Chaining Planar Homographies for Fast and Reliable 3D Plane Tracking *

Manolis I.A. Lourakis and Antonis A. Argyros
Institute of Computer Science, Foundation for Research and Technology - Hellas
Vassilika Vouton, P.O.Box 1385, GR 711 10, Heraklion, Crete, GREECE
{lourakis|argyros}@ics.forth.gr – http://www.ics.forth.gr/cvrl/

Abstract

This paper addresses the problem of tracking a 3D plane over a sequence of images acquired by a free moving camera, a task that is of central importance to a wide variety of vision tasks. A feature-based method is proposed which given a triplet of consecutive images and a plane homography between the first two of them, estimates the homography induced by the same plane between the second and third images, without requiring the plane to be segmented from the rest of the scene. Thus, the proposed method operates by "chaining" (i.e. propagating) across frames the image-to-image homographies due to some 3D plane. The chaining operation represents projective space using a "plane + parallax" decomposition, which permits the combination of constraints arising from all available point matches, regardless of whether they actually lie on the tracked 3D plane or not. Experimental results are also provided.

1 Introduction

Plane tracking is essential for a wide variety of vision applications, ranging from 3D reconstruction [7] to visual localization and augmented reality [10]. Depending on the type of constraints used for driving plane tracking, proposed approaches fall into two categories. The first consists of methods attempting to identify and track *2D layers*, that is image regions comprised of pixels sharing the same parametric 2D motion. Such layers arise from the motion of either piecewise planar imaged objects or of smooth, shallow surfaces whose image motion can be well approximated using a parametric model involving image coordinates. Due to their wide practical applicability, 2D layers have been used by several researchers for representing moving images [12, 13]. In [13], Zelnik-Manor and Irani have applied a linear subspace constraint involving multiple planes across multiple views to estimate the motion of small predefined planar regions. Their method operates in a batch mode, estimating the motion of the entire sequence in a single step and requires the exact spatial extend of each of the tracked regions to be determined in all images. Layer-based methods assume instantaneous camera motion and make extensive use of the spatiotemporal intensity derivatives. Therefore, they are inapplicable when the apparent image displacements are large.

In cases where the inter-frame image motion is not infinitesimal, only sparse feature matches can be reliably extracted from images. Techniques operating under such conditions constitute the second class of plane tracking methods and rely upon geometric constraints arising from matching features. For example, Simon et al [10] track planes by estimating pairwise planar homographies with the aid of tracked interest points. Nevertheless, the approach of [10] requires the image of the tracked plane to be segmented from the scene, calls for manual intervention to bootstrap tracking and is unable to incorporate information arising from points off the plane or from more than two images. Avidan and Shashua [2] follow a direct approach for recovering a set of consistent camera matrices without reconstructing the 3D scene. The main contribution of their work is a merging operation on two consecutive fundamental matrices that uses the trifocal tensor as the connecting thread. Their method is based on tracking a scene plane along an image sequence and provides, as a byproduct, the homography matrices it induces between adjacent views. However, owing to the use of constraints involving algebraic distances, the estimated homographies are not optimal.

In this paper, a novel feature-based approach to plane tracking is presented. The method is based on a homography "chaining" operation that is applied to triplets of consecutive images through a sliding time window and exploits the fact that all images of a planar surface acquired by a rigidly moving observer depend upon the same 3D geometry. Plane tracking is achieved by tracking the 2D projections of points from all over the scene and expressing their image motion as the sum of a homographic transfer plus a residual planar parallax vector. Due to the fact that this "plane + parallax" decomposition simplifies the motion equations by partially factoring out the dependence on the camera relative rotation and intrinsic calibration parameters, it has often been employed for dealing with a wide

*Partially supported by the EU FP6-507752 NoE MUSCLE.

