FORTH-ICS / TR-268 February 2000

Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing Planes†

M.I.A. Lourakis‡, S.V. Tzurbakis, A.A. Argyros
and S.C. Orphanoudakis

Several vision tasks rely upon the availability of sets of corresponding features among images. This paper presents a method which, given some corresponding features in two stereo images, addresses the problem of matching 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 predict the location of image features in the second stereo pair. The resulting scheme handles point and line features in a unified manner and is capable of successfully matching features extracted from stereo pairs acquired from considerably different viewpoints. Experimental results from a prototype implementation demonstrate the effectiveness of the approach.

† This work was funded in part under the VIRGO research network of the TMR Programme (EC Contract No ERBFMRX-CT96-0049)

‡ Currently with RobotVis, INRIA Sophia-Antipolis, France

Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing Planes

M.I.A. Lourakis, S.V. Tzurbakis, A.A. Argyros and S.C. Orphanoudakis

Computer Vision and Robotics Laboratory
Institute of Computer Science (ICS)
Foundation for Research and Technology — Hellas (FORTH)
Science and Technology Park, Heraklio, Crete
POBox 1385, GR-711-10 Greece
http://www.ics.forth.gr/proj/cvrl
E-mail: {lourakis,tzurbak,argyros,orphanou}@ics.forth.gr
Tel: +30 (81) 391703, Fax: +30 (81) 391601

Technical Report FORTH-ICS / TR-268 — February 2000

Several vision tasks rely upon the availability of sets of corresponding features among images. This paper presents a method which, given some corresponding features in two stereo images, addresses the problem of matching 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 predict the location of image features in the second stereo pair. The resulting scheme handles point and line features in a unified manner and is capable of successfully matching features extracted from stereo pairs acquired from considerably different viewpoints. Experimental results from a prototype implementation demonstrate the effectiveness of the approach.

1 Introduction

A fundamental problem in computer vision, appearing in different forms in tasks such as discrete motion estimation [9], feature-based stereo [5], object recognition [24], image registration [3], camera self-calibration [16, 13], image-based rendering [10], etc, is that of determining the correspondence among sets of image features extracted from different views of the same scene. The correspondence problem has proved to be very difficult to solve automatically and a general solution is still lacking. The difficulty mainly stems from the fact that common physical phenomena such as changes in illumination, occlusion, perspective distortion, transparency, etc, might have a tremendous impact on the appearance of a scene in different views, thus complicating their matching. Most approaches for dealing with the correspondence problem rely upon the assumption that the photometric and geometric properties of matching features are similar among images. Thus, feature matching is based on the affinity of pixel intensities and the similarity of 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. [26, 19, 12]), are applicable only to images that have been acquired from adjacent viewpoints, for which disparities are small.

Images whose viewpoints differ considerably have desirable properties for certain types of applications. In such cases, for example, structure from motion estimation becomes more accurate, the flexibility in image acquisition is increased and fewer views are required for effectively sampling the environment. In order to facilitate the matching of features extracted from such images, two alternative strategies have been proposed in the literature. The first is to adopt a semi-automatic approach and assume a priori knowledge of geometric constraints that are satisfied by the different views. For example, Georgis et al [8] require that the projections of four corresponding coplanar points in general position are known. Initially, a "virtual image" is constructed by warping the first image towards the second according to the homography of the known plane. Then, correct matches are found with the aid of a Hough-like scheme exploiting the observation that the virtual and second images share the same epipole. Schmid and Zisserman [23] assume that either the epipolar geometry of two views or the trifocal geometry of three views is known. Then, they match lines across images using a mechanism which combines the available

geometric constraints with graylevel information. Pritchett and Zisserman [21] rely on the existence of coplanar feature groups defined by closed polygons consisting of four edges to estimate local homographies, which are then used to compensate for viewpoint differences and generate putative point matches. Subsequently, the Ransac algorithm is employed to verify consistent matches through the recovery of the epipolar geometry. Faugeras and Robert [7] assume the availability of feature correspondences between two views as well as knowledge of the epipolar geometry among three views to predict the location in the third view of features from the first two views. Related to the previous method are the techniques reported in [11, 1] which synthesize novel views based on a set of reference views and knowledge of the associated epipolar and trifocal geometry respectively.

The second alternative approach for determining feature correspondence in the presence of large disparities, is to exploit quantities that remain unchanged under perspective projection and can be directly computed from the employed image features (i.e. projective invariants). Due to the lack of general-case view invariants [4], such approaches need to make assumptions regarding the structure of the viewed scene. The most common assumption made in previous work is that the features to be matched lie on a single 3D plane in the scene. Planes are very common in human-made environments and have attractive geometric properties. Meer et al [20], for example, employ projective and permutation invariants to obtain representations of coplanar point sets that are insensitive to both projective transformations and permutations of the labeling of the set. Combinatorial search coupled with a voting scheme enables them to identify corresponding points in two views. Lourakis et al [14] treat point and line features in a unified manner and propose a randomized search scheme, guided by geometric constraints, to form hypotheses regarding the correspondence of small subsets of the 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 underlying plane homography and predict more matches.

