Multi-View Subspace Constraints on Homographies

Lihi Zelnik-Manor   **Michal Irani**
Dept. of Computer Science and Applied Math
The Weizmann Institute of Science
76100 Rehovot, Israel
Email: {lihi,irani}@wisdom.weizmann.ac.il

Abstract

The motion of a planar surface between two camera views induces a homography. The homography depends on the camera intrinsic and extrinsic 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 motion of multiple (≥ 2) planes across multiple views.

The paper has three main contributions: (i) We show that the collection of all relative homographies of a *pair* of planes (homologies) across *multiple* views, spans a 4-dimensional linear subspace. (ii) We show how this constraint can be extended to the case of *multiple planes* across *multiple views.* (iii) We suggest two potential application areas which can benefit from these constraints: (a) The accuracy of homography estimation can be improved by enforcing the multi-view subspace constraints. (b) Violations of these multi-view constraints can be used as a cue for moving object detection. All the results derived in this paper are true for *uncalibrated cameras.*

1 Introduction

Homography estimation is used for $3D$ analysis [8, 9, 11, 4, 2, 6, 7], mosaicing [5], camera calibration [12], and more. The induced homography between a pair of views depends on the camera intrinsic and extrinsic parameters, and on the $3D$ plane parameters [1]. 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 multi-view *linear* subspace constraints on the relative motion of multiple (≥ 2) planes.

Linear subspace constraints on homographies have been previously derived by Shashua and Avidan [10]. They showed that the collection of homographies of *multiple planes* between a *pair of views*, spans a 4-dimensional linear subspace. This constraint requires the number of planes in the scene to be greater than 4. In this paper we first derive a "dual" constraint, for a *pair* planes over *multiple (> 4) views* (Section 3). This constraint is then extended to a constraint on homographies of *multiple planes* across *multiple views* (Section 4).

Algorithms for $3D$ analysis which are based on the use of multiple homographies (in scenes with *multiple planes*) have been suggested (e.g., [8, 9, 13, 7]). Most of these algorithms rely on *accurate* precomputation of the homographies. However, in scenes containing multiple planes, the image region corresponding to each plane may be small. In such cases, the homography estimation tends to be highly inaccurate [11] (i.e, when applied to small image regions). In this paper we show how the accuracy of homography estimation can be improved by employing the multi-view subspace constraints (Section 5.1). We also show how *violations* of these multi-view constraints can be used as a cue for moving object detection (Section 5.2).

All the results derived in this paper are true for *uncalibrated camera.*

2 Homographies - Basic Notations

Let $\vec{Q} = (X, Y, Z)^t$ and $\vec{Q'} = (X', Y', Z')^t$ denote a scene point with respect to two different camera views, respectively. Let $\vec{q} = (x, y, 1)^t$ and $\vec{q'} = (x', y', 1)^t$ denote the corresponding points in the two images. We can write:

$$\vec{q} \cong C\vec{Q} \quad , \quad \vec{q'} \cong C'\vec{Q'} \qquad (1)$$

where $\cong$ denotes equality up to a scale factor. C and C' are 3×3 matrices [1] (composed of camera internal parameters and projection).

Let π be a planar surface with plane normal $\vec{n}$, then $\vec{n}^t\vec{Q} = 1$ for all points $\vec{Q} \in \pi$ ($\vec{n} = \frac{\vec{m}}{d_\pi}$, where $\vec{m}$ 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 transformation between the $3D$ coordinates of a scene point $Q \in \pi$ in the two views, can be expressed by:

$$\vec{Q'} = G\vec{Q} \qquad (2)$$

where

$$G = R + \vec{t}\vec{n}^t \qquad (3)$$

R is the rotation matrix and $\vec{t}$ is the translation of the camera. Therefore, the induced transformation between the corresponding *image* points is:

$$\vec{q'} \cong H\vec{q} \tag{4}$$

where

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

is the induced homography between the two views of the plane π. From Eq. (4) it is clear that when H is computed from image point correspondences, it can be estimated only up to a scale factor.

