Multiview Constraints on Homographies

Lihi Zelnik-Manor and Michal Irani, *Member*, *IEEE*

Abstract—The image motion of a planar surface between two camera views is captured by a homography (a $2D$ projective transformation). The homography depends on the intrinsic and extrinsic camera parameters, as well as on the $3D$ plane parameters. While camera parameters vary across different views, the plane geometry remains the same. Based on this fact, we derive *linear* subspace constraints on the relative homographies of multiple (≥ 2) planes across multiple views. The paper has three main contributions: 1) We show that the collection of all relative homographies (homologies) of a *pair of planes* across *multiple views*, spans a 4-dimensional linear subspace. 2) We show how this constraint can be extended to the case of *multiple planes* across *multiple views*. 3) We show that, for some restricted cases of camera motion, linear subspace constraints apply also to the set of homographies of a *single plane* across multiple views. All the results derived in this paper are true for uncalibrated cameras. The possible utility of these multiview constraints for improving homography estimation and for detecting nonrigid motions are also discussed.

Index Terms—Homographies, homologies, motion estimation, multiview analysis.

1 Introduction

Homography estimation is used for $3D$ analysis [18], [10], [21], [25], [11], [7], [14], [17], [16], mosaicing [13], camera calibration [26], [31], and more. The induced homography between a pair of views depends on the intrinsic and extrinsic camera parameters and on the $3D$ plane parameters [10]. While camera parameters vary across different views, the plane geometry remains the same. In this paper, we show how we can exploit this fact to derive multiview *linear* subspace constraints on the relative homographies of multiple (≥ 2) planes, and for restricted cases of camera motion—on the collection of homographies of a single plane.

Linear subspace constraints on homographies have been previously derived by Shashua and Avidan [22]. They showed that the collection of homographies of multiple planes between a pair of views spans a 4-dimensional linear subspace. This constraint, however, requires the number of planes in the scene to be greater than 4. In this paper, we obtain subspace constraints for scenes containing as little as two planes (and often a single plane). Here, the need for multiple planes in [22] is replaced by the need for multiple views, which is often less restrictive. We first derive a constraint for the relative homographies (homologies) of a *pair of planes* over *multiple* (> 4) *views* (Section 2). This constraint is then extended to a constraint on the relative homographies of *multiple planes* across *multiple views* (Section 3). We show how appropriate scaling of the homographies leads to further reduction in the dimensionality of these subspaces (Section 4) and to an extension of the multiplane constraint of [22] to a multiview multiplane constraint (Section 5). We show that, for some cases of restricted camera motion, linear subspace constraints apply also to the homographies of a *single* plane across multiple views (Section 6).

Different video-related applications can benefit from such multiview constraints. For example, many algorithms based on planar homographies (e.g., [18], [21], [11], [28], [17]) or on planar homologies (e.g., [23], [4]) rely on accurate precomputation of these homographies (or homologies). However, the image region corresponding to a planar surface may be small. In such cases, the homography estimation tends to be highly inaccurate [25] (i.e., when applied to small image regions). Adding the multiview subspace constraints presented in this paper as *additional* constraints to existing homography estimation methods can significantly improve the accuracy of the estimated homographies. Furthermore, violations of these multiview constraints can form additional cues for detecting nonrigid motions (e.g., for moving object detection). We provide empirical evaluations of these in Section 7.

All the results derived in this paper are true for uncalibrated cameras. A preliminary version of this paper appeared in [30].

1.1 Homographies—Basic Notations

First, we derive the basic homography notations which will be used later in the paper. Let $\vec{Q} = (X, Y, Z)^T$ and $\vec{Q'} = (X', Y', Z')^T$ denote the $3D$ coordinates of a scene point with respect to two different camera views. Let $\vec{q} = (x, y, 1)^T$ and $\vec{q'} = (x', y', 1)^T$ denote the homogeneous coordinates of its projection point in the two images. We can write

$$\vec{q} \cong C\vec{Q} \quad , \quad \vec{q'} \cong C'\vec{Q'}, \tag{1}$$

where $\cong$ denotes equality up to a scale factor. C and C' are 3×3 matrices [10] which capture the camera's internal parameters and the projection.

Let π be a planar surface with a plane normal $\vec{n}$, then $\vec{n}^T\vec{Q} = 1$ for all points $\vec{Q} \in \pi$ ($\vec{n} = \frac{\vec{n_\pi}}{d_\pi}$, where $\vec{n_\pi}$ is a unit vector in the direction of the plane normal and d_π is the distance of the plane from the first camera center). The

• The authors are with the Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot, 76100, Israel. E-mail: {lihi, irani}@wisdom.weizmann.ac.il.

Manuscript received 21 Mar. 2000; revised 25 Mar. 2001; accepted 26 July 2001.
Recommended for acceptance by H. Christensen.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111742.

0162-8828/02/$17.00 © 2002 IEEE

transformation between the $3D$ coordinates of a scene point $Q \in \pi$, in the two views, can be expressed by

$$\vec{Q}' = G\vec{Q}, \tag{2}$$

where

$$G = R + \vec{t}\vec{n}^T, \tag{3}$$

R is the rotation matrix capturing the change of orientation between the two cameras views, and $\vec{t}$ is the translation between the two camera views. Therefore, the induced transformation between the corresponding *image* points is

$$\vec{q}' \cong A\vec{q}, \tag{4}$$

where

$$A = C'(R + \vec{t} \cdot \vec{n}^T)C^{-1} \tag{5}$$

is the induced homography between the two views of the plane π. From (4), it is clear that, when A is computed from image point correspondences, it can be estimated only up to a scale factor. For more details, see [10].

2 Multiview Rank-4 Constraints on Two Planes

In this section, we show that the induced relative image motion between two planar surfaces over multiple frames spans a low-dimensional (4-dimensional) linear subspace.

Let J be a "reference" image and let $J^1, \ldots, J^F$ be F other images of the same scene taken from different views. Let π_r and π_p be two planar surfaces in the scene with plane normals $\vec{n}_r$ and $\vec{n}_p$, respectively. Let A_r^f and A_p^f denote their corresponding homographies[1] between the reference image J and an image $J^f (f = 1, \ldots, F)$. Composing the homography of π_p with the *inverse* of the homography of π_r yields a "relative homography":

$$H^f = (A_r^f)^{-1} A_p^f. \tag{6}$$

This relative homography captures the induced relative image motion between the two planes and is a "plane homology" [20]. Some properties and invariants of planar homologies have been discussed in [8] and used in [17], [4], [23]. Here, we present a different set of constraints on homologies. Using (5) and the Sherman-Morisson formula[2] [19], it can be shown that, for rigidly moving planes π_r and π_p, the matrix H^f has the form

$$\begin{aligned} H^f &= I + \vec{v}^f \vec{m}^T \\ &\equiv \begin{bmatrix} 1 + h_1 & h_2 & h_3 \\ h_4 & 1 + h_5 & h_6 \\ h_7 & h_8 & 1 + h_9 \end{bmatrix}, \end{aligned} \tag{7}$$

where

$$\vec{v}^f = C\frac{R^{f^{-1}}\vec{t}^f}{1 + \vec{n_r}^T R^{f^{-1}} \vec{t}^f}, \qquad \vec{m}^T = (\vec{n_p}^T - \vec{n_r}^T)C^{-1},$$

I is the 3×3 identity matrix, and $\vec{t}^f, R^f$ are the camera translation vector and the camera rotation matrix, respectively (between the reference image J and the image J^f). C is the camera internal parameter matrix at the reference view J. The matrix C^f (i.e., the calibration matrix of J^f) is *eliminated* by the composition. Note that $\vec{v}^f$ is only *view-dependent*, (i.e., is common to all rigidly moving planes between a pair of views J and J^f), whereas $\vec{m}$ is only *plane-dependent*, (i.e., is common to all views for a pair of planes π_r and π_p).