In this work, we propose a novel method for propagating matching features from two stereo images to another stereo pair that is assumed to have been acquired from a significantly different viewpoint. 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 in images that are related by arbitrary projective transformations can be

determined. In conjunction to the matching method, a new technique for segmenting coplanar image feature sets is developed. The rest of the paper is organized as follows. Section 2 presents an overview of some preliminary concepts that are essential for the development of the proposed method. Section 3 presents the method itself. Experimental results from an implementation of the method applied to real images are presented in section 4. The paper is concluded with a brief discussion in section 5.

2 Notation and Background

In the following, vectors and arrays will be written in boldface and the symbol $\simeq$ will be used to denote equality of vectors up to a scale factor. 3D points or lines are written in capitals, while their image projections are designated by small letters. Using projective (homogeneous) coordinates, image point (p_x, p_y) is represented by the 3×1 column vector $\mathbf{p} = (p_x, p_y, 1)^T$. A line having equation of the form $\mathbf{l}^T \cdot \mathbf{p} = 0$ is also delineated by projective coordinates using the vector $\mathbf{l}$. Since projective coordinates are defined up to a scalar, all vectors of the form $\lambda\mathbf{p}$, with $\lambda \neq 0$, are equivalent, regardless of whether they represent a point or a line. The line defined by two points $\mathbf{p_1}$ and $\mathbf{p_2}$ is given by the cross product $\mathbf{p_1} \times \mathbf{p_2}$, and, due to duality, the point of intersection of two lines $\mathbf{l_1}$ and $\mathbf{l_2}$ is equal to $\mathbf{l_1} \times \mathbf{l_2}$. For a more detailed treatment of the application of projective geometry to computer vision, the interested reader is referred to [6].

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. More specifically, assuming that $\mathbf{p}$ and $\mathbf{p}'$ are two homogeneous 3×1 vectors defining a pair of corresponding points in two images, they satisfy $\mathbf{p}'^T\mathbf{F}\mathbf{p} = 0$. Matrix $\mathbf{F}$ is the fundamental matrix, playing a central role in applications involving the recovery of motion and structure information from uncalibrated images and $\mathbf{F}\mathbf{p}$ is the epipolar line corresponding to $\mathbf{p}$. Since $\mathbf{F}$ is defined up to a scale factor and is of rank two, it has seven degrees of freedom and can be estimated using the epipolar constraint for at least seven pairs of corresponding points. Another important concept from projective geometry is the plane homography (also known as plane

projectivity or plane collineation) $\mathbf{H}$, which relates two uncalibrated views of a plane in 3D. Each pair of views of the same 3D plane Π defines a nonsingular 3×3 matrix $\mathbf{H}$ with the following properties. Assuming that $\mathbf{p}$ is the projection in the first view of a point belonging to Π and $\mathbf{p}'$ is the corresponding projection in the second view, then the two projections are related by $\mathbf{p}' \simeq \mathbf{H}\mathbf{p}$. A similar equation relates a pair of corresponding lines $\mathbf{l}$ and $\mathbf{l}'$ in two views as $\mathbf{l}' \simeq \mathbf{H}^{-T}\mathbf{l}$, where $\mathbf{H}^{-T}$ denotes the inverse transpose of $\mathbf{H}$. Matrix $\mathbf{H}$ can be estimated only up to an unknown scale factor, thus it has eight degrees of freedom. As can be seen from the last two equations, a single pair of corresponding features provides two constraints regarding $\mathbf{H}$, therefore the homography can be recovered using at least four pairs of corresponding coplanar points or lines. Owing to the non-singularity of $\mathbf{H}$, a homography is easier to estimate compared to a fundamental matrix [17]. In practice, the most accurate estimates of both a fundamental matrix and a plane homography are obtained using non linear minimization techniques; more details can be found in [26] and [17] respectively.

3 The Proposed Method

The proposed method starts by identifying in each stereo pair the features lying on the two planes that are assumed to exist in the viewed scene. Then, planar features are matched between stereo pairs by exploiting quantities that are invariant under perspective projection. Finally, using geometric constraints that are imposed by the matched 3D planes, the correspondence among features of the two stereo pairs that do not belong to the planes is established. The steps of the algorithm outlined above are explained in greater detail in the sections that follow.