3 Multi-View Two-Plane Constraint

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

$$H_{pr}^f = (H_r^f)^{-1} H_p^f \tag{6}$$

This is also known as a "plane homology" [3, 7]. Some properties and invariants of planar homologies have been discussed in [3], and used in [7]. Here we present a different set of constraints on homologies.

Using Eq. (5) and the Sherman-Morisson formula[1] [14], it can be shown that, for rigidly moving planes π_r and π_p, the matrix H_{pr}^f has the form:

$$\begin{aligned} H_{pr}^f &= I + \vec{v}^f \vec{n}_{pr}^t \\ &\equiv \begin{bmatrix} 1+h_1 & h_2 & h_3 \\ h_4 & 1+h_5 & h_6 \\ h_7 & h_8 & 1+h_9 \end{bmatrix} \end{aligned} \tag{7}$$

where $\vec{v}^f = C \frac{{R^f}^{-1} \vec{t^f}}{1 + \vec{n_r}^t {R^f}^{-1} \vec{t^f}}$, $\vec{n}_{pr} = (\vec{n_p} - \vec{n_r})\, C^{-1}$, I is a 3×3 identity matrix and $\vec{t^f}, R^f$ are the camera translation and the camera rotation matrix, between the reference image J and the image K^f. C is the camera projection matrix at the reference view J. Note that C^f (i.e., the projection matrix of K^f) is eliminated by the composition. Note that $\vec{v}^f$ is view-dependent, i.e., is common to all rigidly moving planes between a pair of views J and K^f, whereas $\vec{n}_{pr}$ is plane-dependent, i.e., is common to all views for a pair of planes π_r and π_p.

Rearranging the components of the relative-homography (3×3) matrix H_{pr}^f in a single (9×1) column vector $\vec{h}_{pr}^f$, we can rewrite Eq. (7) as:

$$\vec{h}_{pr}^f = \mathcal{N}_{pr} \begin{bmatrix} v_X^f \\ v_Y^f \\ v_Z^f \\ 1 \end{bmatrix} = \mathcal{N}_{pr} \begin{bmatrix} \vec{v}f \\ 1 \end{bmatrix} \tag{8}$$

where

$$\mathcal{N}_{pr} = \begin{bmatrix} n_{pr_X} & 0 & 0 & 1 \\ n_{pr_Y} & 0 & 0 & 0 \\ n_{pr_Z} & 0 & 0 & 0 \\ 0 & n_{pr_X} & 0 & 0 \\ 0 & n_{pr_Y} & 0 & 1 \\ 0 & n_{pr_Z} & 0 & 0 \\ 0 & 0 & n_{pr_X} & 0 \\ 0 & 0 & n_{pr_Y} & 0 \\ 0 & 0 & n_{pr_Z} & 1 \end{bmatrix} \tag{9}$$

In practice $\vec{h}_{pr}^f$ are estimated only up to an unknown scale factor λ_{pr}^f (see Eq. (4)). Hence, the *computed* relative homographies, denoted by $\tilde{\vec{h}}_{pr}^f$, are

$$\tilde{\vec{h}}_{pr}^f = \lambda_{pr}^f \vec{h}_{pr}^f \tag{10}$$

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

$$\begin{aligned} &[\mathcal{H}_{pr}]_{9\times F} = \begin{bmatrix} \tilde{\vec{h}}_{pr}^1 & \ldots & \tilde{\vec{h}}_{pr}^F \end{bmatrix}_{9\times F} = \\ &[\mathcal{N}_{pr}]_{9\times 4} \begin{bmatrix} \vec{v}^1 & \ldots & \vec{v}^F \\ 1 & & 1 \end{bmatrix}_{4\times F} \begin{bmatrix} \lambda_{pr}^1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{pr}^F \end{bmatrix} \end{aligned} \tag{11}$$

