Detecting Planes In An Uncalibrated Image Pair

Manolis I.A. Lourakis, Antonis A. Argyros and Stelios C. Orphanoudakis
Institute of Computer Science
Foundation for Research and Technology - Hellas (FORTH)
Vassilika Vouton, P.O.Box 1385, GR 711 10
Heraklion, Crete, GREECE
{lourakis,argyros,orphanou}@ics.forth.gr

Abstract

Plane detection is a prerequisite to a wide variety of vision tasks. This paper proposes a novel method that exploits results from projective geometry to automatically detect planes using two images. Using a set of point and line features that have been matched between images, the method exploits the fact that every pair of a 3D line and a 3D point defines a plane and utilizes an iterative voting scheme for identifying coplanar subsets of the employed feature set. The method does not require camera calibration, circumvents the 3D reconstruction problem, is robust to the existence of mismatched features and is applicable either to stereo or motion sequence images. Sample results from the application of the proposed method to real imagery are also provided.

1 Introduction

Due to their abundance in man-made environments, as well as to their attractive geometric properties, planes are commonly used in various vision tasks. As reported in the literature, planes have been successfully employed in diverse applications such as feature matching [12, 14] and grouping [6], camera self-calibration [19], obstacle detection [13], 3D reconstruction and scene analysis [9, 3, 1, 2, 10], camera relative positioning [15], object recognition [16], visual measurement [4], image mosaicing [22] and augmented reality [17]. Elegant theoretical results related to planes have also been derived [20]. However, despite their popularity, most vision algorithms based on the existence of planes rely on manual or even unspecified schemes for detecting such planes in images. In this respect, plane detection constitutes a preprocessing step, which precedes the exploitation of any constraints imposed by planarity. As such, this step should be accomplished without resorting to 3D reconstruction of the viewed scene, in contrast to what is usually proposed by traditional approaches such as [8].

Existing methods for plane detection that avoid 3D reconstruction are typically based on the extraction and matching of sparse geometric features from images. Sinclair and Blake [18], for example, detect the planes present in a scene by employing pairs of matched points extracted from a stereo pair and clustering them into coplanar sets using the two projective invariants defined by quintuples of coplanar points. The main shortcoming of this approach is that the values assumed by these invariants are sensitive to errors in the localization of image points. Therefore, their comparison is not a trivial issue. Starting

This work was partially supported by EU IST-2001-34545 project LifePlus.

BMVC 2002 doi:10.5244/C.16.57

with a rough initial estimate, Fornland and Schnörr [5] propose a two-step method for locating the dominant plane[1] present in a scene by iteratively solving for both the plane homography and the stereo point correspondence. Lourakis and Orphanoudakis [13] solve a similar problem by employing a robust estimation technique to identify the dominant plane as the one whose induced homography conforms to the motion of the majority of previously matched features. Both [5] and [13] share the drawback of making implicit assumptions regarding the minimum fraction of point features belonging to the dominant plane, a fact that limits their applicability in certain cases.

In this work, a novel method for detecting coplanar sets of point and line features in two images is proposed. The method is based on the observation that every feature pair comprised of a 3D line and a 3D point defines a plane and employs an iterative voting scheme for identifying coplanar subsets of the set of features that are matched between the two images. The use of projective geometry eliminates the need for camera calibration and 3D reconstruction, while robust estimation techniques guard against feature mismatches. The rest of the paper is organized as follows. Section 2 provides some preliminary concepts along with the notation that will be used for the development of the method. Section 3 presents an analytical result which is exploited in Section 4 for developing the proposed plane detection method. Experimental results from a prototype implementation are reported in Section 5. The paper is concluded with a brief discussion in Section 6.

2 Background and Notation

In the following, vectors and arrays will appear in boldface and the symbol $\simeq$ will be used to denote equality of vectors up to a scale factor. 3D points or lines appear in uppercase letters, while their image projections appear in lowercase letters (e.g. $\mathbf{P}$ and $\mathbf{p}$). Using projective (homogeneous) coordinates, an image point (p_x, p_y) is represented by the 3×1 column vector $\mathbf{p} = (p_x, p_y, 1)^T$. A line defined by the equation $\mathbf{l}^T \cdot \mathbf{p} = 0$ is also represented in terms of projective coordinates using the vector $\mathbf{l}$. The line defined by two points $\mathbf{p_1}$ and $\mathbf{p_2}$ is given by the cross product $\mathbf{p_1} \times \mathbf{p_2}$.