3.1 Segmenting Planar Surfaces

In this subsection, an iterative method for identifying coplanar sets of corresponding features is presented. Suppose that a set of corresponding points and lines extracted from two stereoscopic images is available. We start by computing the homography induced by the plane defined by a 3D line $\mathbf{L}$ and a 3D point $\mathbf{P} \notin \mathbf{L}$. As illustrated in Fig. 1, $\mathbf{L}$ is the common intersection of a pencil of 3D planes containing it. In [23], Schmid

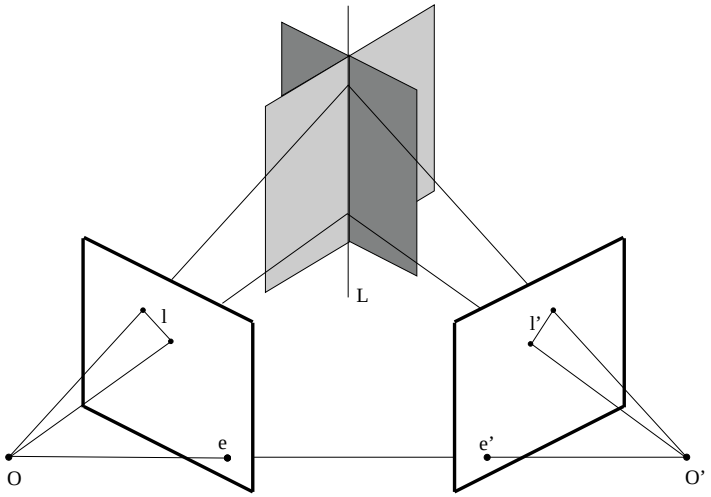


Figure 1: A 3D line $\mathbf{L}$ projects to image lines $\mathbf{l}$ and $\mathbf{l}^{'}$ and defines a pencil of 3D planes that gives rise to a single parameter family of homographies between the two images.

and Zisserman show that the homographies of this pencils' planes are given by a single parameter equation:

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

In Eq.(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 $\mathbf{q} = \mathbf{p} \times \mathbf{p}^{'}$. Obviously, $\mathbf{p}^{'} \cdot \mathbf{q} = 0$, and since $\mathbf{p}^{'} \simeq \mathbf{H}(\mu)\mathbf{p}$, it turns out that

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

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

Based on the above computation, a method for segmenting the two most prominent 3D planes, (i.e. the ones containing the two largest numbers of corresponding features) operates as follows. Initially, the homographies of the planes defined by all pairs of corresponding lines and points are computed. Following this, each homography is used to predict the location of every feature from one image in the other one. A vote is casted in favor of the homography for which the predicted location best approximates the true location of the matching feature. In addition, this feature is assumed to belong to the

plane defined by the homography in question. Upon termination of this voting process, the two planes that receive the largest and second largest numbers of votes are identified as the two most prominent ones. Following this, the homographies of the two most prominent planes are re-estimated using robust regression on the constraints derived from the full sets of features assigned to them [15, 14]. Using those two estimates as initial solutions, the two homographies are refined using the non-linear criterion of [17].

3.2 Matching Coplanar Features

Suppose that two sets of points and lines extracted from a pair of disparate views of a planar surface are available. In order to match those features, the algorithm that we have developed 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 plane homography and predict more matches. Owing to the fact that the algorithm is based on a projective invariant, it is capable of corresponding features that have been extracted from images having considerably different viewpoints. Moreover, it does not make use of any photometric information, since the latter usually varies significantly between disparate views. More details regarding the algorithm can be found in [14].

3.3 Matching Non-Coplanar Features

For clarity of notation, the following conventions are made. Each image is identified by a positive index i, with images I_1 and I_2 assumed to form the first stereo pair and I_3 and I_4 be the stereo pair acquired from the distant viewpoint. The same indices are also used for identifying corresponding points between the images, e.g. the 3D point $\mathbf{P}$ gives rise to two corresponding points $\mathbf{p_1}$ and $\mathbf{p_2}$ in images I_1 and I_2. The plane homography that is induced between images i and j by one of the two 3D planes is denoted by $\mathbf{H_{ij}}$; similarly, $\mathbf{U_{ij}}$ denotes the homography induced by the other 3D plane. Furthermore, it is assumed that intra-stereo point and line matches (i.e. between images I_1 - I_2 and I_3 - I_4) have been obtained using conventional sparse feature matching techniques and that

the two most prominent coplanar feature sets have been identified in both stereo pairs as explained in section 3.1. Also, the features of I_3 that lie on the two planes are assumed to have been matched with those of I_2, as described in section 3.2. In the remainder of this subsection, a geometric construction for solving the inter-stereo matching problem, that is matching the features of I_3 that are not on the two planes with the features of I_1 and I_2, is described. As will soon become clear, image I_4 contributes only to segmenting the features of the two planes in the second stereo pair.