0-7695-2521-0/06/$20.00 (c) 2006 IEEE

spectrum of vision problems [3, 11]. Here, the application of the "plane + parallax" decomposition allows all information conveyed by matching points to be taken into account, without the need for continuously maintaining a segmentation of the tracked plane from the scene. The motion model estimated for the tracked plane is exact and fully projective (i.e. a homography) and no camera calibration information or 3D structure recovery is necessary. Intended for use in time-critical applications, the proposed method is designed to operate in a continuous mode, in which images are processed incrementally as acquired. This is in contrast to methods that process image data in a single batch step and are therefore not suitable for time-sensitive applications. The proposed method follows a strategy similar to that in [2], with an important difference being that it is based on simpler constraints whose derivation is shorter and does not involve the trifocal tensor. Moreover, our method tracks 3D planes by minimizing a geometrically meaningful criterion with respect to a set of four free parameters, which, according to the subspace constraint of [8], is a theoretically minimal one. Compared to [2] and [10] which estimate twelve and eight parameters respectively, the estimation of just four parameters is both faster and more accurate. The rest of the paper is organized as follows. Section 2 explains the notation that will be used and provides some background knowledge. Section 3 derives the constraints that form the basis for the proposed plane tracking method. Experimental results from a prototype implementation are reported in section 4. The paper is concluded with a brief discussion in section 5.

2 Notation and Background

In the following, vectors and arrays appear in boldface and are represented using projective (homogeneous) coordinates [5]; $\simeq$ denotes equality up to an arbitrary scale factor. 3D points are written in uppercase and their image projections in lowercase (e.g. $\mathbf{P}$ and $\mathbf{p}$). $\mathbf{F}$ will designate the fundamental matrix, $\mathbf{e}$ and $\mathbf{e}^{'}$ its associated epipoles and $\mathbf{H}$ will be used for plane homographies.

As shown in [8], the fundamental matrix and plane homographies are tightly coupled. Specifically, the group of all possible homography matrices between two images lies in a subspace of dimension 4, i.e. it is spanned by 4 homography matrices. These homography matrices are such that their respective planes do not all coincide with a single point. Shashua and Avidan show in [1] that given the fundamental matrix $\mathbf{F}$ and the epipoles $\mathbf{e}$ and $\mathbf{e}^{'}$ in an image pair, a suitable basis of 4 homography matrices $\mathbf{H}_1, \ldots, \mathbf{H}_4$, referred to as "primitive homographies", is defined as follows:

$$\mathbf{H}_i = [\epsilon_i]_\times \mathbf{F}, \quad i = 1, 2, 3 \qquad \text{and} \qquad \mathbf{H}_4 = \mathbf{e}^{'}\delta^T, \quad (1)$$

where ϵ_i are the identity vectors $\epsilon_1 = (1,0,0)$, $\epsilon_2 = (0,1,0)$ and $\epsilon_3 = (0,0,1)$, $[.]_\times$ designates the skew symmetric matrix representing the vector cross product and δ is a vector such that $\delta^T\mathbf{e} \neq 0$. Given the primitive homographies, any other homography $\mathbf{H}$ can be expressed as their linear combination $\mathbf{H} = \sum_{i=1}^{4} \lambda_i \mathbf{H}_i,\ \lambda_i \in \mathcal{R}$.

Next, a result due to Shashua and Navab [9] that plays a central role in the development of the proposed method is presented. Let Π be an arbitrary 3D plane inducing a homography $\mathbf{H}$ between two images. Let also $\mathbf{P}_0$ be a 3D point not on Π projecting to image points $\mathbf{p}_0$ and $\mathbf{p}_0^{'}$ and assume that $\mathbf{H}$ has been scaled to satisfy the equation $\mathbf{p}_0^{'} \simeq \mathbf{H}\mathbf{p}_0 + \mathbf{e}^{'}$. Then, for any 3D point $\mathbf{P}$ projecting onto $\mathbf{p}$ and $\mathbf{p}^{'}$, there exists a scalar κ such that

$$\mathbf{p}^{'} \simeq \mathbf{H}\mathbf{p} + \kappa\mathbf{e}^{'}. \quad (2)$$

Equation (2) dictates that the position of projected points in the second image can be decomposed into the sum of two terms, the first depending on the homography induced by Π and the second involving *parallax* due to the deviation of the actual 3D structure from Π. The term κ in Eq. (2) depends on $\mathbf{P}$ but is invariant to the choice of the second image and is termed as *relative affine structure* in [9]. Given $\mathbf{p}$, $\mathbf{p}^{'}$, $\mathbf{H}$ and $\mathbf{e}^{'}$, the term κ corresponding to $\mathbf{P}$ can be computed by cross-multiplying both sides of Eq. (2) with $\mathbf{p}^{'}$, which after some algebraic manipulation yields