A well-known constraint for a pair of perspective views of a rigid scene is the *epipolar constraint*. This constraint states that for each point in one of the images, the corresponding point in the other image must lie on a straight line. Assuming that no calibration information is available, the epipolar constraint is expressed mathematically by a 3×3 singular matrix, known as the *fundamental matrix* and denoted by $\mathbf{F}$. Another important concept in projective geometry is the *plane homography* $\mathbf{H}$, a nonsingular 3×3 matrix which relates two uncalibrated retinal images of a 3D plane. More specifically, if $\mathbf{p}$ is the projection in one view of a point on the plane and $\mathbf{p}'$ is the corresponding projection in a second view, then the two projections are related by a linear projective transformation:

$$\mathbf{p}' \simeq \mathbf{H}\mathbf{p} \tag{1}$$

A similar equation relates a pair of corresponding planar lines $\mathbf{l}$ and $\mathbf{l}'$ in two views:

$$\mathbf{l}' \simeq \mathbf{H}^{-T}\mathbf{l}, \tag{2}$$

where $\mathbf{H}^{-T}$ denotes the inverse transpose of $\mathbf{H}$. For a more detailed treatment of the application of projective geometry to computer vision, the interested reader is referred to [7].

[1] Dominant is the plane on which the majority of the extracted features lie.

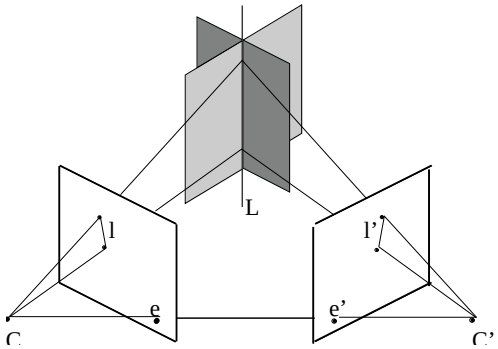


Figure 1: A 3D line $\mathbf{L}$ projects to image lines $\mathbf{l}$ and $\mathbf{l}^{'}$ and defines a pencil of 3D planes that gives rise to a single parameter family of homographies between the two images.

3 The 3D Plane Defined by a Line - Point Pair

In this section, an analytical expression describing the homography induced by the plane defined by a 3D line $\mathbf{L}$ and a 3D point $\mathbf{P} \notin \mathbf{L}$ is derived. This expression will be employed later for identifying sets of coplanar point and line features. In Fig. 1, $\mathbf{L}$ is the common intersection of a pencil of 3D planes containing it. As shown in [7], the homographies induced by these planes are given by an equation depending on a single parameter μ, specifically

$$\mathbf{H}(\mu) = [\mathbf{l}^{'}]_{\times}\mathbf{F} + \mu \mathbf{e}^{'}\mathbf{l}^{T}, \quad \mu \in R \tag{3}$$

In Eq.(3), $\mathbf{l}$ and $\mathbf{l}^{'}$ are the projections of $\mathbf{L}$ in the two images, $\mathbf{F}$ is the underlying fundamental matrix, $\mathbf{e}^{'}$ is the *epipole* in the second image defined by $\mathbf{F}^{T}\mathbf{e}^{'} = \mathbf{0}$ and $[\mathbf{l}^{'}]_{\times}$ is the skew symmetric matrix representing the vector cross product (i.e. $\forall\ \mathbf{x},\ [\mathbf{l}^{'}]_{\times}\mathbf{x} = \mathbf{l}^{'} \times \mathbf{x}$). Assuming that a point $\mathbf{P}$ not on $\mathbf{L}$ projects to the corresponding image points $\mathbf{p}$ and $\mathbf{p}^{'}$, let $\mathbf{q} = \mathbf{p} \times \mathbf{p}^{'}$. Obviously, $\mathbf{p}^{'} \cdot \mathbf{q} = 0$ and, since $\mathbf{p}^{'} \simeq \mathbf{H}(\mu)\mathbf{p}$, it can be shown that

$$([\mathbf{l}^{'}]_{\times}\mathbf{F}\mathbf{p}) \cdot \mathbf{q} + \mu(\mathbf{e}^{'}\mathbf{l}^{T}\mathbf{p}) \cdot \mathbf{q} = 0. \tag{4}$$

Solving Eq. (4) for μ and substituting the solution into Eq. (3) yields the homography of the plane defined by $\mathbf{L}$ and $\mathbf{P}$. Note that all the above computations are based on entities that are directly computed from images.

