Feature Transfer and Matching in Disparate Stereo Views through the Use of Plane Homographies

Manolis I.A. Lourakis, Stavros V. Tzurbakis,
Antonis A. Argyros, and
Stelios C. Orphanoudakis,
Senior Member, IEEE

Abstract—Many vision tasks rely upon the identification of sets of corresponding features among different images. This paper presents a method that, given some corresponding features in two stereo images, matches them with features extracted from a second stereo pair captured from a distant viewpoint. The proposed method is based on the assumption that the viewed scene contains two planar surfaces and exploits geometric constraints that are imposed by the existence of these planes to first transfer and then match image features between the two stereo pairs. The resulting scheme handles point and line features in a unified manner and is capable of successfully matching features extracted from stereo pairs that are acquired from considerably different viewpoints. Experimental results are presented, which demonstrate that the performance of the proposed method compares favorably to that of epipolar and tensor-based approaches.

Index Terms—Feature correspondence, feature transfer, projective transformations, plane homography, fundamental matrix, trifocal tensor, wide baseline matching.

1 Introduction

A fundamental problem in computer vision that appears in different forms in tasks such as discrete motion estimation, 3D scene reconstruction, feature based stereo, object recognition, image registration, camera self-calibration, visual servoing, image-based rendering, augmented reality, etc., is that of determining the correspondence among sets of image features extracted from different views of the same scene. The correspondence problem is very difficult to solve and a general solution is yet to be found. The difficulty stems mainly from the fact that common physical phenomena, such as changes in illumination, occlusion, perspective distortion, transparency, etc., have a significant impact on the appearance of a scene in different views, thus complicating the feature matching process. Most approaches for dealing with the correspondence problem rely upon the assumption that the photometric and geometric properties of matching features in different images are similar. Thus, feature matching is based on the affinity of pixel intensities and the similarity of 2D geometric descriptions, such as image location for points and length or orientation for lines. Such properties, however, are not preserved under general perspective projection, which implies that the correspondence methods that exploit them (e.g., [1], [2], [3]) are applicable only to images that have been acquired from adjacent viewpoints, for which disparities are small.

However, by using images of a certain scene that have been acquired from very different viewpoints, the estimation of structure from motion becomes more accurate, the flexibility in image acquisition is increased, and fewer views are required for effectively sampling the environment. To facilitate the matching of features extracted from such images, two basic alternative strategies have been proposed in the literature. The first, often adopting a semi-automatic approach, is to assume that a priori information regarding the images of the viewed scene is available. For example, Georgis et al. [4] require that the projections of four corresponding coplanar points at arbitrary positions are known. Schmid and Zisserman [5] assume that either the epipolar geometry of two views or the trifocal geometry of three views is known. Faugeras and Robert [6] assume the availability of the epipolar geometry among three views, to predict the location in the third view of features that are matched between the first two views. Related to the previous method are the techniques reported in [7], [8], which synthesize novel views based on a set of reference views and knowledge of the associated epipolar and trifocal geometry, respectively.

The second approach to determining feature correspondence in the presence of large disparities is to exploit quantities that remain invariant under perspective projection and can be directly computed from images. By exploiting the fact that a perspective transformation of a smooth surface can be locally approximated by an affine distortion, affine texture invariants are used in [9] and [10] to match features across widely separated views. Other methods employ geometric constraints by exploiting projective invariants involving the location of image features. The lack of general-case view invariants [11] obliges the latter approaches to make assumptions regarding the structure of the viewed scene. Since planes are very common in human-made environments and have attractive geometric properties, the most popular assumption made by such methods is that the features to be matched lie on a single 3D plane in the scene. Meer et al. [12], for example, employ projective and permutation invariants to obtain representations of coplanar point sets that are insensitive to both projective transformations and permutations in the labeling of the set. Pritchett and Zisserman [13] rely on the existence of suitable coplanar feature structures, namely parallelograms, to estimate local plane homographies which are then used to compensate for viewpoint differences and generate putative point matches. Lourakis et al. [14] propose a "generate and test" scheme for expanding small sets of corresponding planar features to large ones that are related by general projective transformations.

In this work, we propose a novel method for matching features between widely separated uncalibrated stereoscopic views, which is based on transferring matching features from two images of one stereo pair to another. An initial investigation of this method was reported in [15]. The method is based on the assumption that the viewed scene contains two planar surfaces and employs scene constraints that are derived with the aid of projective geometry. Points and lines are treated in a unified manner and their correspondence is determined by feature transfer between images that are related by arbitrary projective transformations. In conjunction to the matching method, a new technique for identifying coplanar feature sets is described. The rest of the paper is organized as follows: The proposed method is presented in Section 2. Experimental results obtained from the application of the proposed method to synthetic and real images are presented and discussed in Section 3. The paper is concluded with a brief discussion in Section 4.

2 The Proposed Method

A high-level block diagram of the proposed method is shown in Fig. 1a, while Fig. 1b illustrates the main transformations involved in the computations.

