\perp (\vec{p_1} - \frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1})^T (\frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})_\perp - \\ (\vec{p_2} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})^T (\frac{A^k\vec{p^k}_1}{a^{k\,T}_3 \vec{p^k}_1} - \frac{A^k\vec{p^k}_2}{a^{k\,T}_3 \vec{p^k}_2})_\perp (\vec{p_1} - \frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1})^T (\frac{A^j\vec{p^j}_1}{a^{j\,T}_3 \vec{p^j}_1} - \frac{A^j\vec{p^j}_2}{a^{j\,T}_3 \vec{p^j}_2})_\perp = 0.\end{aligned} \tag{17}$$

Eq. (17) is a rigidity constraint on a pair of points across three frames. Like the trilinear tensor [23], it involves the parameters of two homographies across three frames. Unlike the trilinear tensor, it does *not* involve the epipoles, but instead, is expressed on *pairs* of points.

The trilinear constraints are based on a initial reference points, and any additional point adds four linearly independent equations to constrain the unknowns of the tensor (which are combinations of the homography parameters and the epipoles). In the generalized parallax rigidity constraint, the basis is a *pair* of points. Here also, any additional point adds four linearly independent rigidity constraints. These are not described here and are beyond the scope of this paper. However, they can be derived through factoring out T_Z from Eq. (14) with the additional third point (still within a pair of frames), followed by very similar steps to those in [23] to form the four linearly independent equations over three frames.

6 Conclusion

In this paper, we described a geometric relationship between the image motion of *pairs* of points over multiple frames. This relationship is based on the *parallax* displacements of points with respect to an arbitrary planar surface, and does not involve epipolar geometry. We derived a constraint over two frames relating the projective structure (with respect to the plane) of any pair of points, based *only* on their image coordinates and their parallax motion. We also derived a $3D$-rigidity constraints between *pairs of points* over *multiple frames* and between *multiple points* over *pairs of frames*, based on their parallax displacements alone, without involving any scene or camera geometry.

We showed applications of these parallax-based constraints to solving three important problems in $3D$ scene analysis: the recovery of $3D$ scene structure, the detection of moving objects in the presence of camera induced motion, and the synthesis of new camera views based on a given set of views. Our approach can handle difficult situations for $3D$ scene analysis, e.g., where there is only a small set of parallax vectors, and in the presence of independently moving objects.

We outlined the generalization of our parallax based constraints to full image motion (as opposed to *parallax* motion), even when the homography is unknown. This is useful for handling scenes that do not contain a physical planar surface.

The use of the parallax constraints derived here provides a continuum between "$2D$ algorithms" and the "$3D$ algorithms" for each of the problems mentioned above. Our future plans include the completion of the algorithms described here, especially the technique for new view generation.

References

[1] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, pages 237–252, Santa Margarita Ligure, May 1992.

[2] J.R. Bergen, P.J. Burt, K. Hanna, R. Hingorani, P. Jeanne, and S. Peleg. Dynamic multiple-motion computation. In Y.A. Feldman and A. Bruckstein, editors, *Artificial Intelligence and Computer Vision: Proceedings of the Israeli Conference*, pages 147–156. Elsevier, 1991.

[3] J.R. Bergen, P.J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14:886–895, September 1992.

[4] P.J. Burt, R. Hingorani, and R.J. Kolczynski. Mechanisms for isolating component patterns in the sequential analysis of multiple motion. In *IEEE Workshop on Visual Motion*, pages 187–193, Princeton, New Jersey, October 1991.

[5] S. Carlsson. Duality of reconstruction and positioning from projective views. In *Workshop on Representations of Visual Scenes*, 1995.

[6] T. Darrell and A. Pentland. Robust estimation of a multi-layered motion representation. In *IEEE Workshop on Visual Motion*, pages 173–178, Princeton, New Jersey, October 1991.

[7] T. Darrell and E. Simoncelli. "nulling" filters and the separation of transparent motions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 738–739, New York, June 1993.

[8] O. Faugeras. *Three-Dimensional Computer Vision.* M.I.T. Press, Cambridge, MA., 1993.

[9] O. Faugeras and L. Robert. What can two images tell us about a third one? In *European Conference on Computer Vision*, pages 485–492, May 1994.

[10] 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.

[11] 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.

[12] K. Hanna. Direct multi-resolution estimation of ego-motion and structure from motion. In *IEEE Workshop on Visual Motion*, pages 156–162, Princeton, NJ, October 1991.

[13] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 1992.

[14] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *European Conference on Computer Vision*, pages 282–287, Santa Margarita Ligure, May 1992.

[15] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, January 1994.

[16] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, Wa., June 1994.

