# Multi-Frame Alignment of Planes

**Lihi Zelnik-Manor** **Michal Irani**
Dept. of Computer Science and Applied Math
The Weizmann Institute of Science
76100 Rehovot, Israel

## Abstract

*Traditional plane alignment techniques are typically performed between* pairs *of frames. In this paper we present a method for extending existing two-frame planar-motion estimation techniques into a* simultaneous multi-frame *estimation, by exploiting multi-frame geometric constraints of planar surfaces. The paper has three main contributions: (i) we show that when the camera calibration does not change, the collection of all parametric image motions of a planar surface in the scene across multiple frames is embedded in a low dimensional linear subspace; (ii) we show that the* relative *image motion of multiple planar surfaces across multiple frames is embedded in a yet* lower *dimensional linear subspace, even with varying camera calibration; and (iii) we show how these multi-frame constraints can be incorporated into simultaneous multi-frame estimation of planar motion, without explicitly recovering any 3D information, or camera calibration. The resulting multi-frame estimation process is more constrained than the individual two-frame estimations, leading to more accurate alignment, even when applied to small image regions.*

## 1 Introduction

Plane stabilization ("2D parametric alignment") is essential for many video-related applications: it is used for video stabilization and visualization, for 3D analysis (e.g., using the Plane+Parallax approach [5]), for moving object detection, mosaicing, etc.

Many techniques have been proposed for estimating the 2D parametric motion of a planar surface between *two* frames. Some examples are [6, 3, 12, 2, 7]. While these techniques are very robust and perform well when the planar surface captures a large image region, they tend to be highly inaccurate when applied to small image regions. Moreover, errors can accumulate over a sequence of frames when the motion estimation is performed between *successive* pairs of frames (as is often done in mosaic construction).

An elegant approach was presented in [9] for automatically estimating an optimal (usually virtual) reference frame for a sequence of images with the corresponding motion parameters that relate each frame to the virtual reference frame. This overcomes the problem associated with error accumulation in sequential frame alignment. However, the alignment method used for estimating the motion between the virtual reference frame and all other frames remains a *two*-frame alignment method.

Sequential two-frame parametric alignment methods do not exploit the fact that all frames imaging the same planar surface share the same plane geometry. In this paper we present a method for extending traditional two-frame planar-motion estimation techniques into a *simultaneous multi-frame* estimation method, by exploiting multi-frame linear subspace constraints of planar motions (Section 4). The use of linear subspace constraints, for motions analysis, has been introduced by Tomasi and Kanade [11]. They used these constraints for factoring $2D$ correspondences into $3D$ motion and shape information. In contrast, here we use linear subspace constraints for *constraining* our $2D$ planar motion estimation and *not* for factoring out any $3D$ information. This results in a multi-frame estimation technique which is more constrained than the individual two-frame estimation processes, leading to more accurate alignment, even when applied to small image regions. Furthermore, multi-frame rigidity constraints relating multiple planar surfaces are applied to further enhance parametric motion estimation in scenes with *multiple* planar surfaces (Section 5).

## 2 Basic Model and Notations

The *instantaneous* image motion of a $3D$ planar surface $\pi$, between two image frames can be expressed as a $2D$ quadratic transformation [8, 1, 3, 5]:

$$\vec{u}(\vec{x};\vec{p}) = X(\vec{x}) \cdot \vec{p}, \tag{1}$$

where $X(\vec{x})$ is a matrix which depends only on the pixel coordinates $(\vec{x}) = (x, y)$:

$$X(\vec{x}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{bmatrix}.$$

and $\vec{p} = (p_1, p_2, ..., p_8)^t$ is a parameter vector:

$$\begin{array}{ll} p_1 = f'(\gamma t_X + \Omega_Y) & p_2 = \frac{f'}{f}(1 + \alpha t_X) - \gamma t_Z - 1 \\ p_3 = -\frac{f'}{f}(\Omega_Z - \beta t_X) & p_4 = f'(\gamma t_Y - \Omega_X) \\ p_5 = \frac{f'}{f}(\Omega_Z + \alpha t_Y) & p_6 = \frac{f'}{f}(1 + \beta t_Y) - \gamma t_Z - 1 \\ p_7 = \frac{1}{f}(\Omega_Y - \alpha t_Z) & p_8 = -\frac{1}{f}(\Omega_X + \beta t_Z) \end{array} \tag{2}$$

where $\vec{n} = (\alpha, \beta, \gamma)^t$ is the normal of the plane $\pi$ (i.e., $\vec{Q}^t\vec{n} = 1$, $\forall \vec{Q} \in \pi$), $\vec{\Omega} = (\Omega_X, \Omega_Y, \Omega_Z)^t$ and $\vec{t} = (t_X, t_Y, t_Z)^t$ are the camera rotation and translation, respectively, and $f$ and $f'$ are the camera focal lengths used for obtaining the two images.