$$\kappa = \frac{(\mathbf{H}\mathbf{p} \times \mathbf{p}^{'})^T(\mathbf{p}^{'} \times \mathbf{e}^{'})}{||\mathbf{p}^{'} \times \mathbf{e}^{'}||^2}. \quad (3)$$

3 The Proposed Method

We start by shedding some light into the function of point $\mathbf{P}_0$ in the derivation of Eq. (2). Recall that $\mathbf{H}$ and $\mathbf{e}^{'}$ are homogeneous entities, defined up to an arbitrary scale factor. Therefore, by fixing $\mathbf{H}$'s scale, $\mathbf{P}_0$ serves to establish a common relative scale between $\mathbf{H}$ and $\mathbf{e}^{'}$. Notice, however, that in the case that $\mathbf{H}$ has not been scaled with the aid of $\mathbf{P}_0$, Eq. (2) continues to hold for some $\kappa^{'}$ that is a fixed scale multiple of the κ given by Eq. (3) for the scaled $\mathbf{H}$.

Suppose now that three consecutive images I_1, I_2 and I_3 are available and that a planar homography between I_1 and I_2 has been estimated. Considering the two image pairs (I_1, I_2) and (I_2, I_3), the key observation upon which the chaining operation is based is the fact that image I_2 is shared by both of them. Hence, the relative affine structure defined when I_2 assumes the role of the first image in Eq. (2) is insensitive to the choice of the second image (i.e. I_1 or I_3) completing the pair. This allows one to extract the relative affine structure from the pair (I_1, I_2) and the corresponding homography and then use it to estimate the plane homography pertaining to the pair (I_2, I_3). More details are given in the next section.

3.1 Chaining Homographies Among Consecutive Frames

Assume that N triplets of matching points $(\mathbf{p}_i,\ \mathbf{p}_i^{'},\ \mathbf{p}_i^{''})$, $i = 1,\ldots,N$, are available across the three images I_1, I_2 and I_3 respectively and that the homography $\mathbf{U}$ from image I_1 to I_2 due to some 3D plane has been estimated. A procedure for estimating the plane homography $\mathbf{V}$ induced by this 3D plane between images I_2 and I_3 will be described in the remainder of this section.

From the set of matching pairs $(\mathbf{p}_i,\ \mathbf{p}_i^{'})$ the epipolar geometry for images I_1 and I_2 and thus the epipole $\mathbf{e}$ in image I_1 can be estimated. In a similar manner, the epipole $\mathbf{e}^{''}$ in I_3 for the camera motion corresponding to frames I_2 and I_3 can be estimated from the set of matching pairs $(\mathbf{p}_i^{'},\ \mathbf{p}_i^{''})$. Recalling that the homography from image I_2 to I_1 is simply $\mathbf{U}^{-1}$, Eq. (2) takes the following form for all point matches in those two images: $\mathbf{p}_i \simeq \mathbf{U}^{-1}\mathbf{p}_i^{'} + \kappa_i\mathbf{e}$. By employing Eq. (3), κ_i can then be estimated as

$$\kappa_i = \frac{(\mathbf{U}^{-1}\mathbf{p}_i^{'} \times \mathbf{p}_i)^T(\mathbf{p}_i \times \mathbf{e})}{||\mathbf{p}_i \times \mathbf{e}||^2}. \quad (4)$$