The dimensionality of the matrices on the right hand side of Eq. (11) implies that the matrix $\mathcal{H}_{pr}$ is of

[1] For a square matrix A, and two column vectors $\vec{u}$, $\vec{w}$, the Sherman-Morrison formula gives:

$$(A + \vec{u}\vec{w}^t)^{-1} = A^{-1} - \frac{(A^{-1}\vec{u})(\vec{w}^t A^{-1})}{I + \vec{w}^t A^{-1} \vec{u}}$$

rank 4 at most[2]. Hence the collection of all relative-homographies 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 [10]. 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 a *pair* (2) of planes over *multiple* (> 4) *views*.

4 Multi-View Multi-Plane Constraints

As explained above, homographies are determined only up to a scale factor. This scale factor differs for every pair of planes and for every pair of views. Therefore, the extension of the two-plane multi-view factorization (Section 3), or the two-view multi-plane factorization [10], into a *multi-view multi-plane* factorization is not straightforward. To extend the low-dimensionality linear subspace constraint to multiple-planes, we constrain the scale factors, denoted by λ^f_{pr}, to be a product of two scalars: one of which is view-dependent and one which is plane-dependent. This can be done with no calibration information.

Let $\pi_1, \ldots, \pi_P$ be P planar surfaces with normals $\vec{n}_1, \ldots, \vec{n}_P$, respectively. Let $H^f_1, \ldots, H^f_P$ be their corresponding homography matrices between the reference view J and the other views $K^f (f = 1 \ldots F)$. Let π_r be a *reference plane* (e.g., could be chosen as the plane occupying the largest image region in the reference image). Assuming the relative homographies $H^f_{pr} (f = 1, \ldots, F; p = 1, \ldots, P)$, with respect to the reference plane π_r and the reference image J, have been computed and are known up to a scale factor, we can arbitrarily set one of the six off-diagonal entries in the *relative homographies* H^f_{pr} to be equal to 1 (i.e., h_2, h_3, h_4, h_6, h_7 or h_8; See Eq. (7)), while the other entries are scaled accordingly. This results in a scale factor λ^f_{pr}, for the relative homographies, which can be factored into a bilinear product of two scalars:

$$\lambda^f_{pr} = \alpha^f \cdot \beta_p$$

where α^f is view-dependent, and β_p is plane-dependent (e.g., if we choose $h_3 = 1$, then we get: $\lambda^f_{pr} = \frac{1}{h_3} = \frac{1}{v^f_X} \cdot \frac{1}{n_{pr_Z}}$, i.e., $\alpha^f = \frac{1}{v^f_X}$ and $\beta_p = \frac{1}{n_{pr_Z}}$). 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 get from Eq. (11):

$$\mathcal{H} = \begin{bmatrix} \underline{\mathcal{H}_{1r}} \\ \vdots \\ \overline{\mathcal{H}_{Pr}} \end{bmatrix}_{9P \times F} = B \cdot \mathcal{N} \cdot V \cdot A \qquad (12)$$

where,

$$\begin{aligned} B &= \begin{bmatrix} \beta_1 & & 0 \\ & \ddots & \\ 0 & & \beta_P \end{bmatrix}_{9P \times 9P} \\ \mathcal{N} &= \begin{bmatrix} \underline{\mathcal{N}_{1r}} \\ \vdots \\ \overline{\mathcal{N}_{Pr}} \end{bmatrix}_{9P \times 4} \\ V &= \begin{bmatrix} \vec{v}^1 & \ldots & \vec{v}^F \\ 1 & & 1 \end{bmatrix}_{4 \times F} \\ A &= \begin{bmatrix} \alpha^1 & & 0 \\ & \ddots & \\ 0 & & \alpha^F \end{bmatrix}_{F \times F} \end{aligned} \qquad (13)$$

The dimensionality of the matrices on the right hand side of Eq. (12) implies that, the matrix $\mathcal{H}$ is of rank 4 at most.