The method starts by establishing feature correspondences within each of the two stereo pairs (blocks labeled (A) in Fig. 1a). Based on these correspondences, the two dominant planes that are assumed to exist in the viewed scene are identified in each stereo pair and their homographies are estimated (blocks (B) in Fig. 1a). Planar features are then matched between the disparate stereo pairs by exploiting appropriate projective invariants (block (C) in Fig. 1a). Finally, using geometric constraints that are imposed by the matched 3D planes, correspondence is established among features of the two stereo pairs that do not belong to the planes (block (D) in

• *The authors are with the Institute of Computer Science, Foundation for Research and Technology, Hellas (FORTH), Vassilika Vouton, PO Box 1385, 71110 Heraklion Crete, Greece.*
E-mail: {lourakis, tzurbak, argyros, orphanou}@ics.forth.gr.

Manuscript received 17 Jan. 2001; revised 24 Jan. 2002; accepted 24 July 2002.
Recommended for acceptance by M. Irani.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113488.

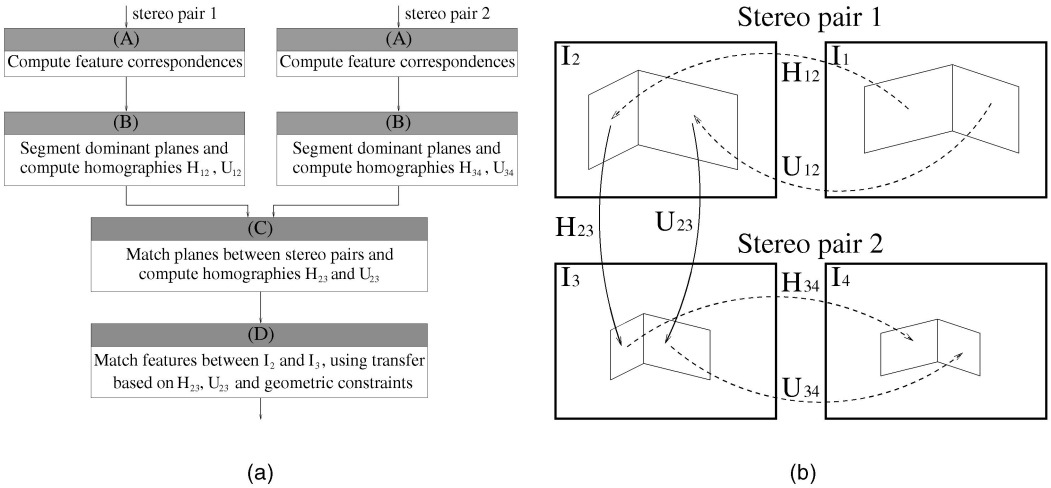

Fig. 1. High-level description of the proposed method: (a) block diagram and (b) projective transformations defined by the two viewed 3D planes.

Fig. 1a). The original contribution of this paper lies in the algorithms related to blocks (B) and (D); blocks (A) and (C) are implemented based on techniques existing in the literature. Assuming a basic familiarity with projective geometry [16], the following sections provide more details on the algorithmic steps outlined above.

2.1 Computing Intrastereo Correspondences

The computation of intrastereo feature correspondences is based on existing techniques. More specifically, point correspondences are obtained using the cross-correlation based technique proposed in [2], whereas line correspondences are obtained by employing the technique presented in [3]. The latter technique extracts line segments by detecting edges and then fitting straight lines to the resulting edge points. Following this, potential matches between line segments are formed and then disambiguated through relaxation labeling.

2.2 Segmenting Planar Surfaces

In this section, an iterative method for identifying coplanar sets of corresponding features is presented; more details can be found in [17]. The method relies upon the availability of a set of corresponding points and lines extracted from two stereoscopic images. First, the fundamental matrix defined by the two stereo images is estimated using [18]. Following this, the homography induced by the plane defined by a 3D line $\mathbf{L}$ and a 3D point $\mathbf{P} \notin \mathbf{L}$ is computed. As illustrated in Fig. 2a, $\mathbf{L}$ is the common intersection of a pencil of 3D planes containing it. The homographies of this pencil's planes are given by a single parameter equation [5]:

$$\mathbf{H}(\mu) = [\mathbf{l}']_{\times}\mathbf{F} + \mu \mathbf{e}'\mathbf{l}^{T}, \quad \mu \in R. \tag{1}$$

In (1), $\mathbf{l}$ 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 $\mathbf{P}$ projects to the corresponding image points $\mathbf{p}$ and $\mathbf{p}'$, let $\mathbf{q} = \mathbf{p} \times \mathbf{p}'$. Clearly, $\mathbf{p}' \cdot \mathbf{q} = 0$ and, since $\mathbf{p}' \simeq \mathbf{H}(\mu)\mathbf{p}$, (1) leads to:

$$([\mathbf{l}']_{\times}\mathbf{F}\mathbf{p}) \cdot \mathbf{q} + \mu(\mathbf{e}'\mathbf{l}^{T}\mathbf{p}) \cdot \mathbf{q} = 0. \tag{2}$$

The parameter μ for the plane defined by $\mathbf{L}$ and $\mathbf{P}$ is determined by solving (2). Then, the corresponding homography is obtained by substituting the solution into (1).