The instantaneous motion model is valid when the camera rotation is small and the forward translation is small relative to the depth.

## 3 Two-Frame Parametric Alignment

In this paper, we extend the *direct* two-frame motion estimation approach of [3, 6] to multiple frames. To make the paper self contained, we briefly outline the basic two-frame technique below.

Two image frames (whose parametric image motion is being estimated) are referred to by the names "reference" image $J$ and "inspection" image $K$. A Gaussian pyramid is constructed for $J$ and $K$, and the motion parameters from $J$ to $K$ are estimated in a coarse-to-fine manner. Within each pyramid level the sum of squared *linearized* differences (i.e., the linearized brightness constancy measure) is used as a match measure. This measure ($Err$) is minimized with respect to the unknown 2D motion parameters $\vec{p}$ of Eq. (1):

$Err(\vec{p}) = \sum_{(\vec{x})} \left( (K(\vec{x}) - J(\vec{x})) \; + \; \nabla J(\vec{x})^t \cdot \vec{u}(\vec{x};\vec{p}) \right)^2$

where $(\vec{u}(\vec{x};\vec{p}))$ is defined in Eq. (1), $J(\vec{x})$ and $K(\vec{x})$ denote the brightness value of image $J$ and $K$ at pixel $\vec{x}$, respectively, and $\nabla J(\vec{x})$ denotes the spatial gradient of $J$ at $\vec{x}$: $\nabla J(\vec{x}) = \left( \frac{\partial J}{\partial x}(\vec{x}), \frac{\partial J}{\partial y}(\vec{x}) \right)^t$. The sum is computed over all the points within a region of interest (often the entire image). Deriving $Err$ with respect to the unknown parameters $\vec{p}$ and setting to zero, yields eight linear equations in the eight unknowns:

$$\mathcal{C} \cdot \vec{p} = \vec{b}, \tag{3}$$

where $\mathcal{C}$ is an $8 \times 8$ matrix:

$$\mathcal{C} = \sum_{(\vec{x})} \left[ X(\vec{x})^t \cdot \nabla J(\vec{x}) \cdot \nabla J(\vec{x})^t \cdot X(\vec{x}) \right],$$

and $\vec{b}$ is an $8 \times 1$ vector:

$$\vec{b} \;\; = \;\; \sum_{(\vec{x})} \left[ X(\vec{x})^t \cdot \nabla J(\vec{x}) \cdot (J(\vec{x}) - K(\vec{x})) \right].$$

This leads to the linear solution $\vec{p} = \mathcal{C}^{-1} \cdot \vec{b}$. Note that $\mathcal{C}$ and $\vec{b}$ are constructed of *measurable image quantities*, hence the word *direct* estimation. This process does not require recovery of any $3D$ information.

Let $\vec{p_i}$ denote the estimate of the quadratic parameters at iteration $i$. $\vec{p_{i+1}} = \vec{p_i} + \vec{\delta p}$ can be solved for by estimating the infinitesimal increment $\vec{\delta p}$. After iterating certain number of times within a pyramid level, the process continues at the next finer level, etc.

Figs. 1.f, 1.i show an example of applying this parametric alignment method. This is an airborne sequence taken from a large distance, hence the camera induced motion can be described by a single $2D$ parametric transformation of Eq. (1). Fig. 1.c was the reference image and Fig. 1.b was the inspection image. Fig. 1.e shows the amount of misalignment between the two input images. When the method is applied to the *entire image region*, it yields accurate alignment (at subpixel accuracy), as can be seen in Figs. 1.f and 1.i. However, once the same method is applied to a small image region (such as the rectangular region marked in Figs. 1.g, 1.j), its accuracy degrades significantly. The farther a pixel is from the region of analysis, the more misaligned it is. In the next section we show how employing *multiple* frames (as opposed to two) can be used to increase accuracy of image alignment even when applied to small image regions.

## 4 Multi-Frame Parametric Alignment

In this section we present a method for extending the *two*-frame technique reviewed in section 3 into a *multi*-frame technique, which exploits *multi-frame constraints* on the image motion of a planar surface. In section 4.1 we derive such a multi-frame constraint, and in section 4.2 we show how it can be incorporated into the $2D$ parametric estimation of planar motion, without requiring any recovery of $3D$ information, or camera calibration.