This implies that when solving for the homographies while consistently setting one of the six off-diagonal entries of the relative homographies to be 1, we are guaranteed that the collection of all relative homographies, of *all planes* across *all views*, lies in a 4-dimensional linear subspace. This scaling of the relative homographies is possible only when at least one of the six off-diagonal entries is different from zero for all planes, in all views. An example where this fails to exist is the identity matrix, which is the case of no motion.

5 Applications

In this section we present two different potential uses of the multi-view subspace constraints presented in Sections 3 and 4. In Section 5.1 it is shown how the accuracy of two-view homography estimation can be improved by constraining it with information from multiple images, using the multi-view subspace constraints. In Section 5.2 we show how *violations* of the multi-view subspace constraints can be used as a cue for detection of moving objects.

The purpose of this section is to convey the strength and potential use of these constraints, and *not* to present a particular algorithm.

5.1 Constrained Homography Computation

Homography estimation techniques perform well when the planar surface captures a large image region.

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

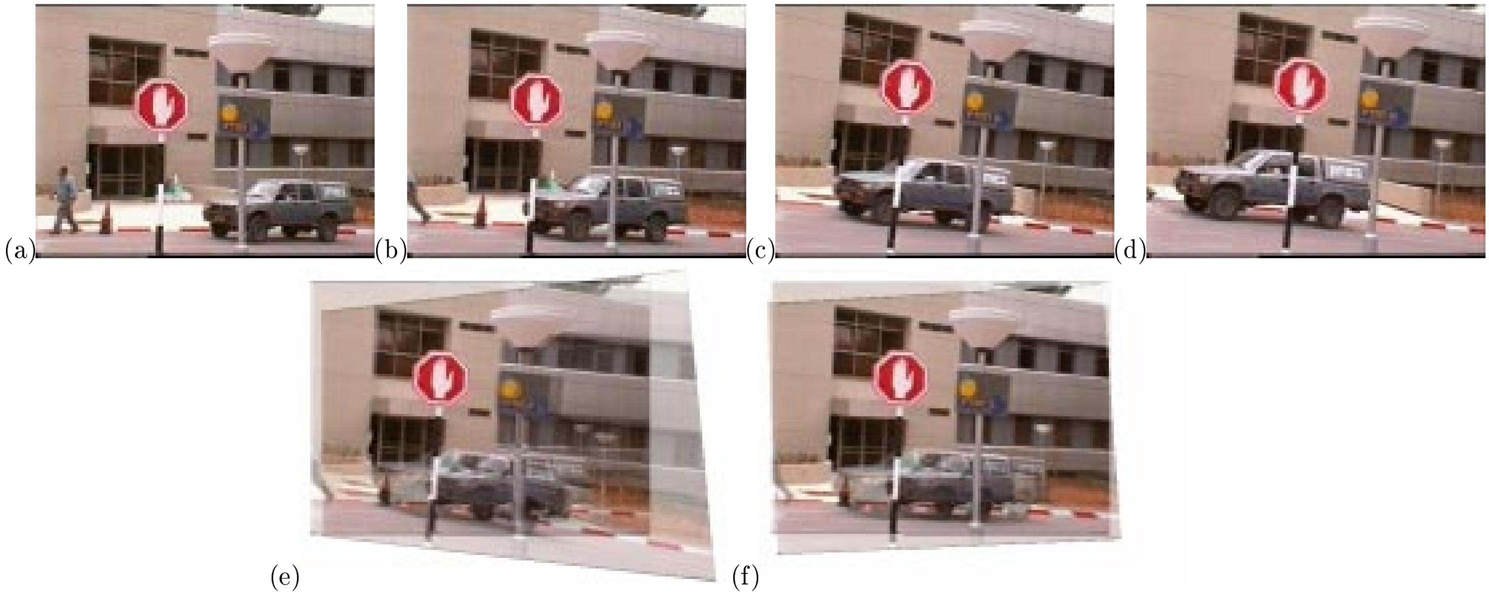

Figure 1: Constrained homography estimation. *(a,b,c,d) sample images from a collection of 25 images obtained from different camera positions. Image 1.b was used as the reference image. (e) Example of bad results from unconstrained two-view homography estimation of the stop sign region. The homography was estimated between the reference view 1.b and view 1.c. All the point correspondences were located on the sign itself and none on the pole. The displayed result is an overlay of the two images after registration according to the unconstrained homography. Although the stop-sign appears aligned, the rest of the image is completely distorted. Note that the pole of the stop-sign is already misaligned, although it lies on the same plane as the sign, and is very close to the region of analysis. (f) The corresponding result from applying the constrained two-plane multi-view homography estimation scheme, to the same region of analysis. The building was used as a reference plane (see text). The stop-sign is well aligned, including the pole, and the building displays accurate 3D parallax.*