4 Coplanar Feature Detection

The key idea behind the proposed method is to compute all the homographies defined by line and point pairs and then use them to identify the features lying on the most prominent 3D plane, that is the one containing the largest set of corresponding features. Following this, the features belonging to the most prominent plane are removed from further consideration and the process is repeated until no more planes can be found. In the remainder of this section, the proposed method is explained in more detail; Fig. 2 lists its steps in pseudocode.

Given a pair of images taken from different viewpoints, the method starts by extracting point and line features from both images. Point and line features are then matched between

1. *Extract and match point and line features*
2. *Estimate* $\mathbf{F}$ *and the epipoles*
3. *For each line - point pair not yet assigned to some plane do*
   3.a Compute the underlying homography with Eqs. (3) - (4)
   3.b Use the computed homography to transfer all features from the first image to the second
   endfor
4. *Choose the line - point pair yielding the homography that correctly transfers the largest number of features*
5. *Reestimate the chosen homography using LMedS on the correctly transferred features*
6. *Refine the LMedS homography estimate with an iterative nonlinear method using the point inliers identified in step 5*
7. *Use the refined estimate to transfer all features between images*
8. *Label the features that are transferred correctly as coplanar and remove them from the feature set*
9. *Repeat steps 3 through 8 until the number of coplanar points is less than a threshold*
10. *Use the estimated homographies to reassign all features to planes*

Figure 2: An overview of the proposed method; see text for details.

images using conventional, correlation based techniques and the underlying fundamental matrix is estimated from matched point features as described in [21]. Next, the epipoles are directly computed from the estimated fundamental matrix using SVD. The homographies induced by each line and point pair are then computed as described in Section 3. To avoid degenerate cases where a point lies too close to its paired line, a line - point pair is ignored if the area of the triangle formed by the point and the endpoints of the line is too small. A line - point pair is also ignored in the case that the homography computed from it is not of rank 3. Each of the computed homographies is used along with Eqs. (1) and (2) to transfer (i.e. predict the location of) every matched feature in one image to the other. For each of the transferred features, a vote is tallied in favor of the homography which most accurately predicts the location of its corresponding feature. In addition, each feature is associated with the homography that most accurately transfers it between images. In the case of points, the proximity between a transferred point feature and its matching counterpart is quantified using the Euclidean distance. For lines, a similar metric based on the Euclidean distances of their corresponding endpoints cannot be employed. This is because line extraction rarely preserves the endpoints of corresponding line segments in two images. Therefore, the proximity of line segments is quantified by a combination of the difference between their distances from the image center and the difference in their orientations. In order for a feature to be considered as accurately transferred, its distance from its matching counterpart should be less than a tight threshold[2].

Upon termination of the above voting process, the homography that receives the largest number of votes is assumed to be induced by the most prominent 3D plane. To improve the accuracy of this initial homography estimate, the full set of features associated with it

[2]Maximum acceptable distances are 2.0 pixels for points and 2.0^o in orientation discrepancy for lines.

is employed to re-estimate it using Least Median of Squares (LMedS) robust regression [13]. This estimate is then further refined by applying the Levenberg-Marquardt algorithm to iteratively minimize a nonlinear criterion that involves the mean symmetric transfer error between actual and transferred points in the two images [7]. In order to safeguard against point mismatches, the previous step is performed using only the point features that correspond to inliers of the LMedS homography estimate. This last homography estimate is used to transfer all features between images. Features whose transferred location is close enough to their matching counterpart are labeled as being coplanar and are removed from the feature set. The process just outlined reiterates for the remaining features until the number of the detected coplanar features drops below a threshold. When the process terminates, all the estimated homographies are used to transfer features between the two images and each feature is assigned to the plane inducing the homography that most accurately transfers it to its matching counterpart. This final step accounts for features that have been assigned to some plane early in the execution of the method but are actually closer to a plane that had not yet been discovered at that point. Features not transferred correctly by any homography are not assigned to a plane.