Taking into account point matches in frames I_2 and I_3, Eq. (2) gives

$$\mathbf{p}_i^{''} \simeq \mathbf{V}\mathbf{p}_i^{'} + \kappa_i\mathbf{e}^{''}, \quad (5)$$

where κ_i are given by Eq. (4). In order for Eq. (5) to hold for the κ_i given by Eq. (4), the scale of $\mathbf{V}$ in it has to be compatible with that of $\mathbf{e}^{''}$. For this reason, $\mathbf{V}$ in Eq. (5) is no longer a homogeneous 3×3 matrix but rather an ordinary, inhomogeneous one. Equation (5) is thus a vector equation linear in $\mathbf{V}$, providing three linear constraints on the nine unknown elements of $\mathbf{V}$. Due to the presence of an arbitrary, unknown scale factor, only two of those three constraints are linearly independent. Denoting the i-th row of matrix $\mathbf{V}$ by $\mathbf{v}_i^T$, writing $\mathbf{p}_i^{''} = (x_i^{''}, y_i^{''}, 1)^T$ and $\mathbf{e}^{''} = (e_x^{''},\ e_y^{''},\ e_z^{''})^T$, those two constraints can be explicitly expressed as [1]

$$\begin{aligned} \mathbf{v}_3^T\mathbf{p}_i^{'}x_i^{''} - \mathbf{v}_1^T\mathbf{p}_i^{'} &= \kappa_i e_x^{''} - \kappa_i e_z^{''} x_i^{''} \\ \mathbf{v}_3^T\mathbf{p}_i^{'}y_i^{''} - \mathbf{v}_2^T\mathbf{p}_i^{'} &= \kappa_i e_y^{''} - \kappa_i e_z^{''} y_i^{''}. \end{aligned} \quad (6)$$

Equations (6) do not require that the employed point matches have been identified as lying on the plane. Therefore, they do not require that the tracked plane has been segmented from the rest of the scene and are applicable even in the case of tracking a virtual (i.e. not physically present in the scene) plane. Concatenating the equations arising from five triplet correspondences, a matrix equation of the form $\mathbf{M}\mathbf{v} = \mathbf{b}$ is generated, where $\mathbf{M}$ is a 10×9 matrix, $\mathbf{v} = (\mathbf{v}_1^T,\ \mathbf{v}_2^T,\ \mathbf{v}_3^T)^T$ and $\mathbf{b}$ is a 10×1 vector. Omitting any row of matrix $\mathbf{M}$, yields a 9×9 linear system with 9 unknowns that may be solved using linear algebra. In the case that more than five triplet matches are available, Eq. (6) gives rise to an over-constrained system from which $\mathbf{V}$ can be estimated in a least squares manner.

As described up to this point, the estimation of $\mathbf{V}$ is achieved with a Direct Linear Transformation (DLT) algorithm (see [5], ch. 3). Since DLT algorithms are not invariant to similarity transformations of the point image coordinates, the accuracy of the estimation can be improved by applying the normalization technique of [4] to matching points prior to feeding them to the DLT algorithm. Independently for each image i, this normalization corresponds to a linear transformation $\mathbf{L}_i$, consisting of a translation followed by an isotropic scaling that maps the average image point to $(1,1,1)^T$. Notice that in this case, the normalized versions of the homography and epipole must be employed in Eq. (4) and the homography estimate $\bar{\mathbf{V}}$ computed with DLT needs to be denormalized using $\mathbf{L}_3{}^{-1}\bar{\mathbf{V}}\mathbf{L}_2$.

In practice, the set of available matching point triplets is almost certain to contain errors due to false matches and errors in the localization of image corners. Hence, in order to prevent such errors from corrupting the computed homography estimate, the group of DLT constraints should be employed within a robust regression framework. In our case, the Least Median of Squares (LMedS) robust estimator is employed to iteratively sample random sets of nine constraints, recover an estimate of matrix $\mathbf{V}$ from each of them and find the estimate that is consistent with the majority of the available constraints. To ensure that those random sets arise from points having a good spatial distribution over the image, random sampling is based on the bucketing technique of [14]. Finally, the set of constraints having the largest support (i.e. the inliers) is employed to recompute $\mathbf{V}$ with least squares.