Based on the above computation, a method for segmenting the two most prominent 3D planes, i.e., the ones containing the two largest sets of corresponding features, can be devised as follows. Initially, the homographies of the planes defined by all pairs of corresponding lines and points are computed. Next, each of these homographies is used to predict, that is transfer, the location of every feature from one image to the other. A vote is cast in favor of the homography for which the predicted location best approximates the true location of the matching feature. In addition, features are associated with the homographies that correctly predict their location in the other image. Upon termination of this voting process, the two planes corresponding to the homographies that receive the largest and second largest numbers of votes are identified as the two most prominent ones. The homographies pertaining to the two most prominent planes are then reestimated using LMedS robust regression [19] on the constraints derived from the full sets of features assigned to them [14]. Using these two estimates as initial solutions, the two homographies are further refined by applying the Levenberg-Marquardt algorithm to the corresponding LMedS inliers for iteratively minimizing a nonlinear criterion that involves the mean symmetric transfer error between actual and transferred point features in the two images [16].

2.3 Matching Coplanar Features between Distant Views

Suppose that point and line features have been extracted in two disparate views of the same planar surface. In order to match these features, the algorithm described in [14] is employed. Briefly, this algorithm employs a randomized search scheme, guided by geometric constraints derived using the *two-line two-point* projective invariant, to form hypotheses regarding the correspondence of small subsets of the two feature sets that are to be matched. The validity of such hypotheses is then verified by using the subsets that are assumed to be matching to recover the associated plane homography and predict more matches. Due to the fact that the algorithm is based on a projective invariant, it is able to correspond features that have been extracted from images acquired from considerably different viewpoints. Moreover, it does not make use of any photometric information, since the latter usually varies significantly between disparate views.

2.4 Matching Noncoplanar Features between Distant Views

For clarity of notation, the following conventions are made. Each image is identified by a positive index i, with I_1 and I_2 representing the first stereo image pair and I_3 and I_4 representing the stereo pair acquired from the distant location (see also Fig. 1b). The same indices are also used for identifying corresponding points between the two images, i.e., a 3D point $\mathbf{P}$ gives rise to two corresponding points $\mathbf{p}_1$ and $\mathbf{p}_2$ in images I_1 and I_2. For a pair of images I_i and I_j, $\mathbf{H}_{ij}$ and $\mathbf{U}_{ij}$ denote the homographies induced between them by the two 3D planes, respectively. Furthermore, it is assumed that intrastereo point and line matches (i.e., between

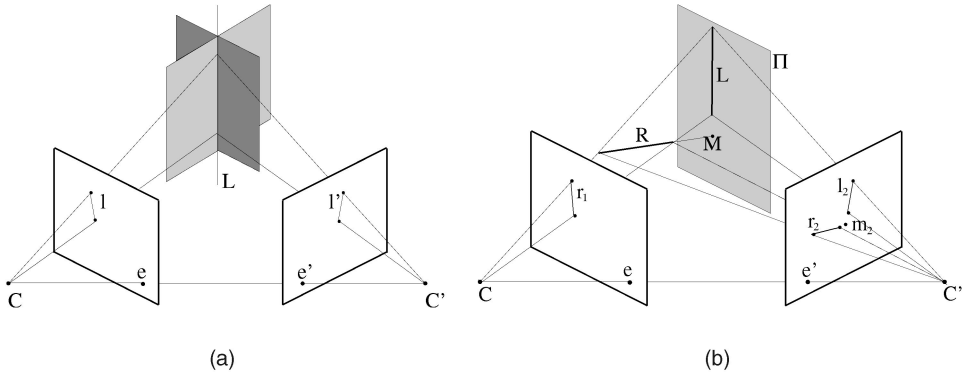

Fig. 2. Stereoscopic views of planes. (a) A 3D line $\mathbf{L}$ projects to image lines l and l' and defines a pencil of 3D planes that gives rise to a single parameter family of homographies between the two images. (b) The projection in the right image of the point of intersection $\mathbf{M}$ of the 3D line $\mathbf{R}$ with the plane Π is given by the intersection of image lines $\mathbf{r}_2$ and l_2. Lines $\mathbf{r}_2$ and l_2 are the projections in the right image of the 3D lines $\mathbf{R}$ and $\mathbf{L}$, $\mathbf{L}$ being $\mathbf{R}$'s projection on Π.

images I_1 - I_2 and I_3 - I_4) have been obtained (as described in Section 2.1), the two most prominent coplanar feature sets have been identified in both stereo pairs (as described in Section 2.2) and the features of I_3 that lie on the two dominant planes have been matched with those of I_2 (as described in Section 2.3). In the remainder of this section, a geometric construction for solving the interstereo matching problem, i.e., matching the features of I_3 that are not on the two planes with the features of I_1 and I_2, is described. Image I_4 contributes only to identifying the features of the two planes in the second stereo pair.

The proposed matching scheme exploits geometric constraints to first transfer nonplanar features from I_2 to I_3. Once this transfer is achieved, matching is realized by examining the features extracted in the neighborhood of the predicted feature location. The following sections describe in more detail the steps involved in feature transfer and matching.