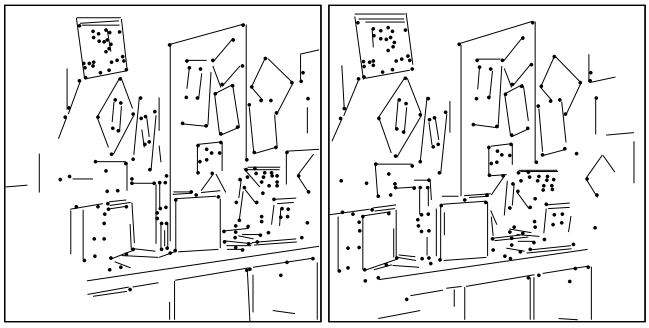
5 Experimental Results

The performance of a prototype implementation of the proposed method has been evaluated based on a set of test images. The point and line features employed throughout all experiments have been extracted and matched automatically using the relaxation labeling techniques presented in [21] and [11] respectively. Representative results from two of the conducted experiments are given in this section.

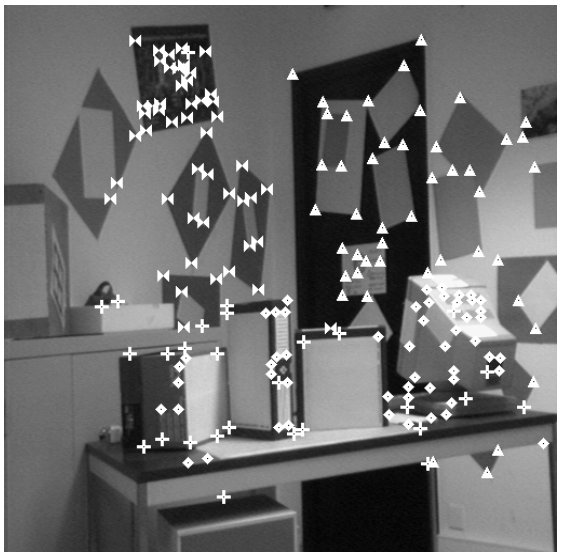
The first experiment refers to the image pair shown in Figures 3(a) and 3(b), depicting an office corner and a table in front of it. In order to facilitate the visualization of disparities between the images, Figure 3(c) shows the two images superimposed in different color channels. Note that the camera motion between the two images is such that the epipoles are located outside the images, estimated at (1684.5, 215.0) and (1623.2, 216.1) for the first and second image respectively. As it is well known, the accurate estimation of the epipoles in this case is not an easy task. The matched features that were fed to the proposed method are shown in Figs. 3(d) and 3(e). The plane detection method segmented those features into three sets, namely the left wall, the right wall and the objects on the table. Although this last set does not correspond to an actual 3D surface, its features roughly belong to a 3D plane and thus their apparent motion between images is fairly accurately captured by a homography. In Fig. 3(f), the point features assigned to the detected planes are marked with different symbols for each plane; to avoid cluttering the image, line features are not shown. The plane detection results are more easily interpreted if the spatial extent of the detected planes is shown. To achieve this for some plane, the second image (Fig. 3(b)) is warped towards the first image (Fig. 3(a)) using the corresponding homography. This results in the cancellation of the plane's motion, which appears stabilized between the first and the warped images. Figures 4(a), 4(b) and 4(c) show the warped images corresponding to the homographies of the three detected planes, superimposed on the image in Fig. 3(a). Clearly, warping the image in Fig. 3(b) according to the homography estimated for each plane, accurately registers the plane's image. This indicates the correctness of the assignment of features to planes.