Since the DLT constraints minimize an algebraic error term with no physical meaning, the estimate computed by LMedS is refined by a nonlinear minimization process that involves a geometric criterion. Letting $d(\mathbf{x},\ \mathbf{y})$ represent the Euclidean distance between the inhomogeneous points represented by $\mathbf{x}$ and $\mathbf{y}$, the nonlinear refinement minimizes the following sum of squared distances

$$\sum_i \left(d(\mathbf{p}_i^{''}, \mathbf{V}\mathbf{p}_i^{'} + \kappa_i\mathbf{e}^{''})^2 + d(\mathbf{p}_i^{'}, \mathbf{V}^{-1}\mathbf{p}_i^{''} - \frac{||\mathbf{p}_i^{''}||}{||\mathbf{V}\mathbf{p}_i^{'} + \kappa_i\mathbf{e}^{''}||}\kappa_i\mathbf{V}^{-1}\mathbf{e}^{''})^2 \right) \quad (7)$$

with respect to $\mathbf{V}$. This criterion involves the mean symmetric transfer error between actual and transferred points in the two images and is minimized by applying the Levenberg-Marquardt iterative algorithm as implemented by MINPACK's LMDER routine, initialized with the estimate provided by LMedS. To safeguard against point mismatches, the nonlinear refinement is performed using only the point features that correspond to inliers of the LMedS homography estimate.

[1] Notice that all available point matches are assumed to originate from actual image points (i.e. corners) and do not include any ideal points whose third coordinate is zero.

Having presented the basic 3-frame chaining operation, it is straightforward to extend it to handle a sequence of more than three views. For example, in order to track the plane in a new image I_4, the homography $\mathbf{V}$ computed in the previous step between frames I_2 and I_3 becomes the new $\mathbf{U}$ for the triplet I_2, I_3 and I_4. Note also that the epipolar geometry of frames I_2 and I_3 has been computed from the previous iteration, therefore only the epipolar geometry between frames I_3 and I_4 needs to be estimated during this step. A final remark concerning the extension of the chaining operation to more than three frames is that the estimation of $\mathbf{V}$ can benefit from point trajectories that are longer than three frames: If, for example, a four-frame point trajectory is available for images I_1, I_2, I_3 and I_4, the constraints generated by the triplet I_1, I_3 and I_4 can be combined with those arising from I_2, I_3 and I_4. This variant of chaining from multiple triplets can be carried out by maintaining a small moving window of past frames.

3.2 Reducing the DOFs of Plane Tracking

In the following, the basic method of the previous section will be modified, aiming to derive a model involving fewer, therefore easier to estimate, degrees of freedom (i.e. free variables). It has already been mentioned that the entire group of all possible homography matrices between two images lies in a subspace spanned by the 4 primitive homographies of Eq. (1). This implies that given the primitive homographies for frames I_2 and I_3, the rows $\mathbf{v}_i^T$ of matrix $\mathbf{V}$ in Eqs. (6) depend on four rather than nine parameters. Therefore, the process described in section 3.1 can be slightly modified to estimate the coefficients λ_i making up $\mathbf{V}$ instead of directly estimating the latter. In other words, both the linear and the nonlinear estimation processes that have been described above are performed with four rather than nine unknowns. This reduction in the dimensionality of the problem is of utmost importance since fewer degrees of freedom entail less computation time for the homography (particularly for the nonlinear refinement) as well as more accurate estimates. We have found experimentally that the execution time for plane tracking using the formulation involving λ_i is by an order of magnitude shorter from that required when estimating $\mathbf{V}$ directly.

4 Experimental Results

The performance of a prototype C implementation of the proposed method has been evaluated based on several test image sequences. The current implementation performs chaining using constraints arising from three frames at a time. Representative results from two of the conducted experiments are given in this section. The point features employed throughout all experiments have been extracted and matched automatically. Using the resulting matching features, the epipoles were computed by finding the kernels of the fundamental matrices estimated using an implementation of [14]. In all experiments, the spatial extend of the employed planes has been defined manually using a polygon in the first frame. Following this, the plane homography between the first two images that is necessary for bootstrapping plane tracking (i.e. $\mathbf{U}$ in Section 3.1), is estimated from the point matches lying within the specified polygon. Alternatively, plane tracking could have been bootstrapped by applying to the first pair of images an automatic plane detection algorithm.