2.4.1 Feature Transfer

The core idea behind transferring point features not belonging to the two dominant planes of the scene between distant views, is to express each such point as the intersection of lines with end points on the planes. Then, these end points can be transferred between the distant views using the available plane homographies, enabling the reconstruction of the related lines in the distant views. The intersection of the reconstructed lines defines the location of the point to be transferred in the distant view. Intuitively, the two scene planes constitute a "reference frame" for the nonplanar 3D points, thus providing an algebraic representation of geometric structure at the image level. Such parameterizations for structure and motion recovery when the imaged scenes are piecewise planar, have appeared only recently [20].

To further clarify the construction, consider Fig. 3, where $\mathbf{P}$ and $\mathbf{O}$ are two 3D points, with $\mathbf{P}$ lying on one of the two most prominent planes and $\mathbf{O}$ not lying on either of these two planes. Using their projections $\mathbf{p}_1$ and $\mathbf{o}_1$ in I_1, the line $\mathbf{r}_1 \simeq \mathbf{p}_1 \times \mathbf{o}_1$ is defined, which corresponds to $\mathbf{r}_2 \simeq \mathbf{H}_{12}\mathbf{p}_1 \times \mathbf{o}_2$ in image I_2. The underlying 3D line $\mathbf{R}$ intersects the other plane at a point projecting to $\mathbf{m}_2$ in I_2. As shown in [21], $\mathbf{m}_2$ is defined by $\mathbf{m}_2 \simeq \mathbf{r}_2 \times \mathbf{U}_{12}^{-T}\mathbf{r}_1$ (see also Fig. 2b). Similarly, if $\mathbf{Q}$ is a second point on the first plane, a line $\mathbf{s}_1 \simeq \mathbf{q}_1 \times \mathbf{o}_1$ is defined in I_1, corresponding to line $\mathbf{s}_2 \simeq \mathbf{H}_{12}\mathbf{q}_1 \times \mathbf{o}_2$ in I_2. The image in I_2 of the intersection of line $\mathbf{S}$ with the second plane is given by $\mathbf{n}_2 \simeq \mathbf{s}_2 \times \mathbf{U}_{12}^{-T}\mathbf{s}_1$. Thus, the projections in I_1 and I_2 of two lines $\mathbf{R}$ and $\mathbf{S}$ that intersect at point $\mathbf{O}$ have been constructed. Since the intersections of lines $\mathbf{R}$ and $\mathbf{S}$ with the two planes are known, their projections in I_3 can also be constructed as

$$\mathbf{r}_3 \simeq (\mathbf{H}_{23}\mathbf{p}_2) \times (\mathbf{U}_{23}\mathbf{m}_2)$$

and

$$\mathbf{s}_3 \simeq (\mathbf{H}_{23}\mathbf{q}_2) \times (\mathbf{U}_{23}\mathbf{n}_2),$$

where $\mathbf{p}_2 \simeq \mathbf{H}_{12}\mathbf{p}_1$ and $\mathbf{q}_2 \simeq \mathbf{H}_{12}\mathbf{q}_1$ are, respectively, the points corresponding to $\mathbf{p}_1$ and $\mathbf{q}_1$ in I_2. Given $\mathbf{r}_3$ and $\mathbf{s}_3$, the projection of point $\mathbf{O}$ in I_3 is simply $\mathbf{o}_3 \simeq \mathbf{r}_3 \times \mathbf{s}_3$.

Notice that the role of points $\mathbf{P}$ and $\mathbf{Q}$ can be assumed by any pair of distinct points lying on either of the two planes. In fact, the projection of point $\mathbf{O}$ in I_3 can be found by intersecting several lines between the two planes, which are formed as explained above. The over-constrained point of intersection of these lines is obtained using the LMedS robust estimator [19] and is tolerant to feature mismatches and errors in feature localization.

So far, only the case of transferring points to I_3 has been examined. In the case of line segments, it suffices to transfer their endpoints in I_3. For increased accuracy, more points on a given line $\mathbf{L}$ can be transferred in I_3, as follows. If $\mathbf{p}_1$ is a point on l_1 in I_1, its corresponding point in I_2 is given by $\mathbf{p}_2 \simeq \mathbf{F}_{12}\mathbf{p}_1 \times \mathrm{l}_2$, where $\mathbf{F}_{12}$ is the fundamental matrix corresponding to I_1 and I_2. Then, the construction of the previous paragraph can be repeated for finding $\mathbf{p}_3$. Assuming that several points lying on $\mathbf{L}$ have been transferred in I_3, the equation of l_3 can be determined by robust line fitting using the set of transferred points.