### 4.1 Single Plane Geometric Constraint

Let $J$ be a reference frame, and let $K^1, ..., K^F$ be a sequence of $F$ inspection frames imaging the *same* planar surface with the *same* focal length $f$. Let $\vec{p}^1, ..., \vec{p}^F$ be the corresponding quadratic parameter vectors of the planar motion (see Eq. (1)). The instantaneous motion model of Eq. (1) is a good approximation of the motion over *short video segments*, as the camera does not gain large motions in short periods of time. In some cases, such as airborne video, this approximation is good also for very long sequences. Choosing the reference frame as the *middle* frame extends the applicability of the model to twice as many frames.

We arrange $\vec{p}^1, ..., \vec{p}^F$ in an $8 \times F$ matrix $P$ where each column corresponds to one frame. From Eq. (2):

$$P = \left[ \; \vec{p}^1 \cdots \vec{p}^F \; \right]_{8 \times F} = S_{8 \times 6} \cdot \left[ \begin{array}{ccc} \vec{t}^1 & \cdots & \vec{t}^F \\ \vec{\Omega}^1 & & \vec{\Omega}^F \end{array} \right]_{6 \times F} \tag{4}$$

where

$$S = \left[ \begin{array}{cccccc} f\gamma & 0 & 0 & 0 & f & 0 \\ \alpha & 0 & -\gamma & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 & 0 & -1 \\ 0 & f\gamma & 0 & -f & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & 1 \\ 0 & \beta & -\gamma & 0 & 0 & 0 \\ 0 & 0 & -\frac{\alpha}{f} & 0 & \frac{1}{f} & 0 \\ 0 & 0 & -\frac{\beta}{f} & -\frac{1}{f} & 0 & 0 \end{array} \right] \tag{5}$$

and $\vec{t}^j, \vec{\Omega}^j$, are the camera translation and rotation, between the reference frame $J$ and frame $K^j$ ($j$ =