A homology is a $2D$ projective transformation, which has a fixed line and a separate fixed point. One can easily show (see [4]) that $\vec{v}^f$ is the fixed point. This is actually the epipole (up to a scale factor) of view J^f in image J. $\vec{m}$, which corresponds to the difference of the two plane normals, is actually the image of the intersection line between the two planes π_r and π_p (again, up to a scale factor). This is the fixed line of the homology H^f. Therefore, the fixed point of the homology H^f is a view-dependent property, while the fixed line of the homology is a plane-dependent property.

We now proceed to derive multiview constraints on homologies. Rearranging the components of the relative-homography (3×3) matrix H^f in a single (9×1) column vector $\vec{h}^f$, we can rewrite (7) as:

$$\vec{h}^f = \mathcal{N}_{9\times4} \begin{bmatrix} \vec{v}^f \\ 1 \end{bmatrix}_{4\times1}, \tag{8}$$

where

$$\mathcal{N} = \begin{bmatrix} \vec{m} & \vec{0} & \vec{0} & \vec{i_1} \\ \vec{0} & \vec{m} & \vec{0} & \vec{i_2} \\ \vec{0} & \vec{0} & \vec{m} & \vec{i_3} \end{bmatrix} \tag{9}$$

and $\vec{i_1} = [1,0,0]^T$, $\vec{i_2} = [0,1,0]^T$, $\vec{i_3} = [0,0,1]^T$. In practice, $\vec{h}^f$ can be estimated only up to an unknown scale factor λ^f because A_r^f and A_p^f are known only up to a scale factor (see (4)). In other words, the *computed* relative homographies (homologies), denoted by $\vec{\tilde{h}^f}$ (We will henceforth denote by ~ all entities computed up to a scale factor), are

$$\vec{\tilde{h}^f} = \lambda^f \vec{h}^f. \tag{10}$$

We now consider *multiple* views $J^f, f = 1 \ldots F$. Since the matrix $\mathcal{N}$ depends only on plane normal parameters and on the camera calibration of the *reference* view, it is common to all views $f = 1 \ldots F$. Hence, we can stack all the computed relative homography vectors (homology vectors) in a $9 \times F$ matrix $\mathcal{H}$, where each column corresponds to a single image view J^f (relative to the reference view J):

[1] The superscripts denote the image/frame index; the subscripts denote the plane index.

[2] For a square matrix A and two column vectors $\vec{u}$, $\vec{w}$, the Sherman-Morrison formula gives $(A + \vec{u}\vec{w}^T)^{-1} = A^{-1} - \frac{(A^{-1}\vec{u})(\vec{w}^T A^{-1})}{I + \vec{w}^T A^{-1}\vec{u}}$.

$$
\begin{aligned}
[\mathcal{H}]_{9\times F} &= \begin{bmatrix} \tilde{\vec{h}}^1 & \cdots & \tilde{\vec{h}}^F \end{bmatrix}_{9\times F} \\
&= \begin{bmatrix} \vec{h}^1 & \cdots & \vec{h}^F \end{bmatrix}_{9\times F} \begin{bmatrix} \lambda^1 & & 0 \\ & \ddots & \\ 0 & & \lambda^F \end{bmatrix} \\
&= [\mathcal{N}]_{9\times 4} \begin{bmatrix} \vec{v}^1 & \cdots & \vec{v}^F \\ 1 & & 1 \end{bmatrix}_{4\times F} \begin{bmatrix} \lambda^1 & & 0 \\ & \ddots & \\ 0 & & \lambda^F \end{bmatrix}.
\end{aligned}
\qquad (11)
$$

The dimensionality of the matrices on the right-hand side of (11) implies that the matrix $\mathcal{H}$ is of rank 4 at most.[3] Hence, the collection of all relative homographies (homologies) of the two planes across all images, resides in a 4-dimensional linear subspace. This constraint is complementary to the constraint shown by Shashua and Avidan [22]. There, it was shown that the collection of homographies of *multiple* (> 4) *planes* between a pair (= 2) *of views*, spans a 4-dimensional linear subspace. In contrast, here we derived a rank-4 constraint for the homologies of a *pair* (= 2) of planes over multiple (> 4) *views*. Note that, even though the two constraints have symmetric properties, they are not dual in the Carlsson-Weinshall-duality sense [3]. The "duality" here is in the *homography parameter space* and not in the $3D$ *space* as in [3]. Simple switching of camera centers and scene points in Shashua and Avidan's constraints will not lead to our constraints. Also, note that the multipleplane constraint shown in [22] applies to plane homographies whereas, the multiview constraint of (11) is restricted to homologies of a pair of planes and does not apply to a single plane homographies in the general case. In Section 5, we further show how the constraint of [22] can be extended to a multiview multiplane constraint on *homographies*.

Here, the dimension 4 derived can also be explained geometrically: When imaging the same pair of planes from multiple views, each new view adds 4 d.o.f. (degrees of freedom): three for the epipole and one scale factor. A similar geometric interpretation can also be given to the rank-4 constraint in [22]: When multiple planes are imaged from the same pair of views, each plane adds 4 degrees of freedom: three for the line of intersection of the new plane with an arbitrary reference plane (could also be the plane at infinity) and one scale factor.

3 Multiview Rank-4 Constraints on Multiple Planes

Homographies are determined only up to a scale factor. This scale factor differs for every pair of planes and for every pair of views, i.e., they are both view dependent and plane-dependent. Therefore, the extension of the *two-plane multiview* factorization ((11) in Section 2), or the *two-view multiplane* factorization [22] into a *multiview multiplane* factorization is not straightforward. We next show how the low-dimensionality linear subspace constraint can be extended to a constraint on multiple-planes across multiple views by enforcing these scale factors (denoted by λ_p^f) to be a product of two scalars: one which is view-dependent and one which is plane-dependent. This can be done without any calibration information, as explained below.