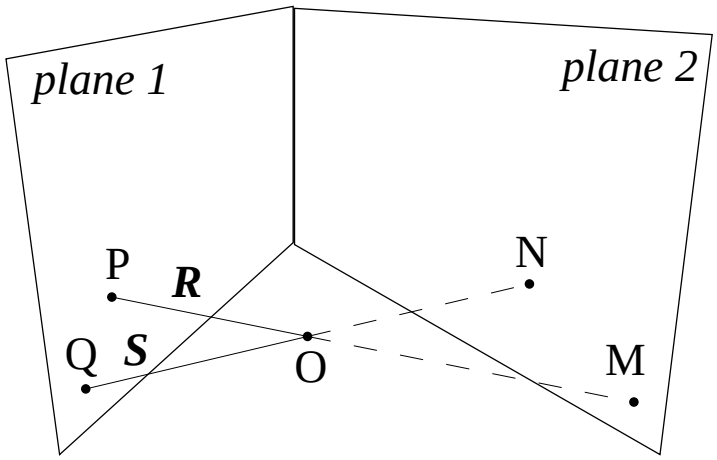


Figure 2: The 3D point **O** defines the lines **R** and **S** with the points **P** and **Q** that lie on the first plane. These lines intersect the second plane at points **M** and **N**.

Referring to Fig.2, let **P** and **O** be two 3D points, with **P** being on one of the two most prominent planes and **O** not being 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 $\mathbf{r_2} \simeq \mathbf{H_{12}p_1} \times \mathbf{o_2}$ in image I_2. Line $\mathbf{r_2}$ intersects the other plane at a point $\mathbf{m_2}$ in I_2, which as shown in [6], is defined by $\mathbf{m_2} \simeq \mathbf{r_2} \times \mathbf{U_{12}}^{-T}\mathbf{r_1}$. Similarly, if **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 $\mathbf{s_2} \simeq \mathbf{H_{12}q_1} \times \mathbf{o_2}$ in I_2. The intersection of $\mathbf{s_2}$ with the second plane in I_2 is given by $\mathbf{n_2} \simeq \mathbf{s_2} \times \mathbf{U_{12}}^{-T}\mathbf{s_1}$. Thus, the projections of two lines **R** and **S** that intersect at point **O** have been constructed in I_1 and I_2. Since the intersections of lines **R** and **S** with the two planes are known, the projections of the former in I_3 can also be constructed as $\mathbf{r_3} \simeq (\mathbf{H_{23}p_2}) \times (\mathbf{U_{23}m_2})$ and $\mathbf{s_3} \simeq (\mathbf{H_{23}q_2}) \times (\mathbf{U_{23}n_2})$, where $\mathbf{p_2} \simeq \mathbf{H_{12}p_1}$ and $\mathbf{q_2} \simeq \mathbf{H_{12}q_1}$ are the the points corresponding to $\mathbf{p_1}$ and $\mathbf{q_1}$ in I_2 respectively. Given $\mathbf{r_3}$ and $\mathbf{s_3}$, the projection of point **O** in I_3 is simply $\mathbf{o_3} \simeq \mathbf{r_3} \times \mathbf{s_3}$. Notice that the role of points **P** and **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 constructed as explained above. Such an overconstrained solution is obtained using robust least squares with the LMedS estimator [22] and is tolerant to errors in feature localization and mismatches. Intuitively, the two scene planes form a "reference frame" for the non planar 3D points, since each of the latter is determined from the intersection of at least two constraint lines defined by pairs of points lying on both of the two planes. Thus, knowledge of the two plane transformations in a new view (i.e. the homographies), permits these constraint lines and, therefore, their points of intersection, to be constructed in the new view.

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 $\mathbf{l_1}$ in I_1, its corresponding point in I_2 is given by $\mathbf{p_2} \simeq \mathbf{F_{12}p_1} \times \mathbf{l_2}$, 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 $\mathbf{l_3}$ can be determined by line fitting using the set of transferred points.

At this point, it should be stressed that noise in the images to be matched will cause 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 with 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 image points or lines.

Compared to [7], the main advantage of the method proposed here is that it relies on constraints arising from the scene structure, and therefore are independent of the camera viewpoints. In contrast, the work of [7] is based on constraints related to the relative positions of the cameras. More specifically, given corresponding features in two images along with the fundamental matrices defined among those two images and a third one, image features can be transferred to the third image by intersecting appropriate epipolar lines. However, certain configurations of the three cameras, for example when all three optical centers are collinear, lead to singular cases for which all epipolar lines are parallel