The main advantage of the method proposed here is that it relies on constraints arising from the scene structure, which are therefore independent of the camera positions. This is in contrast to the *epipolar transfer* method of Faugeras and Robert [6], which is based on intersecting corresponding epipolar lines and, thus, involves constraints related to the relative positions of the cameras. Epipolar transfer fails for 3D points being close to the *trifocal plane* defined by the three image optical centers, since in this case, pairs of corresponding epipolar lines in the third image are almost

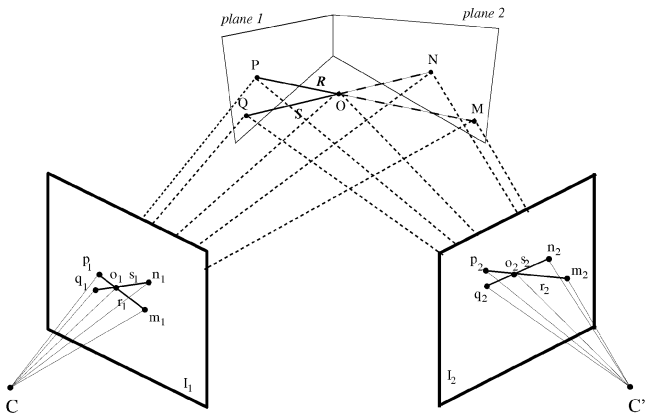

Fig. 3. Point $\mathbf{O}$ (not lying on either plane), together with points $\mathbf{P}$ and $\mathbf{Q}$ (lying on the first plane) defines the lines $\mathbf{R}$ and $\mathbf{S}$, respectively. These lines intersect the second plane at points $\mathbf{M}$ and $\mathbf{N}$. The projections of 3D points and lines in a stereoscopic view are also shown.

TABLE 1
The Effect of Noise on the Performance Index (PI) for Plane Segmentation

Noise sdev	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
PI mean (%)	97.3	97.1	97.3	95.6	91.6	87.2	82.4	79.1	76.2	74.7	73.5
PI sdev (%)	0.87	0.92	0.90	4.67	4.09	1.70	2.09	2.37	2.98	2.09	2.21

coincident. Epipolar transfer will also fail for all points in the degenerate case where all three optical centers are collinear. The limitations of epipolar transfer can be overcome by basing point transfer on the trifocal tensor [22], which elegantly encodes the geometry of three views. However, besides the fact that estimating the trifocal tensor is more complicated compared to estimating plane homographies, it will be demonstrated in Section 3 that, in order for this tensor to be accurately estimated, certain conditions related to the spatial distribution and the cardinality of matches among the three views have to be met.

2.4.2 Feature Matching

Noise in the images to be matched causes the location of transferred features in I_3 to differ slightly from the location of actual features extracted from it, even in the case that these features are indeed correct matches. To overcome this problem, we allow for some error by considering each feature $\mathbf{f}$ of I_2 to match the feature of I_3 that is closest to $\mathbf{f}$'s predicted location in I_3. Feature proximity is quantified by the Euclidean distance between the normalized homogeneous vectors representing transferred and extracted features.

3 Experimental Results

3.1 Experiments on Plane Segmentation

The plane segmentation technique described in Section 2.2 plays a key role in the proposed method. In order to quantitatively study the performance of plane segmentation under increasing amounts of noise, a set of experiments using simulated data has been carried out. Due to the high dimensionality of the problem, it is not feasible to carry out an exhaustive study of the effects induced on the performance of the technique by a systematic variation of all of the parameters involved. Therefore, a realistic scenario corresponding to an obstacle detection task was simulated. A camera was assumed to be overlooking a planar floor at a height of 1.5 meters, with its optical axis at an angle of 15^o with respect to the horizon. The simulated retina is 750×750 pixels and the focal length of the camera is equal to 700 pixels. It is also assumed that the camera is moving rigidly with 3D translational velocity of $(0.100, 0.181, 0.676)$ meters/frame. The synthetic scene viewed by the camera consists of 300 random 3D points and 60 random 3D lines, 100 and 20 of which are respectively assumed to lie on the simulated 3D floor. The heights of the 3D points not lying on the floor are assumed to be uniformly distributed in the range of $[0.15, 2.0]$ meters. The standard deviation of the zero mean Gaussian noise added to retinal projections varied between 0.0 and 2.0 pixels in steps of 0.2 pixels. The proposed plane segmentation technique was applied to the set of the simulated 2D matching features to classify them as being planar or nonplanar. The number of features that are correctly classified, divided by the total number of simulated features, was used as a performance index (PI) for the plane detection technique. To ensure that the results are independent of a particular configuration of the 3D features employed in the simulation, for each noise level the average performance in 100 independent trials is reported, with each trial involving a different, random set of 3D features. The mean and standard deviation of the PI for each noise level is summarized in Table 1, which shows that the method performs satisfactorily even in the presence of large amounts of noise.