[17] Rakesh Kumar, P. Anandan, and K. Hanna. Shape recovery from multiple views: a parallax based approach. In *DARPA IU Workshop*, Monterey, CA, November 1994.

[18] Rakesh Kumar, P. Anandan, M. Irani, J. R. Bergen, and K. J. Hanna. Representation of scenes from collections of images. In *Workshop on Representations of Visual Scenes*, 1995.

[19] J.M. Lawn and R. Cipolla. Robust egomotion estimation from affine motion parallax. In *European Conference on Computer Vision*, pages 205–210, May 1994.

[20] F. Meyer and P. Bouthemy. Region-based tracking in image sequences. In *European Conference on Computer Vision*, pages 476–484, Santa Margarita Ligure, May 1992.

[21] Harpreet Sawhney. 3d geometry from planar parallax. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994.

[22] A. Shashua. Projective structure from uncalibrated images: Structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:778–790, 1994.

[23] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:779–789, 1995.

[24] 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.

[25] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *International Conference on Computer Vision*, pages 920–925, Cambridge, MA, June 1995.

[26] M. Shizawa. On visual ambiguities due to transparency in motion and stereo. In *European Conference on Computer Vision*, pages 411–419, Santa Margarita Ligure, May 1992.

[27] M. Shizawa and K. Mase. Principle of superposition: A common computational framework for analysis of multiple motion. In *IEEE Workshop on Visual Motion*, pages 164–172, Princeton, New Jersey, October 1991.

[28] M. E. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *DARPA IU Workshop*, pages 271–283, 1990.

[29] W.B. Thompson and T.C. Pong. Detecting moving objects. *International Journal of Computer Vision*, 4:29–57, 1990.

[30] P.H.S. Torr and D.W. Murray. Stochastic motion clustering. In *European Conference on Computer Vision*, pages 328–337, May 1994.

[31] P.H.S. Torr, A. Zisserman, and S.J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *International Conference on Computer Vision*, pages 1037–1042, Cambridge, MA, June 1995.

[32] P.H.S. Torr, A. Zisserman, and D.W. Murray. Motion clustering using the trilinear constraint over three views. In *Workshop on Geometric Modeling and Invariants for Computer Vision*, 1995.

[33] 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.

[34] Andrew Zisserman. A case against epipolar geometry. In *Applications of Invariance in Computer Vision*, pages 35–50, Ponta Delgada, Azores, October 1993.

Appendix

In this appendix we rederive the decomposition of image motion into the image motion of a planar surface (a homography) and residual parallax displacements.

Let $\vec{P} = (X, Y, Z)^T$ and $\vec{P'} = (X', Y', Z')^T$ denote the Cartesian coordinates of a scene point with respect to two different camera views, respectively. An arbitrary $3D$ rigid coordinate transformation between $\vec{P}$ and $\vec{P'}$ can be expressed by:

$$\vec{P'} = R\vec{P} + \vec{T'}, \tag{18}$$

where R represents the rotation between the two camera coordinate systems, $\vec{T'} = (T'_X, T'_Y, T'_Z)$ denotes the $3D$ translation in between the two views as expressed in the coordinate system of the second camera, and $\vec{T} = (T_X, T_Y, T_Z) = -R^{-1}\vec{T'}$ denotes the same quantity in the coordinate system of the first camera.

Let Π denote an arbitrary 3D planar surface (real or virtual). Let $\vec{N}$ denote its normal as expressed in the coordinate system of the first camera, and $\vec{N'}$ denote the same quantity in the coordinate system of the second camera. Any point $\vec{P} \in \Pi$ satisfies the equation $\vec{N}^T\vec{P} = d_\pi$ (and similarly $\vec{N'}^T\vec{P'} = d'_\pi$). For a general scene point $\vec{P}$:

$$\begin{aligned} \vec{N}^T\vec{P} &= d_\pi + H \\ \vec{N'}^T\vec{P'} &= d'_\pi + H \end{aligned} \tag{19}$$

where H denotes the perpendicular distance of $\vec{P}$ from the plane Π. Note that H is invariant with respect to the two camera coordinate systems (see Figure 1).