The second experiment employs the image pair in Figures 5(a) and 5(b), showing a

(a) (b) (c)

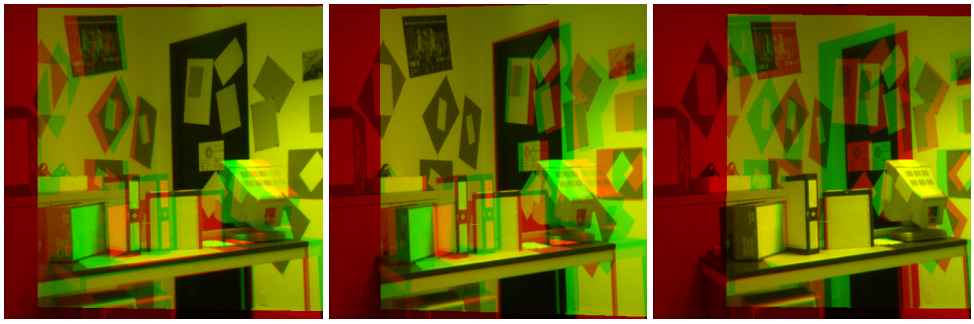


(d) (e)



(f)

Figure 3: (a), (b) two views of an office corner (courtesy of the INRIA RobotVis group), (c) original views superimposed in different color channels, (d), (e) the features matched between the two views: some mismatches, especially for lines, are noticeable, (f) corners segmented into planes by the proposed method: points not assigned to any plane are marked with crosses, points on the right wall with triangles, points on the left wall with "bow ties" and points on the table with rhombi.



(a) (b) (c)

Figure 4: Original images (Figs. 3(a) and 3(b)) superimposed after canceling out the motion corresponding: (a) to the right wall, (b) to the left wall and (c) to the objects on the table. Planes are accurately registered in all three cases.

laboratory scene having multiple planes with chessboard texture. As in the first experiment, the epipoles are found to be located outside the images, namely at (1984.5, 263.5) and (1705.5, 257.7). This experiment is challenging not only due to the rather large number of planes contained in the images, but also due to the less reliable feature matches caused by the repeating patterns present in the scene. To aid in the understanding of the spatial arrangement of planes, Fig 5(c) shows a top view of the imaged scene. It is important to note that this view is supplied for illustration purposes only, i.e. no processing relies upon it. Figure 5(d) shows the two images in Figs. 5(a) and 5(b) superimposed in different color channels. The features matched between images are shown in Figs. 5(e) and (f). The proposed plane detection method segmented those features into four sets, namely the back wall, the left foreground plane, the middle foreground plane and the "floor" plane. As illustrated in Fig. 5(g) which shows the point features assigned to the detected planes using different symbols, the right foreground plane is not detected as a distinct plane. The main reason for this is the special spatial arrangement of the three foreground scene planes. It is clear from Fig. 5(c) that the extensions in space of both the left and middle foreground planes pass through the right foreground one. Thus, the apparent motion of the right foreground plane can be well approximated by the homographies computed solely from features lying on the left and middle foreground planes. As previously explained, Figs. 6(a) through (d) show the warped images corresponding to the homographies of the detected planes, superimposed on the image in Fig. 5(a). Again, it is evident that the proposed method has successfully detected the planes contained in the viewed scene.

6 Conclusion

This paper has presented a fully automatic method for detecting the planes present in a scene using a set of matched point and line features. The method searches for homographies with the aid of an iterative voting scheme based on pairs of point and line features. Neither camera calibration nor 3D scene reconstruction is required, while possible mismatches among the employed feature sets are handled in a robust manner. Experiments on two different image pairs provided evidence regarding the method's performance.

The use of line features results in considerable reductions in the size of the space to