However, they tend to be highly inaccurate when applied to small image regions [11], as is often the case in scenes with *multiple* planar surfaces.

While each independent homography computation is unreliable, all homographies of all planes, across all views must satisfy the multi-view subspace constraints. These constraints can therefore be used to compensate for insufficient spatial information.

Below we suggest one possible approach for taking advantage of the multi-view subspace constraints in the homography estimation:

(i) Define one image as the reference image J. Use any existing method to estimate initial homographies, for all planes and all images, with respect to the reference image.
(ii) Define one of the planes to be a reference plane π_r (e.g., choose π_r to be the plane with the largest image region in J, or the one with the most reliable initial homographies). Compute all the *relative homographies* (homologies, see Eq. (6)), of all other planes for all images, with respect to it.
(iii) If the number of planes = 2, do *not* perform any scaling. Otherwise ($\#planes > 2$) examine the entries $h_2, h_3, h_4, h_6, h_7, h_8$ of all relative homographies and choose the one which consistently differs from zero in all of them. Scale the relative homographies such that the chosen entry becomes 1.
(iv) Stack all relative homographies into a $9P \times F$ matrix $\mathcal{H}$ (see Eq. (12)).
(v) Project the columns of the matrix $\mathcal{H}$ onto a low-dimensional linear subspace, by constraining its rank to be ≤ 4. This gives a matrix $\hat{\mathcal{H}}$. (The choice of the actual rank of $\hat{\mathcal{H}}$, which may be even smaller than 4, can be done by examining the rate of decay of the matrix singular values).
(vi) Refine the estimation of the individual homographies by computing: $H^f_p = H^f_r \hat{H}^f_{pr}$, where H^f_r = *reference* homography and $\hat{H}^f_{pr}$ = is the subspace-projected *relative* homography (both in the regular 3×3 matrix form).
(vii) If you're using an iterative framework to solve for the homographies, repeat steps (i) to (vi) at each iteration.