By inverting Equation (18), we obtain

$$\begin{aligned} \vec{P} &= R^{-1}\vec{P'} - R^{-1}\vec{T'} \\ &= R^{-1}\vec{P'} + \vec{T} \end{aligned} \tag{20}$$

From Equation (19), we derive

$$\frac{\vec{N'}^T\vec{P'} - H}{d'_\pi} = 1 \tag{21}$$

Substituting this in Equation (20) obtains

$$\vec{P} = R^{-1}\vec{P'} + \vec{T}\frac{(\vec{N'}^T\vec{P'} - H)}{d'_\pi} \tag{22}$$

$$= (R^{-1} + \frac{\vec{T}\vec{N'}^T}{d'_\pi})\vec{P'} - \frac{H}{d'_\pi}\vec{T}. \tag{23}$$

Let $\vec{p} = (x, y, 1)^T = \frac{1}{Z}K\vec{P}$ and $\vec{p'} = (x', y', 1)^T = \frac{1}{Z'}K'\vec{P'}$ denote the images of the scene point P in the two camera views as expressed in homogeneous coordinates. K and K' are 3×3 matrices representing the internal calibration parameters of the two cameras. In general K has the following form [8]:

$$K = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & 1 \end{bmatrix}.$$

Also, define $\vec{t} = (t_x, t_y, t_z)^T = K\vec{T}$. (Note that $(K\vec{P})_z = Z$, $(K'\vec{P'})_z = Z'$, and $t_z = T_Z$). Multiplying both sides of Eq. (23) by $\frac{1}{Z'}K$ gives:

$$\frac{Z}{Z'}\vec{p} = K(R^{-1} + \frac{\vec{T}\vec{N'}^T}{d'_\pi})K'^{-1}\vec{p'} - \frac{H}{d'_\pi Z'}\vec{t} \tag{24}$$

Hence,

$$\vec{p} \cong A\vec{p'} - \frac{H}{d'_\pi Z'}\vec{t}, \tag{25}$$

where $\cong$ denotes equality up to an arbitrary scale. $A' = K(R^{-1} + \frac{\vec{T}\vec{N'}^T}{d'_\pi})K'^{-1}$ is a 3×3 matrix which represents the coordinate transformtaion of the planar surface Π between the two camera views, i.e., the homography between the two views due to the plane Π. Scaling both sides by their third component (i.e., projection) gives the equality:

$$\begin{aligned}
\vec{p} &= \frac{A'\vec{p'} - \frac{H}{d'_\pi Z'}\vec{t}}{a'_3\vec{p'} - \frac{HT_z}{d'_\pi Z'}} && (26) \\
&= \frac{A'\vec{p'}}{a'_3\vec{p'}} - \frac{A'\vec{p'}}{a'_3\vec{p'}} + \frac{A'\vec{p'} - \frac{H}{d'_\pi Z'}\vec{t}}{a'_3\vec{p'} - \frac{HT_z}{d'_\pi Z'}} && (27) \\
&= \frac{A'\vec{p'}}{a'_3\vec{p'}} + \frac{\frac{HT_z}{d'_\pi Z'}}{(a'_3\vec{p'} - \frac{HT_z}{d'_\pi Z'})}\frac{A'\vec{p'}}{a'_3\vec{p'}} - \frac{\frac{H}{d'_\pi Z'}}{a'_3\vec{p'} - \frac{HT_z}{d'_\pi Z'}}\vec{t} && (28)
\end{aligned}$$

where a'_3 denotes the third row of the matrix A'. Moreover by considering the third component of the vector Equation (24), we obtain

$$\frac{Z}{Z'} = a'_3\vec{p'} - \frac{HT_z}{d'_\pi Z'} \tag{29}$$

Substituting this into Equation (28), we obtain

$$\vec{p} = \frac{A'\vec{p'}}{a'_3\vec{p'}} + \frac{H}{Z}\frac{T_z}{d'_\pi}\frac{A'\vec{p'}}{a'_3\vec{p'}} - \frac{H}{Zd'_\pi}\vec{t} \tag{30}$$

When $T_z \neq 0$, let $\vec{e} = \frac{1}{T_z}\vec{t}$ denote the epipole in the first image. Then,

$$\vec{p} = \frac{A'\vec{p'}}{a'_3\vec{p'}} + \frac{H}{Z}\frac{T_z}{d'_\pi}(\frac{A'\vec{p'}}{a'_3\vec{p'}} - e) \tag{31}$$

On the other hand, when $T_z = 0$, we obtain

$$\vec{p} = \frac{A'\vec{p'}}{a'_3\vec{p'}} - \frac{H}{Zd'_\pi}\vec{t}. \tag{32}$$

The point denoted by the vector $\frac{A'\vec{p'}}{a'_3\vec{p'}}$ is of special interest, since it represents the location to which the point $\vec{p'}$ is transformed due to the homography A'. In Figure 1 this is denoted as the point $\vec{p_w}$. Substituting these into Equations (31) and (32) yields:
when $T_z \neq 0$:

$$\vec{p} = \vec{p_w} + \frac{H}{Z}\frac{T_z}{d'_\pi}(\vec{p_w} - \vec{e}) \tag{33}$$

and when $T_z = 0$:

$$\vec{p} = \vec{p_w} - \frac{H}{Zd'_\pi}\vec{t} \tag{34}$$