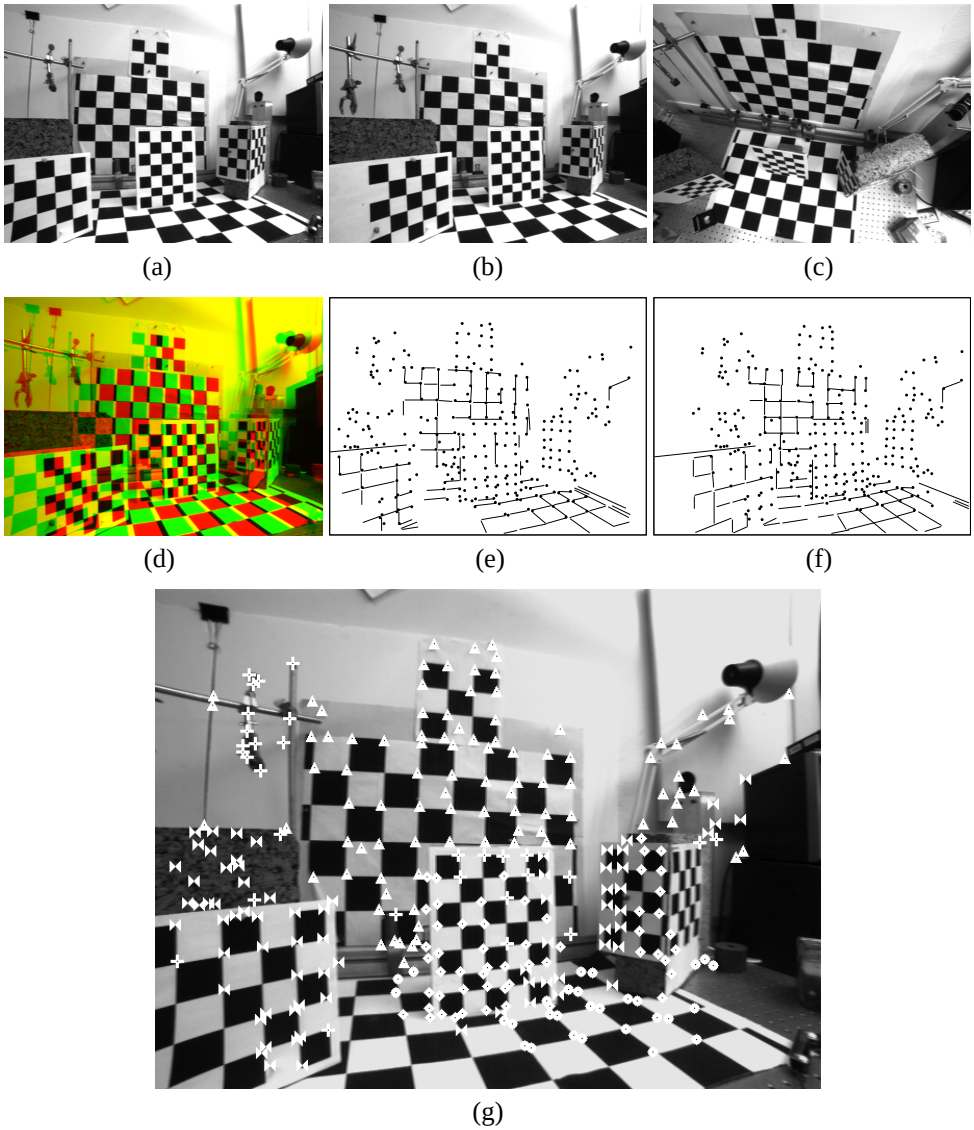
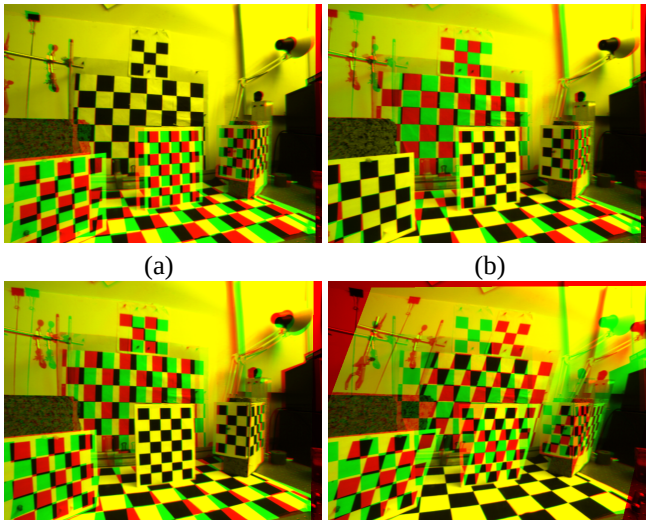


Figure 5: (a), (b) two views of a laboratory scene containing multiple planes (courtesy of the MIT AI lab), (c) top view of the imaged scene included for reference, (d) first two views superimposed in different color channels, (e), (f) the features matched between the two first views, (g) corners segmented into planes by the proposed method: points not assigned to any plane are marked with crosses, points on the back wall with triangles, points on the left foreground plane with "bow ties", points on the middle foreground plane with rhombi and points on the "floor" with circles. Note that the left and right parts of the right foreground plane belong to the same space extended planes defined by the left and the middle foreground planes respectively. Similarly, the points close to the right edge of the middle foreground plane are found to belong to the space extended plane defined by the left foreground plane.