3.2 Experiments on Feature Matching

This section reports two representative experiments involving the application of the proposed matching method to stereo pairs. As demonstrated by these experiments, the proposed method, when applicable, yields results of similar or even better quality compared to those obtained by epipolar or tensor-based feature transfer. The first experiment is performed using two stereo pairs of an indoor scene acquired with a fixed zoom digital camera. The first stereo pair is shown in Figs. 4a and 4b, while Fig. 4c illustrates the left image of the second stereo pair. The latter has been obtained after approaching the viewed scene. To ensure the readability of results, the 95 line segments that have been used by the proposed method are not shown. A total of 227 corresponding point pairs were identified in the first stereo pair, 202 of which were correct (see Figs. 4a and 4b). The two most prominent scene planes (namely the two walls in the background) contained 83 and 69 points in the first stereo pair, respectively. The ones that are visible in Fig. 4c are marked with +s and x's. Fig. 4c also shows the locations of nonplanar points that were matched in the second stereo pair using the proposed method (a total of 50 points). Aiming to assess the accuracy of feature transfer which is critical for matching, true point matches were determined manually in Fig. 4c and then the mean Euclidean distance between transferred and actual point locations was measured. The mean transfer error for the proposed method was found to be 1.59 pixels. By comparison, the mean error obtained from the epipolar transfer of [6] was 2.95 pixels. The estimation of the fundamental matrices required by [6] (specifically between the images of Fig. 4a and Fig. 4c and, Fig. 4b and Fig. 4c) was based on corresponding features located on the two prominent planes only. In order to investigate tensor-based point transfer, corresponding points on the two most prominent planes along with the associated homographies were employed for estimating the tensor defined by the images of Figs. 4a, 4b, and 4c. The tensor was estimated using a "plane plus parallax" approach (see [16, p. 393]), which uniquely determines it using the homographies induced by a reference plane and a pair of points off the plane. Each of the two available planes assumed the role of the reference plane and all point pairs not lying on it were considered for computing the tensor yielding the minimum mean transfer error for points on both planes. The mean error for this tensor was 1.77 pixels, very close to that of the proposed method. For each pair of points in Figs. 4a and 4b, the transferred point in Fig. 4c was computed as the point of intersection of the nine 2D constraint lines arising from the trilinearities defined by the tensor. In addition to the parallax-based approach, the tensor was also estimated using a generic method that does not explicitly exploit the existence of planes. The trifocal tensor in this case was estimated with the aid of the Projective Vision Toolkit[1] using the 124 point features lying on the two prominent planes in Figs. 4a, 4b, and 4c. The mean transfer error corresponding to the tensor estimated in this manner was 1.52 pixels.

Using the point matches determined by the proposed method, the self-calibration technique developed in [24] was employed to estimate the camera intrinsic calibration parameters. Then, the point matches along with the estimated intrinsic parameters were fed to the algorithm described by Zhang in [25] for determining the rigid 3D motion of the camera and reconstructing the scene

1. A free implementation of [23] by the Computational Video Group of the Institute for Information Technology of the National Research Council of Canada; see http://www.cv.iit.nrc.ca/~gerhard/PVT/.

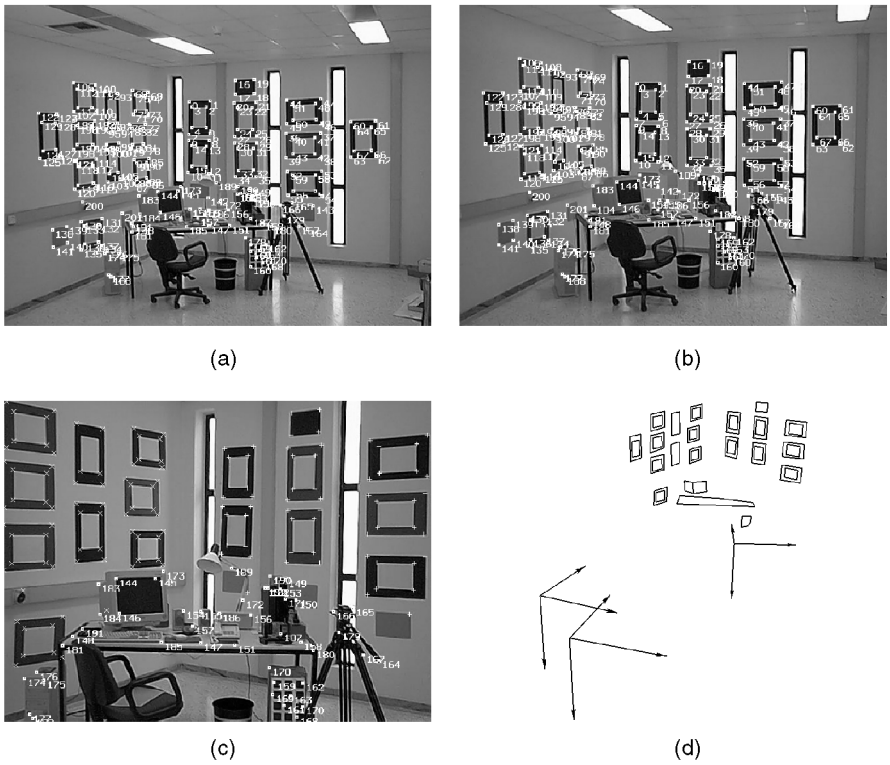

Fig. 4. Indoor scene experiment: (a) and (b) first stereo pair, (c) an image from the second stereo pair; corresponding points are labeled with identical numbers, and (d) side view of the reconstructed wireframe model and the recovered camera positions.

structure up to a scale. This reconstruction is provided as a means to indirectly evaluate the accuracy of the matches produced by the proposed feature matching algorithm. Fig. 4d shows a side view of a rough wireframe model which was obtained by defining a set of polygons which correspond to planar scene patches. The reconstructed 3D locations of the cameras are also shown. Clearly, the structure of the scene has been captured correctly.