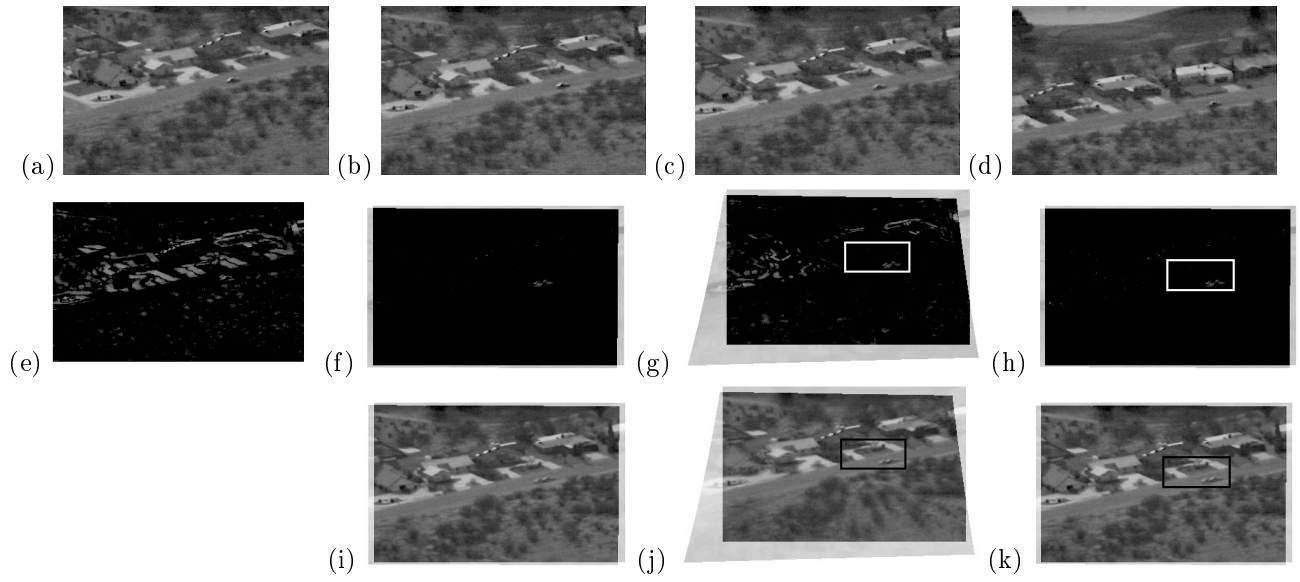


Figure 1: *(a,b,c,d) Sample frames from a 17-frame airborne video clip. Apart from camera motion, there is also a small moving car. Fig. 1.c was the reference frame. (e) The absolute differences between two input frames (1.b and 1.c) indicate initial misalignment. (f,i) High quality alignment (at sub-pixel accuracy) from applying the two-frame technique to the entire image region. 1.f shows absolute differences after alignment, while 1.i shows the average of the two aligned images. Only the independently moving car is misaligned. (g,j) Poor alignment from applying the two-frame alignment to a small image region, marked by a rectangle. Although the rectangular region is well aligned (apart from the moving car), large misalignments can be detected in image pixels which are distant from the analysis region. (h,k) High quality alignment from applying the constrained multi-frame alignment to the same small rectangular region. It was applied simultaneously to all 17 frames. Even pixels distant from the analysis window appear well aligned.*

$1..F$). Note that the shape matrix $S$ is common to all frames, because they all share the same plane normal $\vec{n} = (\alpha, \beta, \gamma)^t$ and focal length $f$. The dimensionality of the matrices on the right hand side of Eq. (4) implies that, without noise, the parameter matrix $P$ is of rank 6 at most. This implies that the collection of all the $p^j$'s ($j = 1..F$) resides in a low dimensional linear subspace. The actual rank of $P$ may be even lower than 6, depending on the complexity of the camera motion over the sequence (e.g., in case of uniform motion it will be 1).

### 4.2 Incorporating Sub-Space Constraint into Multi-Frame Estimation

In this section we show how the low-rank constraint on $P$ can be incorporated into the estimation of $\vec{p}^1, ..., \vec{p}^F$, *without* explicitly solving for any $3D$ information, nor for camera calibration.

It is *not* advisable to first solve for $P$ and then project its columns onto a lower dimensional subspace, because then the individual $\vec{p}^j$'s will already be very erroneous. Instead, we would like to use the low dimensionality constraint to *constrain* the estimation of the individual $\vec{p}^j$'s *a-priori*. We next show how we can apply this constraint *directly* to *measurable image quantities* prior to solving for the individual $\vec{p}^j$'s.

Since all inspection frames $K^1, ..., K^F$ share the same reference frame $J$, Eq. (3) can be extended to multiple frames as:

$$\mathcal{C}_{8\times 8} \cdot \left[ \begin{array}{c} \vec{p}^1 \cdots \vec{p}^F \end{array} \right]_{8\times F} = \left[ \begin{array}{c} \vec{b}^1 \cdots \vec{b}^F \end{array} \right]_{8\times F} \qquad (6)$$

or, in short: $\mathcal{C} \cdot P = B$. Eq. (6) implies that $rank(B) \leq rank(P) \leq 6$. $B$ contains only measurable image quantities. Therefore, instead of applying the low-rank constraint to $P$, we apply it directly to $B$, and only then solve for $P$. Namely: at each iteration $i$ of the algorithm, first compute $B_i = [\vec{b}_i^1 \cdots \vec{b}_i^F]$, and then project its columns onto a lower-dimensional linear subspace by seeking a matrix $\hat{B}_i$ of rank $r$ ($r \leq 6$), which is closest to $B_i$ (in the Frobenius norm). Then solve for $P_i = \mathcal{C}_i^{-1} \hat{B}_i$, which yields the desired $\vec{p}^j$'s.

The advantage of applying the constraint to $B$ instead of $P$ can also be explained as follows: Note that the matrix $\mathcal{C}$ in Eq. (3) is the posterior inverse covariance matrix of the parameter vector $\vec{p}$. Therefore, applying the constraint to $B$ is equivalent to applying it to the matrix $P$, but after normalizing its columns by the inverse covariance matrix $\mathcal{C}$ (Note that all $\vec{p}^j$'s share the same $\mathcal{C}$).

To equalize the effect of subspace projection on all matrix entries, and to further condition the numerical

process, we use the coordinate normalization technique suggested by [4] (Also used in the two-frame method).

Fig. 1 shows a comparison of applying the *two*-frame and *multi*-frame alignment techniques. Figs. 1.g and 1.j show the result of applying the *two*-frame alignment technique (See section 3) to a small image region. The region of interest (marked rectangle) is indeed aligned, but the rest of the image is completely distorted. In contrast, the *multi*-frame constrained alignment (applied to 17 frames), successfully aligned the entire image eventhough applied only to the same small region. This can be seen in Figs. 1.h and 1.k.