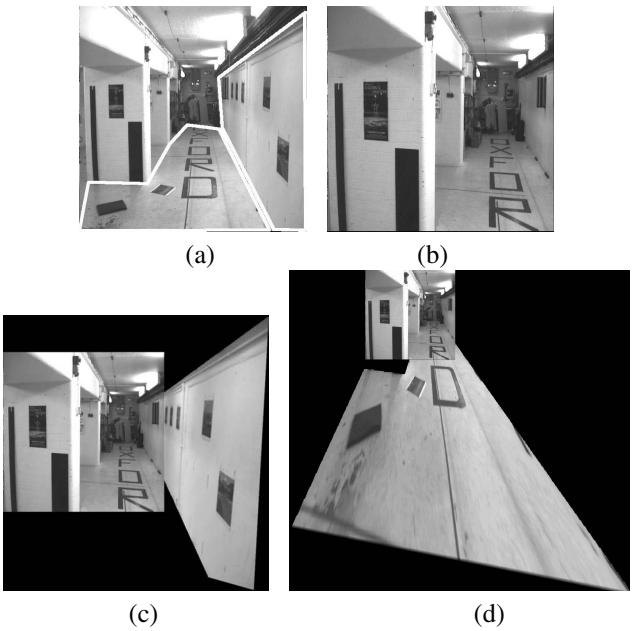

Figure 1. (a), (b) two views of a basement (courtesy of the Oxford Visual Geometry Group). The two polygons in (a) delineate the planar regions tracked between images (a) and (b). (c) right wall warped and stitched with (b), (d) floor warped towards and stitched with (b); see text for explanation.

The first experiment was performed on the well-known "basement" image sequence, two frames of which (namely 0 and 8) are shown in Figs. 1(a) and (b). This sequence consists of 11 frames acquired by a camera mounted on a mobile robot as it approached the scene while smoothly turning left. The plane corresponding to the right corridor wall was tracked from frame 0 to frame 8 using the proposed method. Then, by employing the estimated homography, the right wall from frame 0 was warped towards frame 8. Fig. 1(c) shows the warped wall stitched with frame 8. A second plane, namely the one corresponding to the floor, was also tracked between frames 0 and 8. Fig. 1(d) shows the result of warping the floor plane from frame 0 towards frame 8 and stitching them together. As it is clear from the results, the accuracy of the homographies estimated using

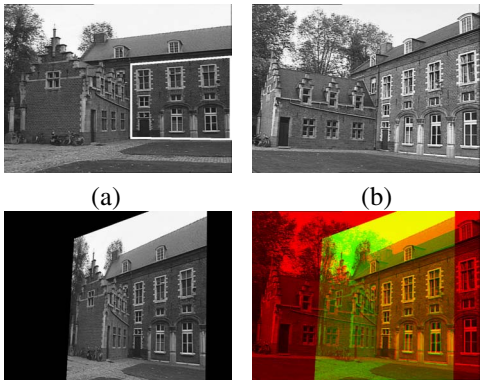

Figure 2. (a), (b) first and last images from the Arenberg castle sequence (courtesy of the University of Leuven VISICS Group), (c) first image warped towards the last using the estimated homography, (d) warped image in (c) superimposed on (b) using the red and green color channels.

the proposed method is satisfactory in both cases. Excluding the time required to detect and match corners between successive frames, the average running times for tracking the wall and floor planes in the whole sequence, were respectively 63 and 66 ms per frame on an Intel P4@1.8 GHz running Linux.

In order to quantitatively evaluate the performance of plane tracking, the floor plane was tracked from frame 0 to frame 10 and then back to frame 0, reversing the order of intermediate frames. This effectively simulates a camera motion that is circular, i.e. ends at the location where it started. Composing the pairwise homographies estimated by the plane tracker, the floor's homography from the first frame through the last and back to itself can be estimated. Ideally, this homography should be equal to the identity matrix. In practice, the deviation in the position of floor points transferred using this homography from their actual locations in the first frame, indicates the accuracy of plane tracking. The RMS error corresponding to the 91 transferred floor points was found to be 19.3 pixels, corresponding to an average RMS error of 0.91 pixels for each of the 21 frames involved in tracking. However, since certain floor points correspond to mismatches or poorly localized corners, a more appropriate error measure is given by the root median square (RMedS) error, which was found to be equal to 8.47 pixels or on average 0.40 pixels per tracked frame.