Let $\pi_1, \dots, \pi_P$ be P planar surfaces with normals $\vec{n}_1, \dots, \vec{n}_P$, respectively. Let $A_1^f, \dots, A_P^f$ be their corresponding homography matrices between the reference view J and each of the other views $J^f (f = 1 \dots F)$. Let π_r be a *reference plane* (e.g., could be chosen as the plane occupying the largest image region in the reference image). We assume that the 3×3 homologies $\tilde{H}_p^f (f = 1, \dots, F; p = 1, \dots, P)$ have been computed with respect to the reference plane π_r and the reference image J, and are each known only up to a scale factor. We can then arbitrarily set one of the six *off*-diagonal entries in each homology $\tilde{H}_p^f$ to be equal to 1 (i.e., $\tilde{h}_2$, $\tilde{h}_3$, $\tilde{h}_4$, $\tilde{h}_6$, $\tilde{h}_7$, or $\tilde{h}_8$; see (7)), and scale all the other entries of the homology accordingly. This results in a new (unknown) scale factor λ_p^f for each homology, which is guaranteed to be a bilinear product of two (unknown) scalars:

$$\lambda_p^f = \alpha^f \cdot \beta_p, \qquad (12)$$

where α^f is *view*-dependent and β_p is *plane*-dependent (e.g., if we set $\tilde{h}_3$ to be 1, then $\tilde{h}_3 = \lambda_p^f h_3 = 1$, i.e., $\lambda_p^f = \frac{1}{h_3}$ and, from (7), we get:

$$\lambda_p^f = \frac{1}{h_3} = \frac{1}{v_X^f} \cdot \frac{1}{m_{p_Z}}.$$

In other words, $\alpha^f = \frac{1}{v_X^f}$ and $\beta_p = \frac{1}{m_{p_Z}}$, where $\vec{v}^f = [v_X^f, v_Y^f, v_Z^f]^T$ and $\vec{m}_p = [m_{p_X}, m_{p_Y}, m_{p_Z}]^T$). Note that α^f is common to all planes and β_p is common to all views. Since all planar surfaces π_p share the same $3D$ camera motion between a pair of views, we can now extend (11) to *multiple planes* to get

$$
\begin{aligned}
\mathcal{G} &= \begin{bmatrix} \mathcal{H}_1 \\ \hline \vdots \\ \hline \mathcal{H}_P \end{bmatrix}_{9P\times F} \\
&= \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_P \end{bmatrix}_{9P\times 9P} \begin{bmatrix} \mathcal{N}_1 \\ \hline \vdots \\ \hline \mathcal{N}_P \end{bmatrix}_{9P\times 4} \begin{bmatrix} \vec{v}^1 & \cdots & \vec{v}^F \\ 1 & & 1 \end{bmatrix}_{4\times F} \\
&\quad \begin{bmatrix} \alpha^1 & & 0 \\ & \ddots & \\ 0 & & \alpha^F \end{bmatrix}_{F\times F}.
\end{aligned}
\qquad (13)
$$

The dimensionality of the matrices on the right-hand side of (13) implies that the matrix $\mathcal{G}$ is of rank 4 at most.

This implies that when solving for the homographies, if one of the six off-diagonal entries of the relative homographies (homologies) is consistently set to 1, we are guaranteed that the collection of all relative homographies, of *all planes* across *all views*, lies in a 4-dimensional linear subspace. Note, however, that the multiview, multiplane constraint has limited applicability because the required

[3] In practice, the actual rank may be even lower than 4, e.g., in cases of degenerate camera motion.

scaling of all the relative homographies is possible only when at least one of the six off-diagonal entries is consistently nonzero for all planes and for all views. An example where this condition fails to exist is when there is at least one view in which $v_X = 0$ (the first component of the epipole), another view in which $v_Y = 0$ (the second component of the epipole), and a third view in which $v_Z = 0$ (the third component of the epipole). (This, however, rarely happens in short segments of real video sequences). The two-plane constraint (Section 2), on the other hand, is *always* applicable.

4 Multiview Rank-3 Constraints

In Section 3, the unknown scale factors λ_p^f were not recovered, but were forced to be a bilinear product (12) of a view-dependent scalar (α^f) and a plane dependent scalar (β_p). This allowed us to extend the multiview linear subspace constraints to multiple planes. Next, we show how this scale factor can actually be recovered (from uncalibrated views), leading to different lower-dimensional linear subspace constraints on the homologies parameters of multiple planes across multiple views. Furthermore, it enables extending the rank-4 constraint on *homographies* of multiple planes between a pair of views [22] to a rank-4 constraint on *homographies* of multiple planes across multiple views (Section 5).

From (7), we have $H_p^f = I + \vec{v}^f \vec{m}_p^T$. It is easy to see that, for any vector $\vec{u} \perp \vec{m}_p$, we have $H_p^f \vec{u} = \vec{u}$, i.e., the *true* (unknown) homology H_p^f has an eigenvalue 1. Because the space of all vectors $\vec{u} \perp \vec{m}_p$ is of dimension 2 (confined to the plane perpendicular to $\vec{m}_p$), the multiplicity of that eigenvalue of H_p^f is two (see also [27], [10], [24]). From (10), we have $\tilde{H}_p^f = \lambda_p^f H_p^f$, which entails that the *computed* (known) homology $\tilde{H}_p^f$ has an eigenvalue λ_p^f of multiplicity 2. Therefore, for each computed homology $\tilde{H}_p^f$, we can compute its three eigenvalues and detect its eigenvalue λ_p^f with multiplicity 2. If we then scale $\tilde{H}_p^f$ by that λ_p^f, we get the *true* homology $H_p^f = \frac{1}{\lambda_p^f} \tilde{H}_p^f$. Next, we subtract I from each homology to get $\hat{H}_p^f = H_p^f - I$ and extend (11) to

$$\begin{aligned}[\hat{\mathcal{H}}]_{9\times F} &= \left[\, \hat{\vec{h}}^1 \quad \dots \quad \hat{\vec{h}}^F \,\right]_{9\times F} \\ &= [\hat{\mathcal{N}}]_{9\times 3} \left[\, \vec{v}^1 \quad \dots \quad \vec{v}^F \,\right]_{3\times F}\end{aligned}, \qquad (14)$$

where

$$\hat{\mathcal{N}} = \begin{bmatrix} \vec{m}_p & \vec{0} & \vec{0} \\ \vec{0} & \vec{m}_p & \vec{0} \\ \vec{0} & \vec{0} & \vec{m}_p \end{bmatrix}.$$

Equation (14) implies that the collection of all relative homographies of two planes across all views (after scaling and subtracting the identity) lies in a 3-dimensional linear subspace.

The extension to a constraint on homologies of multiple planes across multiple views is now trivial:

$$\hat{\mathcal{G}} = \begin{bmatrix} \hat{\mathcal{H}}_1 \\ \hline \vdots \\ \hline \hat{\mathcal{H}}_P \end{bmatrix}_{9P\times F} = \begin{bmatrix} \hat{\mathcal{N}}_1 \\ \hline \vdots \\ \hline \hat{\mathcal{N}}_P \end{bmatrix}_{9P\times 3} \left[\, \vec{v}^1 \quad \dots \quad \vec{v}^F \,\right]_{3\times F}. \qquad (15)$$

The dimensionality of the matrices on the right-hand side of (15) implies that the matrix $\hat{\mathcal{G}}$ is of rank 3 at most.

5 Multiview Constraints on Homographies of Multiple Planes

So far, we derived multiview linear subspace constraints only on the parameters of *homologies* (relative homographies) of pairs of planes. Next, we show how we can derive multiview linear subspace constraints on parameters of the *homographies* of the individual planes.