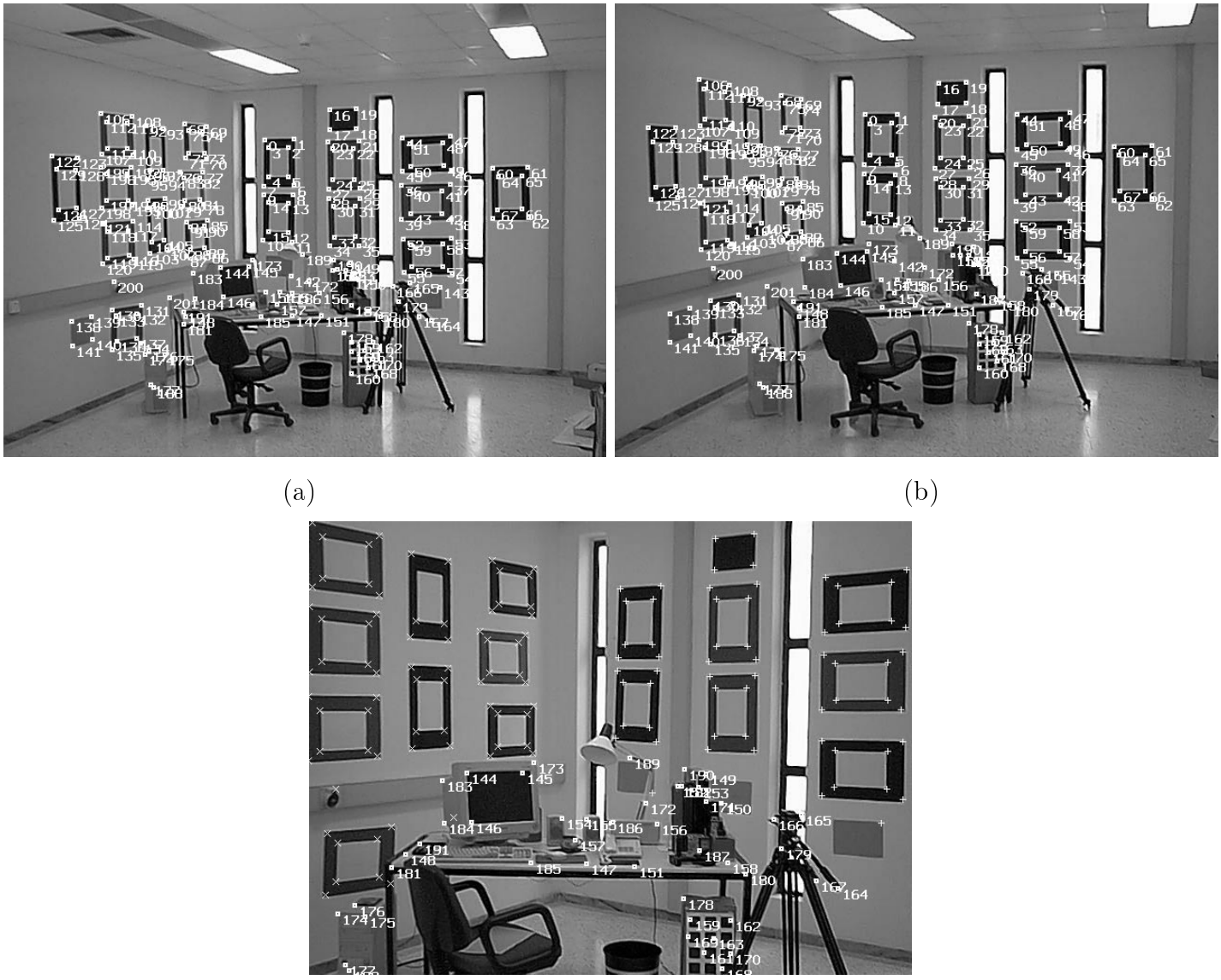
and thus intersect at infinity. It can also be argued that the trifocal tensor (see [1] and references therein) which elegantly encodes the geometry of three views could have been employed for matching the latter in a manner similar to that described in [2]. However, besides the fact that the estimation of the trifocal tensor is not a trivial task, some matches must be known a priori among the three views in order to bootstrap the estimation of the tensor.

4 Experimental Results

The proposed method has been tested with the aid of several stereo pairs. Results from two representative experiments are reported in this section. Throughout all experiments, intra-stereo point and line matches were obtained using [26] and [12] respectively.

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.3(a) and 3(b). Due to space considerations, only one of the images forming the second stereo pair is shown in Fig.3(c). The second stereo pair has been obtained after approaching the viewed scene. Notice that due to the large change in the apparent size of objects among the images of Fig.3(a)-(b) and (c), correlation-based schemes would fail to match them. Corresponding points in the first stereo pair are labeled with identical numbers. The points that have been found to belong to the two most prominent planes of the scene (namely the two walls in the background) are marked with `+` and `x` in Fig.3(c). Fig.3(c) also shows the locations of points that were reprojected to the second stereo pair using the proposed method. To ensure the readability of results, the line segments used by the proposed method are not shown. In order to assess the accuracy of the reprojection, true point matches were determined manually in Fig.3(c) and then the mean Euclidean distance between reprojected and actual matches was measured. The mean reprojection error was found to be 1.59 pixels. For comparison, the mean error obtained from an implementation of [7] was 2.95 pixels. The estimation of the fundamental matrices required by [7] (specifically between the images of Fig. 3(a) - Fig. 3(c) and Fig. 3(b) - Fig. 3(c)) was based on corresponding features located on the two prominent planes only.