The second experiment employs another well-known image sequence, the first and last frames of which are shown in Figs. 2(a) and (b). The sequence depicts the Arenberg castle in Belgium and consists of 22 frames acquired with a handheld camera. Applying the proposed method, the 3D plane defined by the rightmost wall outlined in Fig. 2(a) was tracked throughout the whole sequence. Fig. 2(c) illustrates the result of warping the first frame towards the last using the estimated homography. To aid in the evaluation of this result, Fig. 2(d) shows it superimposed on Fig. 2(b), using different color channels for each image. As can be clearly seen, image warping according to the estimated homography successfully registers the plane's image in Fig. 2(a) with that in Fig. 2(b). In this case, the average running time for plane tracking was 84 ms per frame. The plane of the right wall was again tracked from the first to the last frame and back (for a total of 43 frames) and the RMS and RMedS errors in this case were 31.7 and 18.3 pixels, amounting to average errors of 0.73 and 0.43 pixels per frame respectively.

5 Conclusion

This paper has presented a method for tracking a 3D plane across an image sequence. The method is based on a homography "chaining" operation that exploits geometric constraints arising from all available point matches, without the need of segmenting them into planar and non planar ones. The "chaining" operation involves the estimation of a quadruple of plane parameters, which is achieved using a combination of linear and nonlinear optimization techniques. No restrictions are imposed on plane motion or the viewed spatial extend of the tracked plane, which is actually allowed to be outside the field of view of all but the first pair of images. Moreover, possible mismatches among the employed point features are handled in a robust manner. A scheme for camera tracking that builds upon the proposed method is given in [6].

References

[1] S. Avidan and A. Shashua. Tensor Embedding of the Fundamental Matrix. In *Proc. of post-ECCV SMILE'98*, LNCS 1506, pages 47–62, 1998.

[2] S. Avidan and A. Shashua. Threading Fundamental Matrices. *IEEE Trans. on PAMI*, 23(1):73–77, Jan. 2001.

[3] A. Criminisi, I. Reid, and A. Zisserman. Duality, Rigidity and Planar Parallax. In *Proc. of ECCV'98*, pages 846–861, 1998.

[4] R. Hartley. In Defense of the 8-Point Algorithm. *IEEE Trans. on PAMI*, 19(6):580–593, Jun. 1997.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[6] M. Lourakis and A. Argyros. Efficient, Causal Camera Tracking in Unprepared Environments. *Computer Vision and Image Understanding Journal*, 99(2):259–290, Aug. 2005.

[7] C. Rother and S. Carlsson. Linear Multi View Reconstruction and Camera Recovery Using a Reference Plane. *IJCV*, 49(2/3):117–141, 2002.

[8] A. Shashua and S. Avidan. The Rank-4 Constraint in Multiple View Geometry. In *Proc. of ECCV'96*, volume 2, pages 196–206, 1996.

[9] A. Shashua and N. Navab. Relative Affine Structure: Canonical Model for 3D from 2D Geometry and Applications. *IEEE Trans. on PAMI*, 18(9):873–883, Sep. 1996.

[10] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless Tracking using Planar Structures in the Scene. In *Proc. of Int'l Symp. on Augm. Reality*, 2000.

[11] B. Triggs. Plane + Parallax, Tensors and Factorization. In *Proc. of ECCV'00*, pages 522–538, 2000.

[12] J. Wang and E. Adelson. Representing Moving Images with Layers. *IEEE Trans. on Image Processing*, 3(5):625–638, Sep. 1994.

[13] L. Zelnik-Manor and M. Irani. Multi-Frame Estimation of Planar Motion. *IEEE Trans. on PAMI*, 22(10):1105–1116, Oct. 2000.

[14] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. *AI Journal*, 78:87–119, 1995. Detailed version in INRIA RR-2273.