Let A_p^f be the homography of plane p at view f (with $p = 1 \dots P$ and $f = 1 \dots F$). This can be estimated only up to an unknown scale factor γ_p^f, i.e., $\tilde{A}_p^f = \gamma_p^f A_p^f$.

Unlike the case of homologies H_p^f, where we could uniquely recover the scale factors λ_p^f (Section 4), in the case of homographies A_p^f, the scale factor γ_p^f cannot be uniquely recovered. However, we can use the procedure described in Section 4 in order to recover all the scale factors $\{\gamma_p^f\}_{p=1}^P$ of all homographies corresponding to a single view f up to a *single consistent view-dependent* scale factor. This is shown next.

We arbitrarily choose a plane r to be a reference plane and estimate all corresponding homologies $\tilde{H}_p^f$ using (6). Putting together (6) and (10), we get

$$\tilde{H}_p^f = \frac{\gamma_p^f}{\gamma_r^f} (A_r^f)^{-1} A_p^f = \lambda_p^f H_p^f.$$

From this, it can be easily seen that

$$\frac{\gamma_p^f}{\gamma_r^f} = \lambda_p^f.$$

We can recover the scale factor λ_p^f of each homology using the procedure described in Section 4. Next, by dividing each computed homography $\tilde{A}_p^f$ by the recovered scale factor λ_p^f of its corresponding homology $\tilde{H}_p^f$, we get

$$\hat{A}_p^f = \gamma_r^f A_p^f,$$

i.e., all homographies $\hat{A}_p^f$ at each view are now determined up to a single view-dependent scale factor γ_r^f. Using (5), we can write

$$\hat{A}_p^f = B^f + \vec{v}^f \vec{m}_p,$$

where $B^f = \gamma_r^f C^f R C^{-1}$, $\vec{v}^f = \gamma_r^f C^f \vec{t}$ is the (scaled) epipole and $\vec{m}_p = \vec{n}_p^T C^{-1}$ (See (5) for notations). This leads to an extension of the multiplane constraint of [22] to a multiplane *multiview* linear subspace constraint on the "normalized" homographies $\hat{A}_p^f$ of the individual planes:

$$\hat{\mathcal{A}} = \left[\begin{array}{c} \hat{a}_1^1 \mid \cdots \mid \hat{a}_P^1 \\ \hline \vdots \\ \hline \hat{a}_1^F \mid \cdots \mid \hat{a}_P^F \end{array}\right]_{9F\times P} = \left[\begin{array}{cccc} & \vec{v^1} & \vec{0} & \vec{0} \\ \vec{b^1} & \vec{0} & \vec{v^1} & \vec{0} \\ & \vec{0} & \vec{0} & \vec{v^1} \\ \hline & & \vdots & \\ & & \vdots & \\ \hline & \vec{v^F} & \vec{0} & \vec{0} \\ \vec{b^F} & \vec{0} & \vec{v^F} & \vec{0} \\ & \vec{0} & \vec{0} & \vec{v^F} \end{array}\right]_{9F\times 4} \begin{bmatrix} 1 & \cdots & 1 \\ \vec{m}_1 & \cdots & \vec{m}_P \end{bmatrix}_{4\times P}, \tag{16}$$

where $\hat{a}_p^f$ is the ninth vector corresponding to the 3×3 homography matrix $\hat{A}_p^f$ and $\vec{b^f}$ is the ninth vector corresponding to the 3×3 homography matrix B^f. The dimensionality of the matrices on the right-hand side of (16) implies that the matrix $\hat{\mathcal{A}}$ is of rank 4 at most. Hence, the collection of all normalized homographies $\hat{A}_p^f$ of all planes across all views lies in a 4-dimensional linear subspace.

Note that, in the matrix $\hat{\mathcal{A}}$ of (16), each column corresponds to a single plane and each 9-row block corresponds to a single view. The rank-4 constraint on general homographies is useful only when each dimension of the matrix $\hat{\mathcal{A}}$ ($9F\times P$) exceeds 4, i.e., when the number of planes is larger than 4. This is in contrast to the subspace constraints on the homologies ((11), (13), (14), (15)), where each column corresponds to a view and each 9-row block corresponds to a plane. The constraints on homologies are therefore useful even for a single pair of planes as long as the number of views is greater than 4, which is often less restrictive than the requirement for multiple planes (> 4).

6 Multiview Constraints on a Single Plane

All the derived multiview subspace constraints (on homologies or homographies) assumed that the scene contains at least two planes. The derived subspace constraints are useful when the dimensionality of the subspace (e.g., 4) is significantly lower than the dimensionality of the original space of homographies (i.e., 9). Such reduced-dimension subspace constraints *cannot* be derived in general for multiple views of a scene containing a *single* plane (i.e., the $9\times F$ matrix obtained by putting together all homography vectors of that plane such that each column corresponds to one view does not necessarily have a rank lower than 9). This is because, for a general homography, according to (5)

$$\begin{aligned} A^f &= C^f(R^f + t^f n^T)C^{-1} \\ &= C^f R^f C^{-1} + (C^f t^f)(n^T C^{-1}) \\ &= B^f + u^f m^T, \end{aligned} \tag{17}$$

where $B^f = C^f R^f C^{-1}$, $\vec{u^f} = C^f \vec{t^f}$ and $\vec{m^T} = \vec{n^T} C^{-1}$. Since, in the uncalibrated case, B^f can vary from one view to another in an unconstrained manner (e.g., each view can have a different rotation matrix R and/or a different calibration matrix C^f), this already spans a 9-dimensional linear space (when putting all homography vectors in a matrix where each column corresponds to a single view). In the previous sections, we showed how, by looking at the *relative* motion of a pair of planes, we can eliminate all effects of rotation and changes in calibration (see (7)), hence leaving only 4 degrees of freedom for each new view, resulting in a rank-4 constraint on all "relative homographies" (homologies).

However, for some restricted but common cases of camera motion, one can still get linear subspace constraints even for the case of a single plane. We first show a low-dimensionality constraint for the case of pure translation with fixed (but unknown) camera calibration (Section 6.1), and then show a similar constraint for the case when small camera rotation is also allowed (Section 6.2).

6.1 Pure Translation

Let π be a planar surface in the scene with plane normal $\vec{n}$. Let A^f denote its corresponding homography between the reference image J and an image J^f ($f = 1, \ldots, F$). Assuming the camera performs only translational motion and assuming that the (unknown) camera calibration is *fixed* across all views, (17) reduces to