The second experiment refers to the outdoor images of Fig. 5, showing part of FORTH's premises in Heraklion, Crete. Intrastereo point matches are again shown with identical labels in Figs. 5a and 5b, while points lying on the two scene planes are marked with the symbols + and x in the more distant view of Fig. 5c. Compared to the first, the second stereo pair has been acquired from a viewpoint that is further away from the imaged objects. In this experiment, the proposed algorithm did not employ the two most prominent planes of the scene (i.e., the two walls of the building in the background). The plane of the left building wall and the plane corresponding to the car in the foreground were used instead. This choice of scene planes was intended to test the performance of the method in the case of planes having small spatial extent and containing minute numbers of features. A total of 186 corresponding point pairs were identified in the first stereo pair, 165 of which were correct. The employed scene planes contained 43 and 29 points in the first stereo pair, respectively. The 78 nonplanar points that were matched in the second stereo pair are shown numbered in Fig. 5c. In this experiment, the mean transfer error was 1.79 pixels, while epipolar transfer [6] yielded a mean transfer error of 6.15 pixels. This difference in the performance of the two techniques is due to the fact that, for several of the transferred points, the associated interstereo epipolar lines are almost parallel and, thus, their points of intersection cannot be accurately computed. The method proposed in this paper is independent of the relative camera positions and therefore its accuracy is not affected in this case. The mean point transfer error pertaining to the tensor estimated with the parallax-based approach was 2.59 pixels. On the other hand, point transfer through the trifocal tensor estimated with the generic approach yielded a large mean transfer error, namely 9.21 pixels. This last result is due to the fact that the planar points available for estimating the tensor covered a small part of the image plane and were only 71, thus providing insufficient constraints for accurately estimating it. This is further exemplified by considering that when the set of corresponding point triplets used to estimate the tensor was expanded using 82 more matches that were determined *manually*, the mean transfer error dropped to 1.43 pixels only. At this point it should be stressed that a set of matched features having limited spatial extent and small cardinality is not uncommon when dealing with disparate views, since in this case, the large baselines and the limited overlap among viewing locations increase the difficulty of feature matching. As demonstrated by this experiment, the proposed method should be preferred over the generic tensor-based one when the matches among the three views are concentrated in small portions of the images or their number is small. The proposed method relies on the computation of fundamental matrices and homographies which have modest degrees of freedom (7 and 8, respectively). On the other hand, the trifocal tensor has 18 degrees of freedom and is therefore more difficult to estimate accurately from a set of weak constraints. As in the previous experiment, the point matches determined by the proposed method were used for self-calibrating the camera and reconstructing the 3D scene. A top view of the obtained model is shown in Fig. 5d. Notice that the right angle between the two background walls has been correctly recovered.

4 Conclusions

In this paper, a fully automatic method for matching image features between two disparate stereo pairs has been presented. The proposed method exploits geometric constraints arising from the structure of a scene, which are valid regardless of the viewpoints of images and can be computed without any knowledge of camera calibration. Multiple such constraints are defined for each feature, thereby increasing robustness by over-determining the solution. Second, the method is capable of handling images

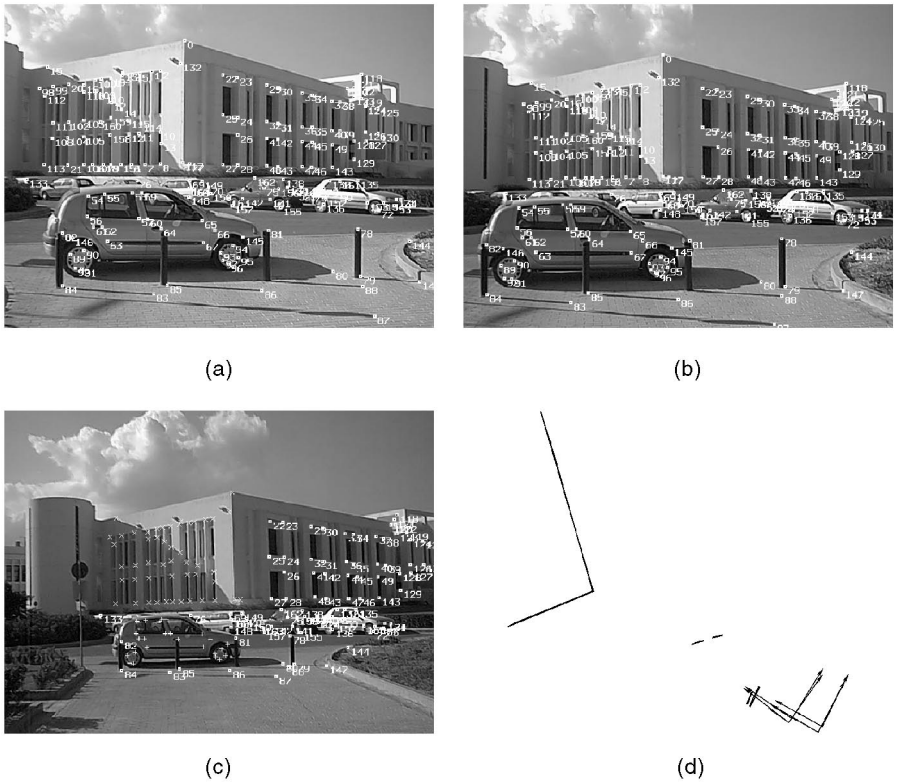

Fig. 5. Outdoor scene experiment: (a) and (b) first stereo pair, (c) an image from the second stereo pair, and (d) top view of the reconstruction.

that have been captured from significantly different viewpoints, despite effects due to illumination changes, perspective foreshortening, etc. Therefore, it is applicable in cases where tracking methods assuming small motion between images would fail. Third, it does not rely on the epipolar or trifocal geometry, whose recovery might be difficult in certain cases. Finally, the method handles points and lines in a unified manner by relying on the same principles for determining their correspondence.