Using the point matches determined by the proposed method, the self-calibration technique developed in [13] was employed to estimate the camera intrinsic calibration



(a) (b)

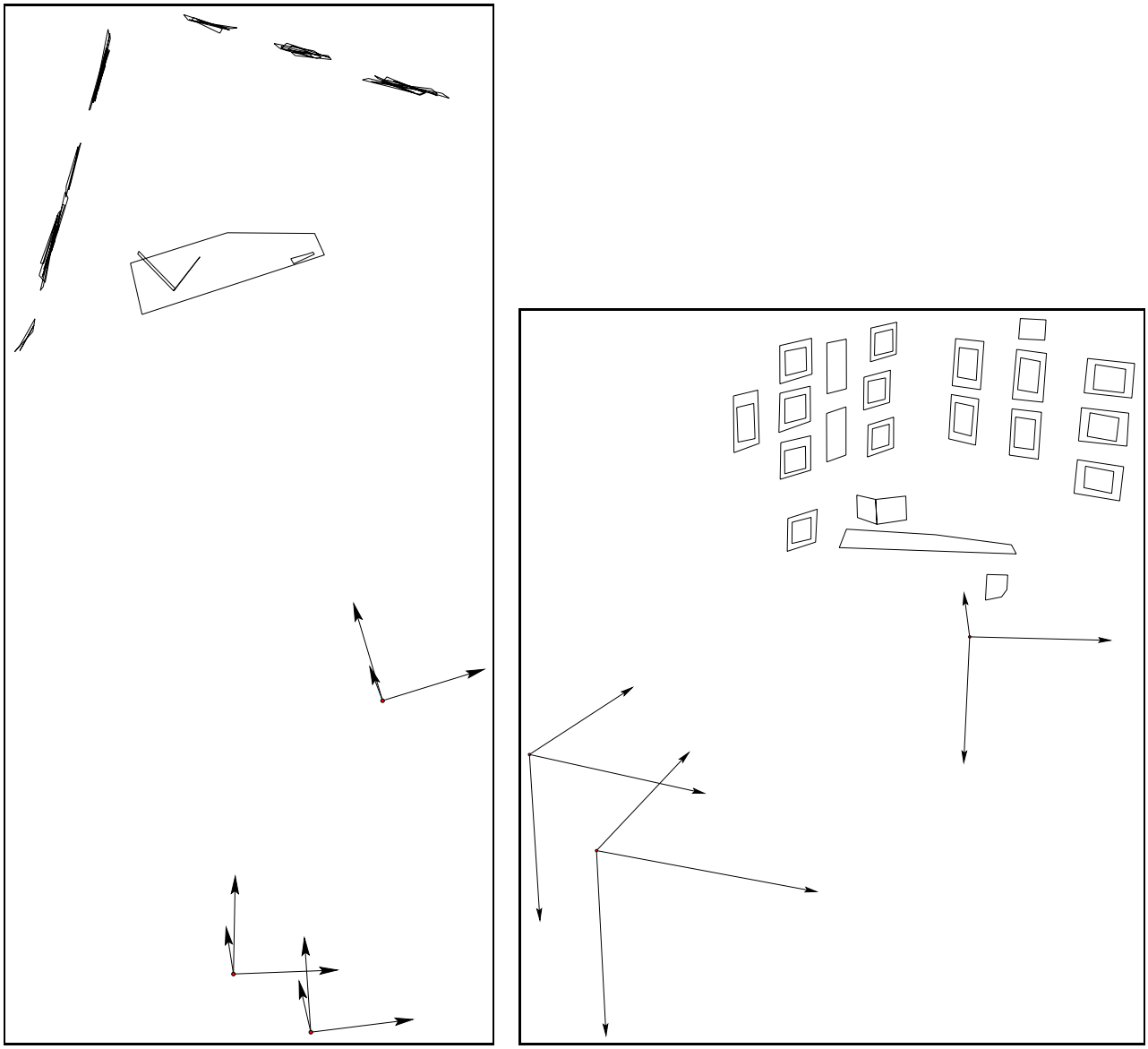
(c)

Figure 3: Indoor scene experiment: (a)-(b) first stereo pair, (c) an image from the second stereo pair.

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 structure up to a scale. Figure 4 shows a top and 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 camera are also shown. Clearly, the structure of the scene has been captured correctly, despite the errors in the reconstructed rectangles on the walls that are mainly due to the small number of images employed. These errors are more evident in the top view of Fig.4(a). The large polygon in front of the two walls corresponds to the top surface of the table, the two polygons above it correspond to the front and left side of the monitor and the polygon below it corresponds to the front side of the box with the calibration pattern appearing in the bottom part of Figs.3.