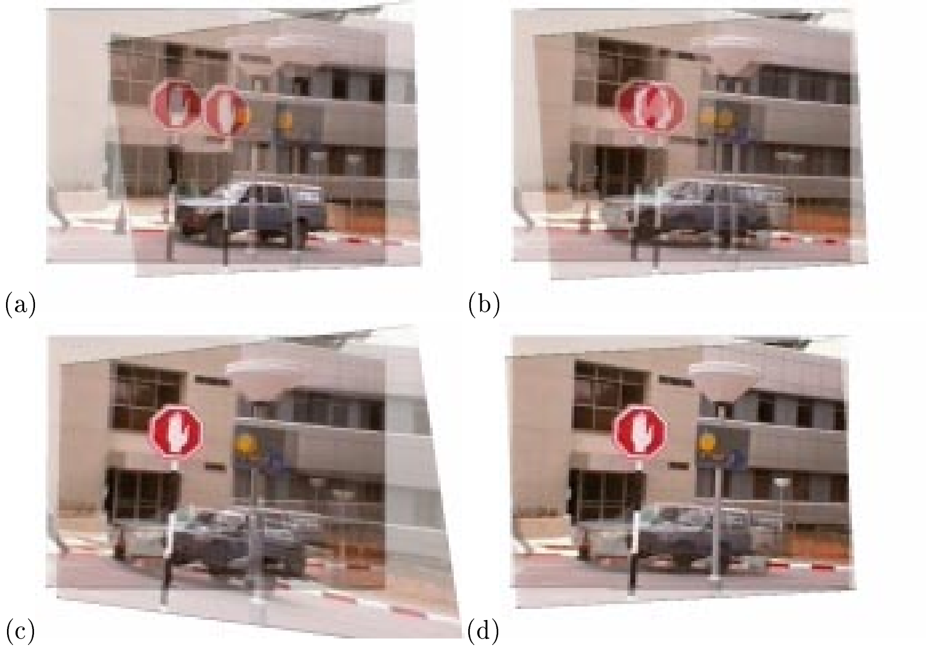

Figure 2: Moving object detection. *(a) Example of unconstrained two-view homography estimation of the car region, between the reference view 1.b and view 1.c. The car appears well aligned after registration and overlay of the two images according to the unconstrained homography. (b,c) The corresponding results from applying the constrained multi-plane multi-view homography estimation scheme, to the car and the stop-sign simultaneously, using the building as a reference plane. Since the car motion is inconsistent with those of the stop-sign and building motion, applying the multi-plane multi-view constraint spoils the homography estimation of both the car (shown in (2.b), where the car is no longer aligned), and the stop-sign (shown in (2.c), where the pole is not aligned). (d) In contrast, the corresponding result from applying the constrained multi-plane multi-view homography estimation scheme, to the stop-sign and the other road-sign simultaneously. The stop-sign is now well aligned, and the rest of the image displays accurate 3D parallax.*

Fig. 1 shows a comparison of applying *two*-view and *multi*-view homography estimation to small image regions. 25 images were taken from different viewing positions. Because the camera is imaging the scene from a short distance, and because its motion contains a translation, different planar surfaces (e.g., the building, the stop-sign, etc.) induce different homographies. The induced homographies of the building were first estimated, using a two-view estimation method. Since it occupies a large image region, these were computed accurately enough. The building was chosen as the reference plane, hence, its computed homographies were used as inputs for constraining the estimation of the homographies corresponding to the other planes in the scene, using the approach described above. Because the stop-sign occupies a very small image region, an unconstrained *two-view* homography estimation of the stop-sign gave distorted results (see Fig. 1.e). The *two-plane multi-view* homography estimation, on the other hand, gave good results for all images, eventhough applied to the same small region (see Fig. 1.f). For the purpose of these experiments the homographies where estimated using Least-Squares fit to precomputed point correspondences.

5.2 Moving Object Detection

The multi-view subspace constraints of Sections 3 and 4 are true only for planar surfaces moving rigidly with respect to each other. Planar surfaces with different $3D$ motions will not necessarily comply with these constraints. Given two planar surfaces (π_r and π_p), we can construct the matrix $\mathcal{H}_{pr}$ of their relative homographies (see Eq. (11)) and examine its rank. If $rank(\mathcal{H}_{pr}) > 4$ then the two planes cannot be rigid with respect to each other. Note that this is a *sufficient* condition, but not a necessary one. In the case of *multiple* planes, we can do the same with the matrix $\mathcal{H}$ of Eq. (12), after appropriate scaling (see Section 4).