REFERENCES

[1] C. Medioni and R. Nevatia, "Matching Images Using Linear Features," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 675-686, 1984.

[2] Z. Zhang, "A New and Efficient Iterative Approach to Image Matching," *Proc. Int'l Conf. Pattern Recognition,* pp. 563-565, 1994.

[3] M.I.A. Lourakis, "Establishing Straight Line Correspondence," Technical Report 208, Inst. of Computer Science, Foundation for Researh and Technology, 1997.

[4] N. Georgis, M. Petrou, and J. Kittler, "On the Correspondence Problem for Wide Angular Separation of Non-Coplanar Points," *Image and Vision Computing,* vol. 16, pp. 35-41, 1998.

[5] C. Schmid and A. Zisserman, "Automatic Line Matching Across Views," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 666-671, 1997.

[6] O. Faugeras and L. Robert, "What Can Two Images Tell Us about a Third One?" *Int'l J. Computer Vison,* vol. 18, no. 1, pp. 5-20, 1996.

[7] S. Laveau and O. Faugeras, "3-D Scene Representation as a Collection of Images and Fundamental Matrices," Technical Report RR-2205, INRIA, 1994.

[8] S. Avidan and A. Shashua, "Novel View Synthesis by Cascading Trilinear Tensors," *IEEE Trans. Visualization and Computer Graphics,* vol. 14, no. 4, pp. 293-306, 1998.

[9] A. Baumberg, "Reliable Feature Matching Across Widely SeparatedViews," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 774-781, 2000.

[10] T. Tuytelaars and L. Van Gool, "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions," *Proc. 11th British Machine Vision Conf.,* pp. 421-425, 2000.

[11] J.B. Burns, R.S. Weiss, and E.M. Riseman, "The Non-Existence of General-Case View Invariants," *Geometric Invariance in Computer Vision,* J.L. Mundy and A. Zisserman, eds., chapter 6, pp. 120-134, MIT Press, 1992.

[12] P. Meer, R. Lenz, and S. Ramakrishna, "Efficient Invariant Representations," *Int'l J. Computer Vison,* vol. 26, no. 2, pp. 137-152, 1998.

[13] P. Pritchett and A. Zisserman, "Wide Baseline Stereo Matching," *Proc. Int'l Conf. Computer Vision,* pp. 754-760, 1998.

[14] M.I.A. Lourakis, S.T. Halkidis, and S.C. Orphanoudakis, "Matching Disparate Views of Planar Surfaces Using Projective Invariants," *Image and Vision Computing,* vol. 18, pp. 673-683, 2000.

[15] M.I.A. Lourakis, S.V. Tzurbakis, A.A. Argyros, and S.C. Orphanoudakis, "Using Geometric Constraints for Matching Disparate StereoViews of 3D Scenes Containing Planes," *Proc. Int'l Conf. Pattern Recognition,* pp. 419-422, 2000.

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

[17] M.I.A. Lourakis, A.A. Argyros, and S.C. Orphanoudakis, "Detecting Planes In an Uncalibrated Image Pair," *Proc. 13th British Machine Vision Conf.),* pp. 587-596, 2002.

[18] 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," *Artifical Intelligence J.,* vol. 78, pp. 87-119, 1995.

[19] P.J. Rousseeuw, "Least Median of Squares Regression," *J. Am. Statistics Assoc.,* vol. 79, pp. 871-880, 1984.

[20] A. Bartoli and P. Sturm, "Structure and Motion from Two Views of a Piecewise Planar Scene," *Proc. Int'l Conf. Computer Vision,* vol. 1, pp. 593-598, 2001.

[21] O. Faugeras, "Stratification of 3-D Vision: Projective, Affine, and Metric Representations," *J. Optical Soc. Am. A,* vol. 12, no. 3, pp. 465-484, 1995.

[22] R.I. Hartley, "Lines and Points in Three Views and the Trifocal Tensor," *Int'l J. Computer Vision,* vol. 22, no. 2, pp. 125-140, 1997.

[23] P.H.S. Torr and A. Zisserman, "Robust Parameterization and Computation of the Trifocal Tensor," *Image and Vision Computing,* vol. 15, pp. 591-607, 1997.

[24] M.I.A. Lourakis and R. Deriche, "Camera Self-Calibration Using the Singular Value Decomposition of the Fundamental Matrix: From Point Correspondences to 3D Measurements," Technical Report RR-3748, INRIA, 1999.

[25] Z. Zhang, "Motion And Structure From Two Perspective Views: From Essential Parameters to Euclidean Motion Via Fundamental Matrix," *J. Optical Soc. Am. A,* vol. 14, no. 11, pp. 2938-2950, 1997.