The second experiment refers to the outdoor images of Fig.5, showing part of FORTH's building in Heraklion. Intra-stereo point matches are again shown with identical labels in Figs.5(a) and (b), while points lying on the two scene planes are marked with the symbols `+` and `x` in the distant view of Fig.5(c). 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). Only the plane of the left wall along with the plane corresponding to the car in the foreground were used instead. This choice of scene planes intends to test the performance of the method in cases of planes having small spatial extents and are defined by a small number of points. Reprojected points in the second stereo pair are drawn enumerated in Fig.5(c). In this experiment, the mean reprojection error was 1.79 pixels, while the same error when the reprojection is made by means of [7] was 6.15 pixels. This discrepancy in the performance of the two schemes is due to the fact that for some of the points not on the two planes, the associated inter-stereo epipolar lines are almost parallel and thus their points of intersection cannot be accurately computed. On the other hand, the proposed method is independent of the relative camera positions, therefore its accuracy in this case is unaffected.

Similarly to the previous experiment, the point matches determined by the proposed method were exploited for self-calibrating the employed camera and then using the cap-



(a) (b)

Figure 4: Indoor scene experiment: Top and side views of the reconstructed wireframe model. Note the errors in the reconstructed wall rectangles in (a), which are mainly caused by the small number of images used.

(a) (b)

(c)

Figure 5: Outdoor scene experiment: (a)-(b) first stereo pair, (c) an image from the second stereo pair.

tured images for 3D reconstruction. A top and a side view of the obtained model are shown in Fig. 6. Notice that the angle between the two background walls has been correctly recovered. The two rectangles closest to the camera locations correspond to the car in the foreground of the images in Fig. 5 and the virtual plane defined by the four poles in front of it. The remaining two small rectangles correspond to the two cars found in the right part of the background in Fig. 5.

5 Conclusions

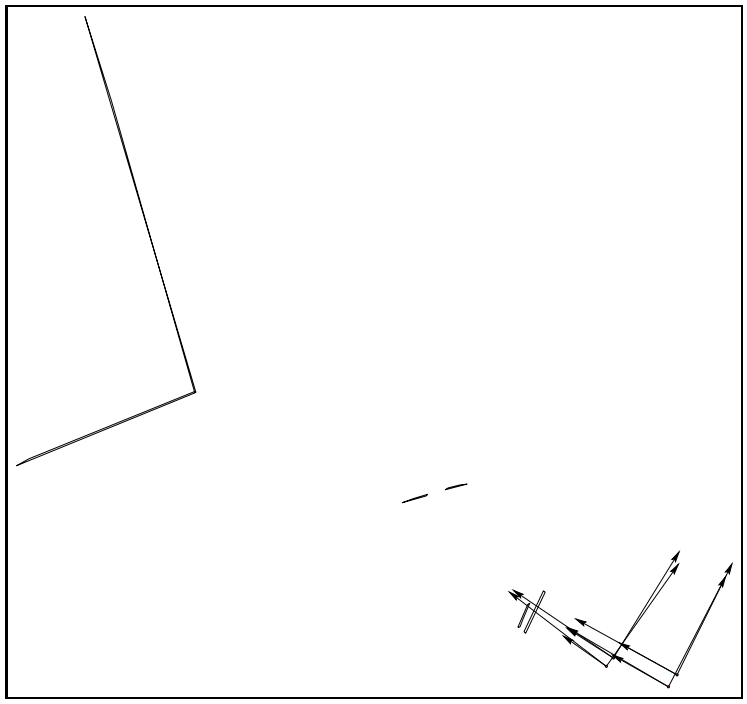
In this paper, a fully automatic method for matching image features between two disparate stereo pairs has been presented. The proposed method has several advantages. First, it exploits geometric constraints arising from the structure of a scene, which are valid regardless of the viewpoints of images and can be computed without any knowledge of camera calibration. Multiple such constraints are defined for each feature, thereby increasing robustness by overdetermining the solution. Second, the method is capable of handling images 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 estimates of the epipoles or the epipolar lines whose accurate computation is known to be difficult in certain cases [18]. Finally, the method handles points and lines in a unified manner by relying on the same principles for deducing their correspondence.

References

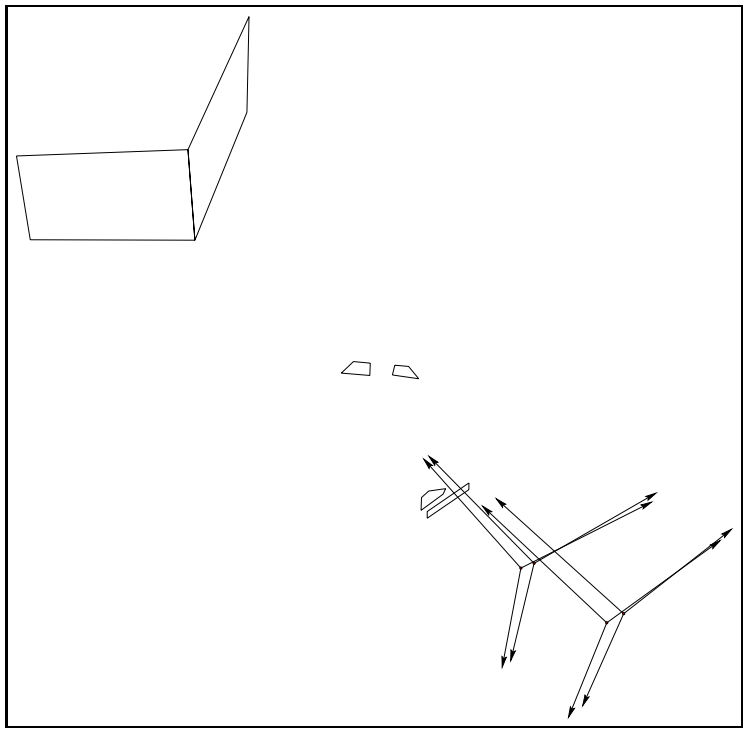
[1] S. Avidan and A. Shashua. Novel View Synthesis in Tensor Space. In Proc. of CVPR'97, pages 1034–1040, Jun. 1997.