In the presence of noise, however, the rank of the matrix $\mathcal{H}$ may appear to be larger than 4 even for rigid planes. To avoid misinterpretation due to errors in the homography estimation, and to detect rank violations which are truly due to inconsistent $3D$ motion, we take the following approach: We apply the multi-view rigidity scheme presented in Section 5.1 to *improve* the estimation of the individual homographies and their relative-homographies, as if the planes were rigid with respect to each other. If the planes are in fact rigid with respect to each other, (i.e., have the same $3D$ motions across all views) then this process will improve their homography estimation. (This can be verified e.g., by comparing the accuracy of alignment before and after applying the rank estimation). If, on the other hand, the planes are not rigid with respect to each other (i.e., have different $3D$ motions across *some* views), then forcing the multi-view low-dimensionality constraint will *spoil* the homography estimation, leading to larger misalignment errors. This is detected as a case of inconsistent $3D$ motion.

Fig. 2 presents the results of applying the multi-plane multi-view homography estimation to non-rigidly moving objects. The scene contains a car, moving independently of the camera motion. Using the previously computed homographies of the building region as a reference plane, we applied the multi-plane multi-view scheme, of Section 5.1, to the car and the stop-sign simultaneously (this time using the parameter scaling of Section 4). Since the car is not moving rigidly with respect to the stop-sign and the building, applying the constrained estimation resulted in *worse* homography estimation for the car as well as for the stop-sign, than those found by two-view unconstrained process (See Figs. 2.b and 2.c). In contrast, the same *multi-plane* multi-view scheme, was applied simultaneously to the stop-sign and the *other* road-sign. Accurate homography estimation is now achieved (See Fig. 2.d). Hence, the degradation in the homography estimation observed in Figs. 2.b and 2.c, indicates that the car is moving $3D$-inconsistently with respect to the other planes (the building and the two signs).

6 Concluding Remarks

In this paper we showed that the collection of homologies of multiple planar surfaces across multiple views, are embedded in a low dimensional linear subspace. We further showed that these constraints can be used to improve homography estimation in multi-planar scenes, and serve as a cue for moving object detection. While the paper presented the core constraints and the core elements of such approaches, the integration of these elements into a single end-to-end algorithm remains a task for our future research.

References

[1] Olivier Faugeras. *Three-Dimensional Computer Vision – A Geometric Viewpoint.* MIT Press, Cambridge, MA, 1996.

[2] A.W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*, pages 310–326, 1998.

[3] Van Gool, L. Proesmans, and A. Zisserman. Grouping and invariants using planar homologies. In *Workshop on Geometrical Modeling and Invariants for Computer Vision*, 1995.

[4] M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multi-view parallax geometry application. In *European Conference on Computer Vision*, pages 829–845, 1998.

[5] Michal Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *International Conference on Computer Vision*, pages 605–611, Cambridge, MA, November 1995.

[6] K. Kanatani. Optimal homograhpy computation with a reliability measure. In *Proc. of the IAPR Workshop on Machine Vision*, Makuhari, Chiba, Japan, November 1998.

[7] P. Pritchett and A. Zisserman. Matching and reconstruction from widely separated views in 3d structure from multiple images of large-scale environments. In *LNCS 1506*, Springer-Verlag, 1998.

[8] Q.T.Luong and O.Faugeras. Determining the fundamental matrix with planes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 489–494, New York, June 1993.

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

[10] A. Shashua and S. Avidan. The rank 4 constraint in multiple ($\geq$3) view geometry. In *European Conference on Computer Vision*, 1996.

[11] R. Szeliski and P.H.S Torr. Geometrically constrained structure from motion: Points on planes. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pages 171–186, Freiburg, Germany, June 1998.

[12] Bill Triggs. Autocalibration from planar scenes. In *European Conference on Computer Vision*, pages 89–105, 1998.

[13] T. Vieville, C.Zeller, and L.Robert. Using collineations to compute motion and structure in an uncalibrated image sequence. *International Journal of Computer Vision*, 20:213–242, 1996.

[14] W.T. Vetterling W.H. Press, S.A. Teukolsky and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, Cambridge, MA, 1992.