$$A^f = I + \vec{u^f}\vec{m^T}. \tag{18}$$

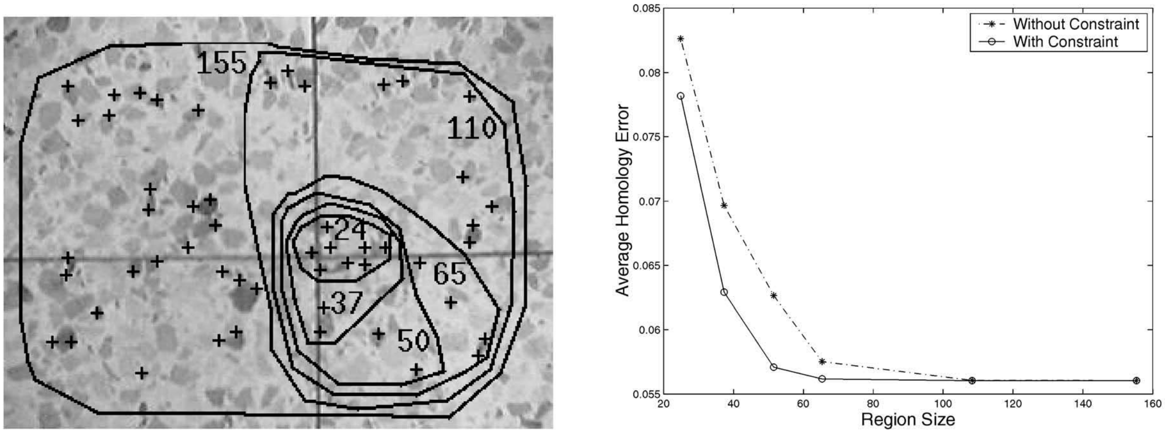
Note that $\vec{m}$ is only plane-dependent, (i.e., is common to all views for the plane π), whereas $\vec{u^f}$ is only view-dependent (here $\vec{u^f} = C\vec{t^f}$, since $C^f \equiv C$ for all views). The homography for the case of pure translation (18) has the same algebraic structure as the "relative homography" in (17) since, in both cases, the rotation is canceled. Nevertheless, here we had to assume constant calibration for all views; whereas, in the multiplane case, the calibration could vary for each view. Here, on the other hand, we are examining homographies of a single plane; whereas, there, we examined homologies (*relative* homographies) of pairs of planes.

Hence, the same rank-4 constraint that was derived in Section 2 applies for the set of all homographies of a *single* plane across multiple views, when there is *no* camera rotation and the camera calibration is constant. Furthermore, this also implies that, for this restricted case, a rank-4 constraint applies directly to the collection of all homographies of *multiple planes* across *multiple-views*, as was true for the homologies of multiple planes (This, of course, requires the proper scale normalization as was done in Section 3, only this time there is no need for a selected reference plane).

6.2 Small Camera Rotation

Next, we show that subspace constraints on the homographies of a single plane can be derived when rotation is also introduced, but is restricted to small angles. This, however, increases the dimensionality from 4 to 7 (but is still smaller than 9).

The small camera rotation assumption implies that the rotation matrix R of (3) has the form [5]



(a)

(b)

Fig. 1. A quantitative comparison of homology estimation with and without enforcing the multiview subspace constraint of (11). (a) Shows the image (of floor tiles) from which the 19-frame sequence was synthetically generated with ground-truth known homologies. The tracked feature points are marked in black crosses. The different regions-of-interest in which homography estimation was applied are marked by black curves. The numbers correspond to the maximum distance between any two feature points within each region. (b) The graph displays the average error in homology estimation of all frames in the sequence as a function of the size of the region of interest. The error for each 3×3 homology H is defined as: $\| I - H^{-1}_{Ground\,Truth} H \|$, where I is with the 3×3 identity matrix. The smaller the region is, the larger the initial error is (dashed curve). Enforcing the multiview subspace constraint of (11) improves the accuracy, especially in small regions (solid curve).

$$R = \begin{bmatrix} 1 & -\Omega_Z & \Omega_Y \\ \Omega_Z & 1 & -\Omega_X \\ -\Omega_Y & \Omega_X & 1 \end{bmatrix}. \tag{19}$$

Now, assuming the calibration is *fixed* for all views, the homography matrix A^f has the form

$$A^f = C(R^f + \vec{t}^f \vec{n}^T)C^{-1}. \tag{20}$$

Therefore,

$$C^{-1}A^f C = R^f + \vec{t}^f \vec{n}^T. \tag{21}$$

Under these assumptions, each additional frame adds 7 new degrees of freedom: three translation parameters ($\vec{t}^f$), three rotation parameters ($\Omega^f_X, \Omega^f_Y, \Omega^f_Z$), and 1 scale factor λ^f (since A^f can be recovered only up to a scale factor). Rearranging the components of the homography matrix A^f in a single (9×1) column vector a^f, we can rewrite (21) as

$$M\vec{a}^f = \mathcal{N}_{9\times7} \begin{bmatrix} \vec{t}^f \\ \vec{\Omega}^f \\ 1 \end{bmatrix}, \tag{22}$$

where

$$\mathcal{N} = \begin{bmatrix} \vec{n} & \vec{0} & \vec{0} & \vec{0} & \vec{i}_3 & -\vec{i}_2 & \vec{i}_1 \\ \vec{0} & \vec{n} & \vec{0} & -\vec{i}_3 & \vec{0} & \vec{i}_1 & \vec{i}_2 \\ \vec{0} & \vec{0} & \vec{n} & \vec{i}_2 & -\vec{i}_1 & \vec{0} & \vec{i}_3 \end{bmatrix}, \tag{23}$$

$\vec{i}_1, \vec{i}_2, \vec{i}_3$ are as defined in (9), $\vec{\Omega}^f = [\Omega^f_X, \Omega^f_Y, \Omega^f_Z]^T$, and M is a 9×9 matrix which depends on the camera intrinsic parameters (of C). Note that M is invertible.[4]

Since the matrix $\mathcal{N}$ depends only on plane normal parameters, it is common to all views $f = 1 \ldots F$ whose homographies are estimated relative to the reference frame J. Similarly, since the calibration is fixed, the matrix M is also common to all views. We assume that the homographies $\tilde{A}^f_p = \lambda^f A^f_p$ have been computed and are each known up to a scale factor. Hence, we can stack all the computed homography vectors in a $9 \times F$ matrix $\mathcal{A}$, where each column corresponds to a single image view J^f (relative to the reference view J):

$$\begin{aligned} [\mathcal{A}]_{9\times F} &= \begin{bmatrix} \vec{\tilde{a}}^1 & \cdots & \vec{\tilde{a}}^F \end{bmatrix}_{9\times F} \\ &= [M^{-1}]_{9\times9} [\mathcal{N}]_{9\times7} \begin{bmatrix} \vec{t}^1 & \cdots & \vec{t}^F \\ \vec{\Omega}^1 & \cdots & \vec{\Omega}^F \\ 1 & & 1 \end{bmatrix}_{7\times F} \\ & \begin{bmatrix} \lambda^1 & & 0 \\ & \ddots & \\ 0 & & \lambda^F \end{bmatrix}_{F\times F}. \end{aligned}$$

The dimensionality of the matrices on the right-hand side of (24) implies that the matrix $\mathcal{A}$ is of rank 7 at most. Hence, the collection of all homographies of a plane across all images (for the case of small camera rotation, arbitrary translation, and fixed calibration) resides in a 7-dimensional linear subspace.

7 Empirical Evaluation

Next, we explore the utility of the derived multiview constraints. The purpose here is to examine the strength and potential uses of these constraints and *not* to propose a particular algorithm.