[2] P. Beardsley, P.H.S. Torr, and A. Zisserman. 3D Model Acquisition From Extended Image Sequence. In Proc. of ECCV'96, pages 683–695, Cambridge, UK, 1996.

[3] L.G. Brown. A Survey of Image Registration Techniques. ACM Comp. Surveys, 24(4):325–376, 1992.



(a)



(b)

Figure 6: Outdoor scene experiment: Top and side views of the reconstructed wireframe model.

[4] J.B. Burns, R.S. Weiss, and E.M. Riseman. The Non-Existence of General-Case View Invariants. In J.L. Mundy and A. Zisserman, editors, Geometric Invariance in Computer Vision, chapter 6, pages 120–134. MIT Press, Cambridge, MA, 1992.

[5] U.R. Dhond and J.K. Aggarwal. Structure from Stereo: A Review. IEEE Trans. on SMC, 19(6):1489–1510, Nov. 1989.

[6] O. Faugeras. Stratification of 3-D Vision: Projective, Affine, and Metric Representations. J. of the Opt. Soc. of Am. A, 12(3):465–484, Mar. 1995.

[7] O. Faugeras and L. Robert. What Can Two Images Tell Us About a Third One? IJCV, 18(1):5–20, Apr. 1996.

[8] N. Georgis, M. Petrou, and J. Kittler. On the Correspondence Problem for Wide Angular Separation of Non-Coplanar Points. IVC, 16:35–41, 1998.

[9] T.S. Huang and A.N. Netravali. Motion and Structure from Feature Correspondences: A Review. Proc. of the IEEE, 82(2):252–268, Feb. 1994.

[10] S.B. Kang. A Survey of Image-Based Rendering Techniques. In Videometrics VI (SPIE Int'l Symp. on Elec. Imaging: Science and Technology), volume 3641, pages 2–16, Jan. 1999.

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

[12] 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.

[13] 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, Aug. 1999.

[14] M.I.A. Lourakis, S.T. Halkidis, and S.C. Orphanoudakis. Matching Disparate Views of Planar Surfaces Using Projective Invariants. In Proc. BMVC'98, volume 1, pages 94–104, Sep. 1998. Longer version to appear in IVC journal, 2000.

[15] M.I.A. Lourakis and S.C. Orphanoudakis. Visual Detection of Obstacles Assuming a Locally Planar Ground. In Proc. of ACCV'98, LNCS No. 1352, volume 2, pages 527–534, Hong Kong, China, Jan. 1998.

[16] Q.-T. Luong and O. Faugeras. Self-Calibration Of a Moving Camera From Point Correspondences and Fundamental Matrices. IJCV, 22(3):261–289, 1997.

[17] Q.-T. Luong and O.D. Faugeras. Determining the Fundamental Matrix with Planes: Instability and New Algorithms. In Proc. of CVPR'93, pages 489–494, 1993.

[18] Q.-T. Luong and O.D. Faugeras. On The Determination of Epipoles Using Cross-Ratios. CVIU, 71(1):1–18, Jul. 1998.

[19] C. Medioni and R. Nevatia. Matching Images Using Linear Features. IEEE Trans. on PAMI, PAMI-6(6):675–686, 1984.

[20] P. Meer, R. Lenz, and S. Ramakrishna. Efficient Invariant Representations. IJCV, 26(2):137–152, Feb. 1998.

[21] P. Pritchett and A. Zisserman. Wide Baseline Stereo Matching. In Proc. of ICCV'98, pages 754–760, Jan. 1998.

[22] P.J. Rousseeuw. Least Median of Squares Regression. J. of the Am. Stat. Assoc., 79:871–880, 1984.

[23] C. Schmid and A. Zisserman. Automatic Line Matching Across Views. In Proc. of CVPR'97, pages 666–671, 1997.

[24] I. Weiss. Geometric Invariants and Object Recognition. IJCV, 10(3):207–231, 1993.

[25] Z. Zhang. Motion And Structure From Two Perspective Views: From Essential Parameters to Euclidean Motion Via Fundamental Matrix. Journal of the Optical Society of America A, 14(11):2938–2950, 1997.

[26] 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, Oct. 1995.