Fig. 2 shows a quantitative comparison of the *two*-frame and *multi*-frame alignment techniques. When applying two-frame motion estimation to the small region, the farther the pixel is from the center of the region the larger the error is. However, when applying multi-frame motion estimation to the same small region, the errors everywhere are at sub-pixel level.

Fig. 3 shows another comparison of applying *two*-frame alignment and *multi*-frame alignment to small image regions. The sequence contains 34 frames taken by a moving camera. Because the camera is imaging the scene from a short distance, and because its motion also contains a translation, therefore different planar surfaces (e.g., the house, the road-sign, etc.) induce different $2D$ parametric motions. As long as the house was not occluded, the *two*-frame alignment, when applied only to the house region, stabilized the house reasonably well. However, once the house was partially occluded by the road-sign, and was not fully in the camera's field of view, the quality of the two-frame alignment degraded drastically (see Figs. 3.f and 3.j). The *multi*-frame constrained alignment, on the other hand, successfully aligned the house, even in frames where only a small portion of the house was visible (see Figs. 3.g and 3.k). In this case, the actual rank used was much smaller than 6 (it was 2), and was detected from studying the rate of decay of the eigenvalues of the matrix $B$.

# 5 Extending to Multiple Planes

In this section we show that in scenes containing multiple planar surfaces, even stronger geometric constraints can be derived and used to improve the parametric motion estimation. We present two different multi frame geometric constraints for sequences with multiple planes. The second constraint does not require constant camera calibration.

## 5.1 The Multi-Plane Rank-6 Constraint

Let $\pi_1, ..., \pi_m$ be $m$ planar surfaces with normals $\vec{n}_{\pi_1} = [\alpha_1, \beta_1, \gamma_1]^t$, ... , $\vec{n}_{\pi_m} = [\alpha_m, \beta_m, \gamma_m]^t$, respectively. Let $P_{\pi_1}, ..., P_{\pi_m}$ be the corresponding quadratic motion parameter *matrices*, and let $S_{\pi_1}, ..., S_{\pi_m}$ be the corresponding shape matrices, as defined by Eqs. (4) and (5). We can stack the $P_{\pi_\eta}(\eta = 1...m)$ matrices to form an $8m \times F$ matrix $\mathcal{P}$, where each column corresponds to one frame. Since all planar surfaces $\pi_\eta$ share the same $3D$ camera motion between a *pair of frames*, we get from Eq. (4):

$$\mathcal{P} = \begin{bmatrix} \frac{P_{\pi_1}}{\vdots} \\ \hline P_{\pi_m} \end{bmatrix}_{8m \times F} = \begin{bmatrix} \frac{S_{\pi_1}}{\vdots} \\ \hline S_{\pi_m} \end{bmatrix}_{8m \times 6} \begin{bmatrix} \vec{t}^1 & \dots & \vec{t}^F \\ \vec{\Omega}^1 & & \vec{\Omega}^F \end{bmatrix}_{6 \times F} \tag{7}$$

The dimensionality of the matrices on the right hand side of Eq. (7) implies that, without noise, the parameter matrix $\mathcal{P}$ is also of rank 6 at most. (As before, the actual rank of $\mathcal{P}$ may be even lower than 6, depending on the complexity and variability of the camera motion over the sequence).

## 5.2 Incorporating Multi-Plane Rank-6 Constraint into Estimation

Again we would like to apply the constraint to measurable image quantities. We show next how this can be done. Let $\mathcal{C}_{\pi_1}, ..., \mathcal{C}_{\pi_m}$ be the matrices corresponding to planes $\pi_1, ..., \pi_m$. Note that the matrices $\mathcal{C}_{\pi_\eta}$'s are different from each other due to the difference in the region of summation, which is the region of each planar surface in the reference frame. We can write:

$$\begin{bmatrix} \mathcal{C}_{\pi_1} & 0 & \cdots & 0 \\ 0 & \mathcal{C}_{\pi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{C}_{\pi_m} \end{bmatrix} \cdot \begin{bmatrix} P_{\pi_1} \\ \vdots \\ P_{\pi_m} \end{bmatrix}_{8m \times F} = \begin{bmatrix} B_{\pi_1} \\ \vdots \\ B_{\pi_m} \end{bmatrix}_{8m \times F} \tag{8}$$

or, in short: $\mathcal{C} \cdot \mathcal{P} = \mathcal{B}$. Note that here, as opposed to Eq. (6), $\mathcal{P}, \mathcal{B}$ and $\mathcal{C}$ contain information from all planar surfaces. Eq. (8) implies that $rank(\mathcal{B}) \leq rank(\mathcal{P}) \leq 6$.