[4] Proof: Let $B = C^{-1}AC$. Since C^{-1} and C are regular matrices, we get that $B = 0$ iff $A = 0$. Rearranging the (3×3) matrix B into a (9×1) column vector $\vec{b}$, we get $\vec{b} = M\vec{a}$. If M were singular, then $\exists \vec{a} \neq 0$ such that $\vec{b} = M\vec{a} = 0$. This implies that the corresponding $B = 0$, while the corresponding $A \neq 0$. This, however contradicts the observation that $B = 0$ iff $A = 0$, hence, M must be a nonsingular matrix.

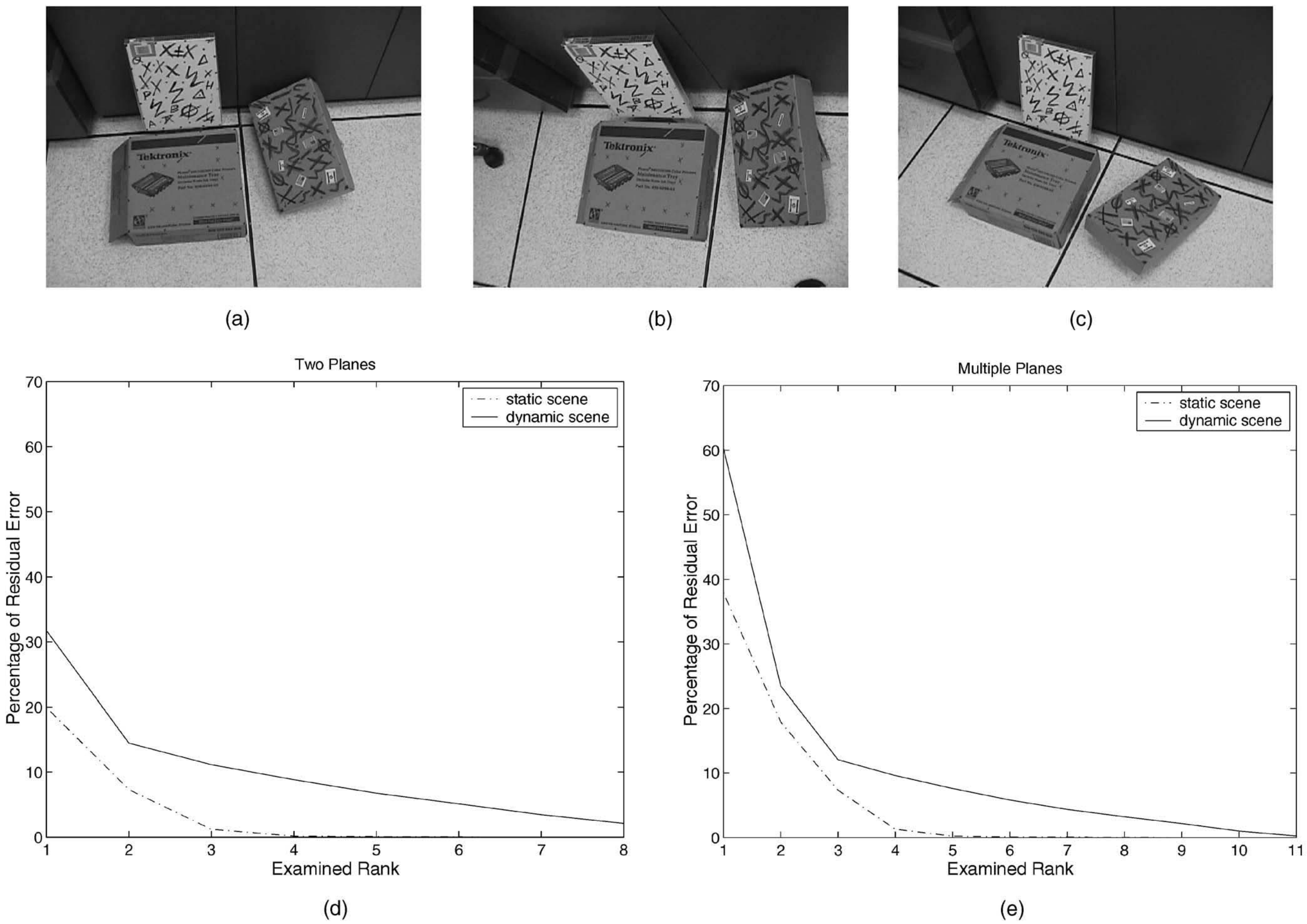


Fig. 2. Detection of nonrigid motion: (a), (b), and (c) Sample images from a sequence of 20 images, containing three different planes (the three boxes). In the first nine images, all planes are static and only the camera moves, i.e., the planes move rigidly with respect to each other (Fig. 2a and 2b). In images 10 to 20, the right box moved relative to the other two planes (Fig. 2c). (d) Displays the results of nonrigid motion detection for the case of two planes using (11). The relative homographies are estimated for the right box where the white box is used as a reference plane. The graph shows the percentage of error for each rank of the relative-homographies matrix $\mathcal{H}$ of (11), once for the static scene (dashed curve) and once for the dynamic scene (solid curve). (e) Displays results of nonrigid motion detection for the case of multiple planes using (13). The relative homographies are estimated for the two brown boxes and the white box is used as a reference plane. The graph displays the percentage of error for each of the relative-homographies matrix $\mathcal{G}$ of (13), once for the static scene and once for the moving object scene. In both cases, (d) and (e), the error is practically zero at rank 4 for the static scene (dashed curve); whereas, for the dynamic scene, the rank 4 constraint is clearly violated (solid curve). See text for more details.

7.1 Constrained Homography Estimation

Homography estimation techniques produce reliable homography parameters when the planar surface captures a large image region. However, they tend to be highly inaccurate when applied to small image regions [25], as is often the case in scenes with *multiple* planar surfaces. While each independent homography computation is unreliable, all relative homographies of all pairs of planes across all views, must satisfy the multiview subspace constraints derived above, i.e., they must all reside in a 4-dimensional (or even smaller) linear subspace. These multiview constraints can therefore be used as additional constraints to compensate for insufficient spatial information, leading to more accurate homography estimation. This is illustrated next for the subspace constraints derived in (11).

Fig. 1 shows a quantitative comparison of homography estimation with and without enforcing the multiview subspace constraint of (11). A synthetic sequence was generated by geometrically warping a reference image by known ground truth homologies (The case of pure translation results in homologies even for a single plane. The "reference plane" in this case is the plane at infinity. This is further elaborated in Section 6.1. Feature points were automatically selected in the reference image (marked points in Fig. 1a) and were tracked using the KLT algorithm [15], [2]. The homologies were estimated between the reference image and each of the other images using Least-Squares fit to the computed point correspondences. They were then compared against the ground truth homologies and the error for each 3×3 homology H was defined as $||I - H^{-1}_{GroundTruth}H||$, where I is the 3×3 identity matrix. The experiment was repeated for varying region sizes, by taking the feature points from within different regions of interest (also marked in Fig. 1a). The smaller the region is, the larger the error is (see dashed curve in Fig. 1b). All these homologies where then stacked into a $9 \times F$ matrix $\mathcal{H}$ (see (11)). The columns of $\mathcal{H}$ were then projected onto the closest 4-dimensional linear subspace. Enforcing the multiview