(a) (b)

(c) (d)

Figure 6: Original images (Figs. 5(a) and 5(b)) superimposed after canceling out the motion corresponding: (a) to the back wall, (b) to the foreground left plane, (c) to the foreground middle plane and (d) to the "floor" plane.

be searched, since in the case where only point features were used, the computation of the homography induced by a single plane would require a triplet of features (i.e. three corresponding points) and the epipoles or a quadruple of features if the fundamental matrix is not available. Besides, real images usually contain less lines than points, making the exhaustive search for homographies among all line - point pairs computationally feasible. It should also be pointed out that compared to point features, line segments are more robust to occlusions and can be more accurately localized in images.

References

[1] C. Baillard and A. Zisserman. Automatic Reconstruction of Piecewise Planar Models from Multiple Views. In *Proc. of CVPR'99*, pages 559–565, 1999.

[2] A. Bartoli, P. Sturm, and R. Horaud. A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene. Technical Report RR-4070, MOVI/INRIA, Oct. 2000.

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

[4] A. Criminisi, I. Reid, and A. Zisserman. A Plane Measuring Device. *IVC*, 17:625–634, 1999.

[5] P. Fornland and C. Schnörr. A Robust and Convergent Iterative Approach for Determining the Dominant Plane from Two Views Without Correspondence and Calibration. In *Proc. of CVPR'97*, pages 508–513, Jun. 1997.

[6] L. Van Gool, M. Proesmans, and A. Zisserman. Planar Homologies as a Basis for Grouping and Recognition". *IVC*, 16:21–26, 1998.

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

[8] A. Imiya and I. Fermin. Voting Method for Planarity and Motion Detection. *IVC*, 17:867–879, 1999.

[9] M. Irani, P. Anandan, and D. Weinshall. From Reference Frames to Reference Planes: Multi-View Parallax Geometry and Applications. In *Proc. of ECCV'98*, pages 829–845, 1998.

[10] R. Kaucic, R. Hartley, and N. Dano. Plane-based Projective Reconstruction. In *Proc. of ICCV'01*, volume I, pages 420–427, 2001.

[11] M.I.A. Lourakis. Establishing Straight Line Correspondence. Technical Report 208, ICS/FORTH, Aug. 1997. Available at ftp://ftp.ics.forth.gr/tech-reports/1997.

[12] M.I.A. Lourakis, S.T. Halkidis, and S.C. Orphanoudakis. Matching Disparate Views of Planar Surfaces Using Projective Invariants. *IVC*, 18:673–683, 2000.

[13] M.I.A. Lourakis and S.C. Orphanoudakis. Visual Detection of Obstacles Assuming a Locally Planar Ground. In *Proc. of ACCV'98*, volume 2, pages 527–534, Jan. 1998.

[14] M.I.A. Lourakis, S.V. Tzurbakis, A.A. Argyros, and S.C. Orphanoudakis. Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing Planes. In *Proc. of ICPR'00*, pages 419–422, 2000.

[15] R. Mohr, L. Morin, and E. Grosso. Relative Positioning with Uncalibrated Cameras. In J.L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 440–460. MIT Press, Cambridge, MA, 1992.

[16] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. Planar Object Recognition Using Projective Shape Representation. *IJCV*, 16(1):57–99, Sep. 1995.

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

[18] D. Sinclair and A. Blake. Quantitative Planar Region Detection. *IJCV*, 18(1):77–91, Apr. 1996.

[19] B. Triggs. Autocalibation from Planar Scenes. In *Proc. of ECCV'98*, pages 89–105, 1998.

[20] L. Zelnik-Manor and M. Irani. Multiview Constraints on Homographies. *IEEE Trans. on PAMI*, 24(2):214–223, 2002.

[21] 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.

[22] I. Zoghlami, O. Faugeras, and R. Deriche. Using Geometric Corners to Build a 2D Mosaic from a Set of Images. In *Proc. of CVPR'97*, pages 420–425, 1997.