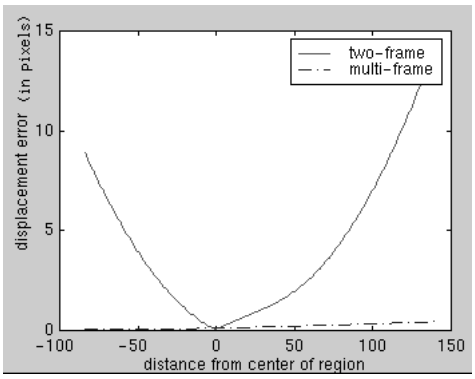


Figure 2: *A quantitative comparison of* two-*frame and* multi-*frame alignment. The values in the graph correspond to misalignments in Figs. 1.g and 1.h. These errors are displayed as a function of the distance from the center of the rectangular region in Fig. 1. The results of two-frame alignment applied to the entire image region (Fig. 1.f) were used as ground truth.*

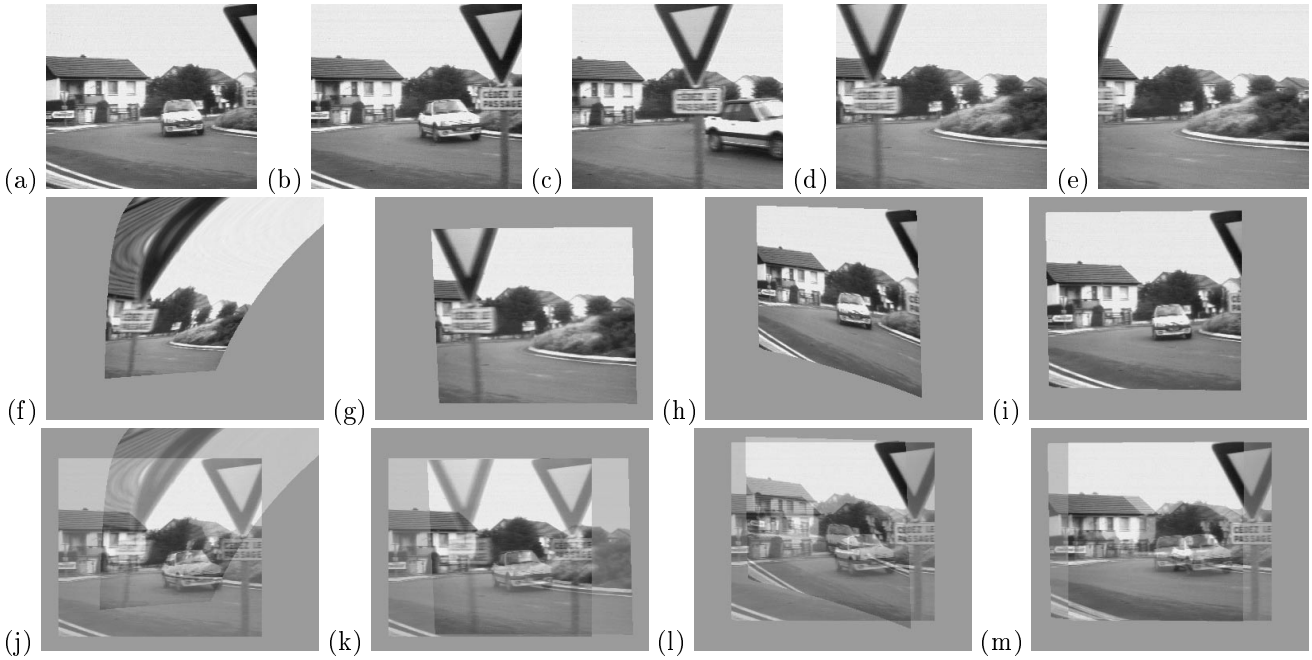


Figure 3: *(a,b,c,d,e) sample images from a sequence of 34 frames. Image 3.b was used as the reference frame. (f) Bad two-frame alignment of the house region between the reference frame 3.b and frame 3.d. The frame was completely distorted because the house region was significantly occluded by the road-sign; (j) shows same result overlayed on top of the reference image 3.b. (g) The corresponding result from applying the constrained multi-frame alignment. The house is now well aligned even though only a small portion of the house is visible (see overlay image (k)), while the rest of the image is not distorted. The road sign is not aligned because is at a different depth, and displays accurate 3D parallax. (h,l) Badly distorted two-frame alignment applied to the road-sign between frames 3.b and 3.a (where the sign is barely visible). Although the sign appears aligned, the rest of the image is distorted, and the house displays wrong parallax. (i,m) The corresponding result (to (h,l)) from applying the constrained multi-plane (multi-frame) alignment to the sign (see text). The house displays now accurate 3D parallax.*