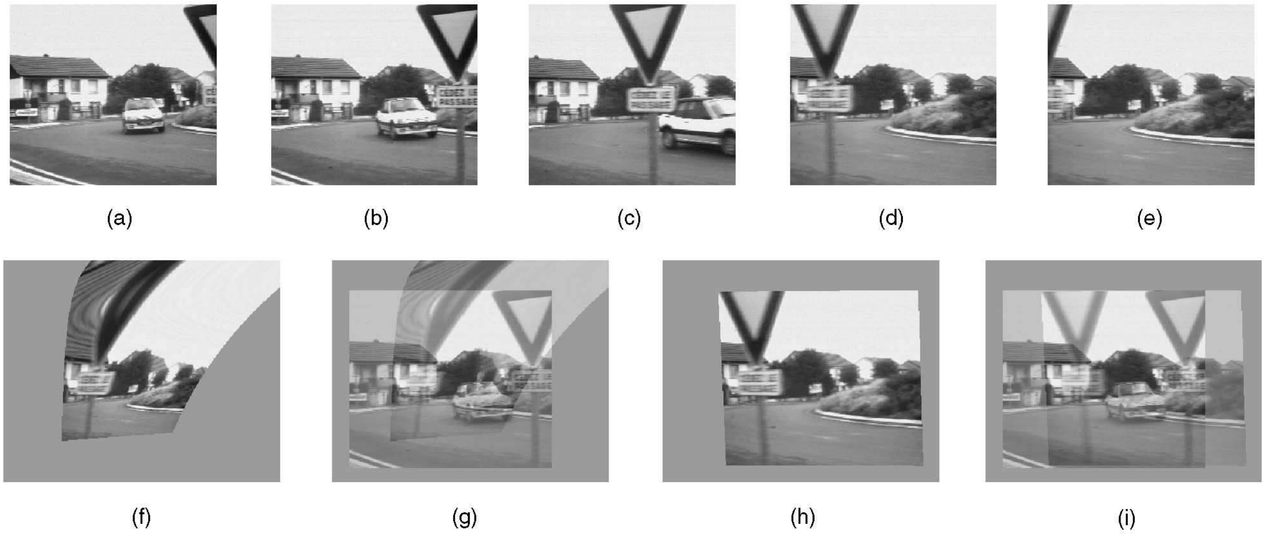


Fig. 3. Constrained multiframe homography estimation. (a), (b), (c), (d), and (e) Sample frames from a sequence of 34 frames. Image (b) was used as the reference frame. (f) Poor alignment resulting from bad two-frame homography estimation of the house region between the reference frame (b) and frame (d). The frame was completely distorted because the house region is small and was significantly occluded by the road sign. (g) Shows the same result overlayed on top of the reference image (b). (h) Good alignment as a result of applying the constrained multiframe homography estimation. The house is now well aligned even though only a small portion of the house is visible (see overlay image (i)), while the rest of the image is not distorted. The road sign is not aligned because it is at a different depth and displays accurate $3D$ parallax.

subspace constraint reduced the errors significantly for small regions (see solid curve in Fig. 1b). As expected, when the region is small, the improvement is more significant than when it is large.

The subspace projection was done by applying SVD to the matrix $\mathcal{H}$ of (11) and setting to zero all but the four largest singular values. To guarantee uniform error in all matrix entries before subspace projection and to further condition the numerical process, we first applied coordinate normalization similar to that suggested by Hartley [9].

7.2 Detecting Nonrigid Motion

The multiview subspace constraints on multiple planes are true only for planes moving rigidly with respect to each other. Planar surfaces with different $3D$ motions will not necessarily comply with these constraints. This can be used as an additional cue for detecting independently moving objects.

Given two planar surfaces (π_r and π_p), we can construct the matrix $\mathcal{H}$ of their relative homographies (see (11)) and examine its rank. If $rank(\mathcal{H}) > 4$ (beyond reasonable noise), then the two planes cannot be moving rigidly with respect to each other. Note that this is a *sufficient* condition, but not a necessary one. Fig. 2 displays a comparison of the rank of the relative-homography matrix $\mathcal{H}$, once constructed from homographies of two rigidly moving planes, and once constructed from homographies of two nonrigidly moving planes. The sequence contains 20 images taken from different view points and different orientations (Figs. 2a, 2b, and 2c). The scene contains three different planar surfaces, which are rigid with respect to each other in images 1 to 9 (Figs. 2a and 2b). In images 10-20, one of the planes moved relative to the other two planes (the right box in Fig. 2c). The graph in Fig. 2d shows the percentage of residual error ("noise") assuming different ranks of the relative-homographies matrix $\mathcal{H}$. The error is defined as

$$\sqrt{\sum_{i=r+1}^{n} S_i^2} / \sqrt{\sum_{i=1}^{n} S_i^2},$$

where S_i are the singular values of $\mathcal{H}$, r is the assumed rank of $\mathcal{H}$, and n is the total number of singular values. The dashed curve shows the decay of error as a function of the rank for the case of two rigid boxes (frames 1-9), and the solid curve shows the decay of error as a function of rank for the same two boxes when they move nonrigidly (frames 10-20) The rank of $\mathcal{H}$ for the static scene is clearly no more than 4; whereas, for the moving plane case, it is well above 4.

Similarly, the multiplane, multiview subspace constraint of (13) is also true only for static scenes and is violated in the dynamic scene. Fig. 2e displays a comparison of the rank of the relative-homography matrix $\mathcal{G}$ once constructed from the homographies of the three rigidly moving planes and once constructed from the homographies of the three nonrigidly moving planes. The graph displays the percentage of error for each possible rank of the relative-homographies matrix of the three boxes for the static scene subsequence and for the moving object subsequence. The rank of $\mathcal{G}$ for the static scene is clearly no more than 4 (dashed curve); whereas, for the moving plane case, it is well above 4 (solid curve).

7.3 Homography Estimation for a Single Plane

We have successfully developed and implemented an end-to-end algorithm, which takes advantage of the multiview subspace constraints on homographies of a *single* plane across multiple views, for the simpler and more restricted cases discussed in Section 6. Such assumptions are valid for multiple consecutive frames in *short* video segments (where the assumption of small camera rotation is still valid). The algorithm we developed [29] combines the multiframe linear subspace constraints with "direct" methods, leading to a *simultaneous multiframe constrained homography estimation* algorithm directly from brightness variations across multiple frames. We refer the reader to [29] for more details. Fig. 3 shows a comparison of applying *two*-frame homography estimation and *multi*frame constrained homography estimation on small image regions, under the assumption of instantaneous camera motion. The sequence contains 34 frames taken by a moving camera. Because the moving camera is imaging the scene from a short distance, different planar surfaces (e.g., the house, the road-sign, etc.) induce different homographies. In frames where the house was not occluded, the *two*-frame homography estimation (when applied only to the house region) aligned the house region reasonably well. However, in frames where only a small portion of the house was visible (e.g., when the house was partially occluded by the road-sign and was not fully in the camera field of view), the quality of the alignment resulting from two-frame homography estimation degraded drastically (see Figs. 3f and 3g). The alignment resulting from the multiframe constrained homography estimation, on the other hand, successfully aligned the house, even in frames where only a small portion of the house was visible (see Figs. 3h and 3i). For further details, see [29].