We thus project the columns of matrix $\mathcal{B}$ onto a lower-dimensional subspace, at each iteration, resulting in $\hat{\mathcal{B}}$ (which is closest to $\mathcal{B}$ in the Frobenius norm), and then solve for $\mathcal{P} = \mathcal{C}^{-1} \cdot \mathcal{B}$. In other words, we solve for all parametric motions of all planar surfaces, across all frames, simultaneously. The low-rank constraint is stronger here than in the single plane case, because the matrix $\mathcal{B}$ is of a larger dimension ($8m \times F$).

### 5.3 Relative Motion Rank-3 Constraint

Moreover, by looking at the *relative* motion of planar surfaces we can get an even *stronger* geometric constraint, which is true *even for the case of varying camera focal length.*

Let $f$ and $f^j$ be the focal length of the frame $J$ and frame $K^j$, respectively. Let $\pi_r$ be an arbitrary planar surface ($\pi_r$ could be one of $\pi_1, ..., \pi_m$). Denote $\Delta\vec{p}^j_\eta = \vec{p}^j_{\pi_\eta} - \vec{p}^j_{\pi_r}$ ($\eta = 1...m$). It is easy to see from Eq. (2) that taking the difference $\Delta\vec{p}^j_\eta$ eliminates all effects of camera rotation, leaving only effects of camera translation and the focal length:

$$[\Delta\vec{p}^j_\eta]_{8\times 1} = [\Delta\tilde{S}_\eta]_{8\times 3}[\vec{\tau}^j]_{3\times 1} \tag{9}$$

where $\vec{\tau}^j = [f^j t^j_X, f^j t^j_Y, T = t^j_Z]^t$, and:

$$\Delta\tilde{S}_\eta = \begin{bmatrix} (\gamma_\eta - \gamma_r) & 0 & 0 \\ \frac{1}{f}(\alpha_\eta - \alpha_r) & 0 & -(\gamma_\eta - \gamma_r) \\ \frac{1}{f}(\beta_\eta - \beta_r) & 0 & 0 \\ 0 & (\gamma_\eta - \gamma_r) & 0 \\ 0 & \frac{1}{f}(\alpha_\eta - \alpha_r) & 0 \\ 0 & \frac{1}{f}(\beta_\eta - \beta_r) & -(\gamma_\eta - \gamma_r) \\ 0 & 0 & -\frac{1}{f}(\alpha_\eta - \alpha_r) \\ 0 & 0 & -\frac{1}{f}(\beta_\eta - \beta_r) \end{bmatrix}$$

$\Delta\tilde{S}_\eta$ is common to all *frames*. The camera translation $\vec{t}^j$ and focal length $f^j$ are common to all *planes* (between the reference frame $J$ and frame $K^j$). We can therefore extend Eq. (9) to *multiple planes* and *multiple frames* as follows:

$$\Delta\mathcal{P} = \begin{bmatrix} \Delta P_{\pi_1} \\ \vdots \\ \Delta P_{\pi_m} \end{bmatrix}_{8m\times F} = \begin{bmatrix} \Delta\tilde{S_{\pi_1}} \\ \vdots \\ \Delta\tilde{S_{\pi_m}} \end{bmatrix}_{8m\times 3} \cdot [\vec{\tau}^1 ... \vec{\tau}^F]_{3\times F} \tag{10}$$

The dimensionality of the matrices on the right hand side of Eq. (10) implies that, without noise, the differ-

ence parameter matrix $\Delta\mathcal{P}$ is of rank 3 at most.

It is possible to obtain a similar constraint (with $rank \leq 4$), for *general homographies* case [13] (as opposed to the instantaneous case). The rank-4 constraint is an extension to the constraint shown by [10]. Shashua and Avidan presented a rank-4 constraint on the collection of homographies of *multiple planes* between a *pair of frames*. In our case [13] the constraints are on *multiple planes* across *multiple frames*. We refer the reader to [13] for more details.

In the next section (5.4) we show how the multi-plane rank-3 constraint can be incorporated into the multi-frame estimation process to further enhance planar-motion estimation.