8 Concluding Remarks

In this paper, we showed that, for scenes containing multiple (at least two) planes, multiview linear subspace constraints on their homographies can be derived. Similar constraints can also be derived for scenes containing a single planar surface under restricted assumptions on camera motion (which are usually valid in short video sequences). We further showed that these constraints can potentially be used to improve homography estimation of existing algorithms and can serve as an additional cue for detection of nonrigid motion.

Acknowledgments

This was supported by the Israeli Ministry of Science (Grant no. 1229).

References

[1] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation," *Proc. European Conf. Computer Vision,* pp. 237–252, May 1992.

[2] S. Birchfield, "KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker," http://robotics.stanford.edu/~birch/klt.

[3] S. Carlsson and D. Weinshall, "Dual Computation of Projective Shape and Camera Positions from Multiple Images," *Int'l J. Computer Vision,* vol. 27, no. 3, 1998.

[4] A. Criminisi, I. Reid, and A. Zisserman, "Duality, Rigidity and Planar Parallax," *Proc. European Conf. Computer Vision,* pp. 846–861, June 1998.

[5] J.Q. Fang and T.S. Huang, "Solving Three-Dimensional Small-Rotation Motion Equations," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 253-258, 1983.

[6] O. Faugeras, *Three-Dimensional Computer Vision—A Geometric Viewpoint.* Cambridge, Mass.: MIT Press, 1996.

[7] A.W. Fitzgibbon and A. Zisserman, "Automatic Camera Recovery for Closed or Open Image Sequences," *Proc. European Conf. Computer Vision,* pp. 310–326, 1998.

[8] L. Van Gool, L. Proesmans, and A. Zisserman, "Grouping and Invariants Using Planar Homologies," *Proc. Workshop Geometrical Modeling and Invariants for Computer Vision,* 1995.

[9] R.I. Hartley, "In Defence of the 8-Point Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19. no. 6, pp. 580–593, June 1997.

[10] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, ISBN: 0521623049, 2000.

[11] M. Irani, P. Anandan, and D. Weinshall, "From Reference Frames to Reference Planes: Multi-View Parallax Geometry Application," *Proc. European Conf. Computer Vision,* pp. 829–845, 1998.

[12] M. Irani, B. Rousso, and S. Peleg, "Computing Occluding and Transparent Motions," *Int'l J. Computer Vision,* vol. 12, no. 1, pp. 5–16, Jan. 1994.

[13] M. Irani, P. Anandan, and S. Hsu, "Mosaic Based Representations of Video Sequences and Their Applications," *Proc. Int'l Conf. Computer Vision,* pp. 605–611, Nov. 1995.

[14] K. Kanatani, "Optimal Homograhpy Computation with a Reliability Measure," *Proc. IAPR Workshop Machine Vision,* pp. 17–19, Nov. 1998.

[15] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique With an Application to Stereo Vision," *Proc. Image Understanding Workshop,* pp. 121–130, 1981.

[16] J. Ma and N. Ahuja, "Dense Shape and Motion from Region Correspondences by Factorization," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 219–224, 1998.

[17] P. Pritchett and A. Zisserman, "Matching and Reconstruction from Widely Separated Views In 3D Structure from Multiple Images of Large-Scale Environments," Lecture Notes in Computer Science 1506, Springer-Verlag, pp. 219–224, 1998.

[18] Q.T. Luong and O. Faugeras, "Determining the Fundamental Matrix with Planes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 489–494, June 1998.

[19] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C.* Cambridge, Mass.: Cambridge Univ. Press, 1992.

[20] J.G. Semple and G.T. Kneebone, *Algebraic Projective Geometry.* New York: Oxford Univ. Press, 1952.

[21] A. Shashua, "Projective Structure from Uncalibrated Images: Structure from Motion and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, pp. 778–790, 1994.

[22] A. Shashua and S. Avidan, "The Rank 4 Constraint in Multiple (≥ 3) View Geometry," *Proc. European Conf. Computer Vision,* pp. 196–206, 1996.

[23] D. Sinclair, H. Christensen, and C. Rothwell, "Using the Relation Between a Plane Projectivity and the Fundamental Matrix," *SCIA,* pp. 181–188, 1996.

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

[25] R. Szeliski and P.H.S. Torr, "Geometrically Constrained Structure from Motion: Points on Planes," *Proc. European Workshop 3D Structure from Multiple Images of Large-Scale Environments,* pp. 171–186, 1998.

[26] B. Triggs, "Autocalibration from Planar Scenes," *Proc. European Conf. Computer Vision,* pp. 89–105, 1998.

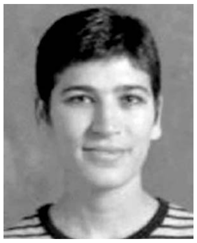
[27] B. Triggs, "Plane + Parallax, Tensors and Factorization," *Proc. European Conf. Computer Vision,* pp. 522–538, 2000.

[28] T. Vieville, C. Zeller, and L. Robert, "Using Collineations to Compute Motion and Structure in an Uncalibrated Image Sequence," *Int'l J. Computer Vision,* vol. 20, pp. 213–242, 1996.

[29] L. Zelnik-Manor and M. Irani, "Multi-Frame Estimation of Planar Motion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1105–1116, Oct. 2000.

[30] L. Zelnik-Manor and M. Irani, "Multi-View Subsapce Constraints on Homograhpies," *Proc. Int'l Conf. Computer Vision,* pp. 710–715, Sept. 1999.

[31] Z. Zhang, "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations," *Proc. Int'l Conf. Computer Vision,* Sept. 1999.



Lihi Zelnik-Manor received the BSc degree in mechanical engineering from the Technion, Israel, in 1995, where she graduated summa cum laude. In 1998, she recieved the MSc degree with honors in computer science from the Weizmann Institute of Science, Israel. Today she is studying toward a PhD degree in the Department of Computer Science and Applied Mathematics in the Weizmann Institute of Science, Rehovot, Israel. Her research focuses on the analysis of video sequences and its applications.

Michal Irani received the BSc degree in mathematics and computer science in 1985 and the MSc and PhD degrees in computer science in 1989 and 1994, respectively), all from the Hebrew University of Jerusalem. During 1993-1996, she was a member of the technical staff in the Vision Technologies Laboratory at David Sarnoff Research Center, Princeton, NJ. Dr. Irani is currently a member of the faculty of the Computer Science and Applied Mathematics department at the Weizmann Institute of Science, Israel. Her research interests are in the area of computer vision and video information analysis. Her work includes analysis of space-time visual information, motion analysis, 3D scene reconstruction, compact representations of video sequences, analysis of dynamic events, video enhancement, synthesis and visualization, video compression, video browsing and indexing, and multisensor alignment and fusion. Dr. Irani received the David Sarnoff Research Center Technical Achievement Award in 1994 and the Yigal Allon three-year fellowship for outstanding young scientists in 1998. She received the best paper award at the 2000 European Conference on Computer Vision (ECCV) and the honorable mention for the Marr Prize at the 2001 IEEE International Conference on Computer Vision (ICCV). She is a member of the IEEE and the IEEE Computer Society and an associate editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence*.