### 5.4 Incorporating the Rank-3 Constraint into Multi-Frame Estimation

Assume that for one planar surface, $\pi_r$, we know the collection of all its parametric motions, $P_{\pi_r}$. (This is either given to us, or estimated at previous iteration). We would like to use the ($rank \leq 3$) constraint to refine the estimation of the collection of parametric motions $P_{\pi_1}, ..., P_{\pi_m}$, of all other planes. Using Eq. (8) we derive:

$$\mathcal{C} \cdot \Delta\mathcal{P} = \begin{bmatrix} \underline{B_{\pi_1} - \mathcal{C}_{\pi_1} \cdot P_{\pi_r}} \\ \vdots \\ \underline{B_{\pi_m} - \mathcal{C}_{\pi_m} \cdot P_{\pi_r}} \end{bmatrix}_{8m \times F} = \mathcal{B}^* \qquad (11)$$

Therefore $rank(\mathcal{B}^*) \leq rank(\Delta\mathcal{P}) \leq 3$. To Incorporate the constraint into the estimation of the individual $P_{\pi_\eta}$'s, we project the columns of the matrix $\mathcal{B}^*$ onto a lower-dimensional ($\leq 3$) subspace at each iteration, resulting in $\hat{\mathcal{B}}^*$ (which is closest to $\mathcal{B}^*$ in Frobenius norm). Therefore, (from Eq. (11)) we can estimate a new matrix $B^{**}$

$$\mathcal{B}^{**} = \mathcal{C} \cdot \mathcal{P} = \begin{bmatrix} \underline{\mathcal{C}_{\pi_1}} \\ \vdots \\ \underline{\mathcal{C}_{\pi_m}} \end{bmatrix} \cdot P_{\pi_r} + \hat{\mathcal{B}}^* \qquad (12)$$

and then solve for $P_{\pi_\eta} = \mathcal{C}_{\pi_\eta}^{-1} \cdot B_\eta^{**}$. Note that here $\mathcal{B}^*$ and $\mathcal{B}^{**}$ are constructed from measurable image quantities ($\mathcal{B}$ and $\mathcal{C}$), as well as from the parameters $P_{\pi_r}$, which are either known or else estimated at previous iteration. The process is repeated at each iteration. Note: (i) If we know that the focal length does not change along the sequence, we can also apply the constraint $rank(\mathcal{B}^{**}) \leq rank(\mathcal{P}) \leq 6$ prior to solving for $P_{\pi_\eta}$; (ii) $\pi_r$ can alternate between the planes, but we found it to work best when $\pi_r$ was chosen to be a "dominant" plane (i.e., one whose matrix $\mathcal{C}_{\pi_r}$ is best conditioned).

Fig. 3 also presents a comparison of regular (unconstrained) two-frame alignment with the *multi-plane* constrained alignment, applied to the road-sign. The motion parameters of the house region were first estimated using the single-plane multi-frame constrained alignment (see Section 4). These were then used as inputs for constraining the estimation of the sequence of 2D motion parameters of the road-sign. The two-frame alignment technique did not perform well in cases when the sign was only partially visible (see Figs. 3.h and 3.l). The multi-plane (multi-frame) constrained alignment, on the other hand, stabilized the sign well even in cases when the sign was only partially visible (see Figs. 3.i and 3.m).


[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *PAMI*, 7(4):384–401, July 1985.

[2] S. Ayer and H. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding. In *ICCV*, pages 777–784, 1995.

[3] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *ECCV*, pages 237–252, 1992.

[4] R. I. Hartley. In defence of the 8-point algorithm. *PAMI*, 19(6):580–593, June 1997.

[5] M. Irani, B. Rousso, and P. peleg. Recovery of ego-motion using region alignment. *PAMI*, 19(3):268–272, March 1997.

[6] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *IJCV*, 12:5–16, February 1994.

[7] S. Ju, M. J. Black, and A. D. Jepson. Multi-layer, locally affine optical flow and regularization with transparency. In *CVPR*, pages 307–314, 1996.

[8] H. Longuet-Higgins. Visual ambiguity of a moving plane. *Proceedings of The Royal Society of London B*, 223:165–175, 1984.

[9] H. Sawhney and R. Kumar. True multi-image alignment and its application to mosaic[k]ing and lens distortion correction. In *CVPR*, pages 450–456, 1997.

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

[11] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *IJCV*, 9:137–154, November 1992.

[12] J. Wang and E. Adelson. Layered representation for motion analysis. In *CVPR*, pages 361–366, 1993.

[13] L. Zelnik-Manor. Multi-frame alignment of planes. MSc thesis, The Weizmann Institute of Science, Israel, March 1999.