In Proc. of the International Conf. on Pat. Recogn. (ICPR'00), Vol. 1, Barcelona, Spain, Sep. 3-8, 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
{lourakis,tzurbak,argyros,orphanou}@ics.forth.gr

Institute of Computer Science
FORTH
P.O.Box 1385, 711 10
Heraklion, Crete, Greece

AND

Department of Computer Science
University of Crete
P.O.Box 1470, 714 09
Heraklion, Crete, Greece

Abstract

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, feature-based stereo, object recognition, image registration, camera self-calibration, image-based rendering, 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.

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

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 [4, 10, 3, 1]. 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, 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 [9, 6].

In this paper, 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. 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 [2].

A well-known constraint for a pair of perspective views of a rigid scene, is the *epipolar constraint*. 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 $\mathbf{p}'^T \mathbf{F} \mathbf{p} = 0$, where $\mathbf{F}$ is the fundamental matrix. Another important concept from projective geometry is the *plane homography* $\mathbf{H}$, which relates two uncalibrated views of a plane Π in 3D. 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 $\mathbf{p}' \simeq \mathbf{H}\mathbf{p}$. Matrix $\mathbf{H}$ can be estimated only up to an unknown scale factor, thus it has eight degrees of freedom.

Owing to the non-singularity of $\mathbf{H}$, a homography is easier to estimate compared to a fundamental matrix [8]. 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 [13] and [8] 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.

3.1 Segmenting Planar Surfaces

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 $\mathbf{P} \notin \mathbf{L}$. Clearly, $\mathbf{L}$ is the common intersection of a pencil of 3D planes containing it. As shown in [12], the homographies of this pencils' planes are given by a single parameter equation:

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

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 $([\mathbf{l}']_{\times}\mathbf{F}\mathbf{p}) \cdot \mathbf{q} + \mu(\mathbf{e}'\mathbf{l}^T\mathbf{p}) \cdot \mathbf{q} = 0$. The parameter μ for the plane defined by $\mathbf{L}$ and $\mathbf{P}$ is determined by solving this last equation and then the corresponding homography is given 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. Finally, the homographies of the two most prominent planes are refined using robust regression on the constraints derived from the full sets of features assigned to them [6].

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 [6] 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

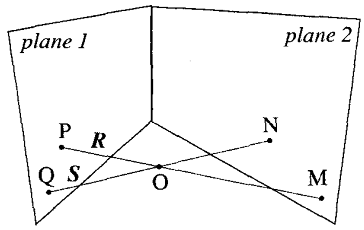


Figure 1. The 3D point O defines the lines R and S with the points P and Q that lie on the first plane. These lines intersect the second plane at points M and N.

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. More details regarding the algorithm can be found in [6].

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 $\mathbf{P}$ has 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_1 and I_2, as described in section 3.2.

Referring to Fig.1, let $\mathbf{P}$ and $\mathbf{O}$ be two 3D points, with $\mathbf{P}$ being on one of the two most prominent planes and $\mathbf{O}$ 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 [2], is defined by $\mathbf{m_2} \simeq \mathbf{r_2} \times \mathbf{U_{12}}^{-T}\mathbf{r_1}$. 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 $\mathbf{s_2} \simeq \mathbf{H_{12}q_1} \times \mathbf{o_2}$ 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 $\mathbf{R}$ and $\mathbf{S}$ that intersect at point $\mathbf{O}$ have been constructed in I_1 and I_2. Since the intersections of lines $\mathbf{R}$ and $\mathbf{S}$ 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 $\mathbf{O}$ 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 constructed as explained above. Such an overconstrained solution is obtained using robust least squares with the LMedS estimator [11] 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, theoretically, it suffices to transfer their endpoints in I_3. For increased accuracy however, more points on a given line $\mathbf{L}$ can be transferred in I_3. Then, the equation of $\mathbf{l_3}$ can be determined by line fitting using the full set of transferred points.

4 Experimental Results

Due to space limitations, results from only one representative experiment are reported in this paper. More results can be found in [7]. Throughout all experiments, intra-stereo point and line matches were obtained using [13] and [5], respectively.

The reported experiment refers to the outdoor images shown in Fig.2. Intra-stereo point matches are shown with identical labels in Figs.2(a) and (b), while points lying on the two scene planes are marked with the symbols + and x in the distant view of Fig.2(c). 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.2(c). In this experiment, the mean reprojection error is 1.79 pixels, while the error of reprojection by means of [3] is 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.

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. 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. Third, it does not rely on estimates of the epipoles or the epipolar lines whose accurate computation is known to be difficult. 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] 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.

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

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

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

[6] M. Lourakis, S. Halkidis, and S. Orphanoudakis. Matching Disparate Views of Planar Surfaces Using Projective Invariants. In *Proc. BMVC'98*, volume 1, pages 94–104, Sep. 1998.

[7] M. Lourakis, S. Tzurbakis, A. Argyros, and S. Orphanoudakis. Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing Planes. Technical Report 268, ICS/FORTH, Feb. 2000. ftp://ftp.ics.forth.gr/tech-reports/2000.

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

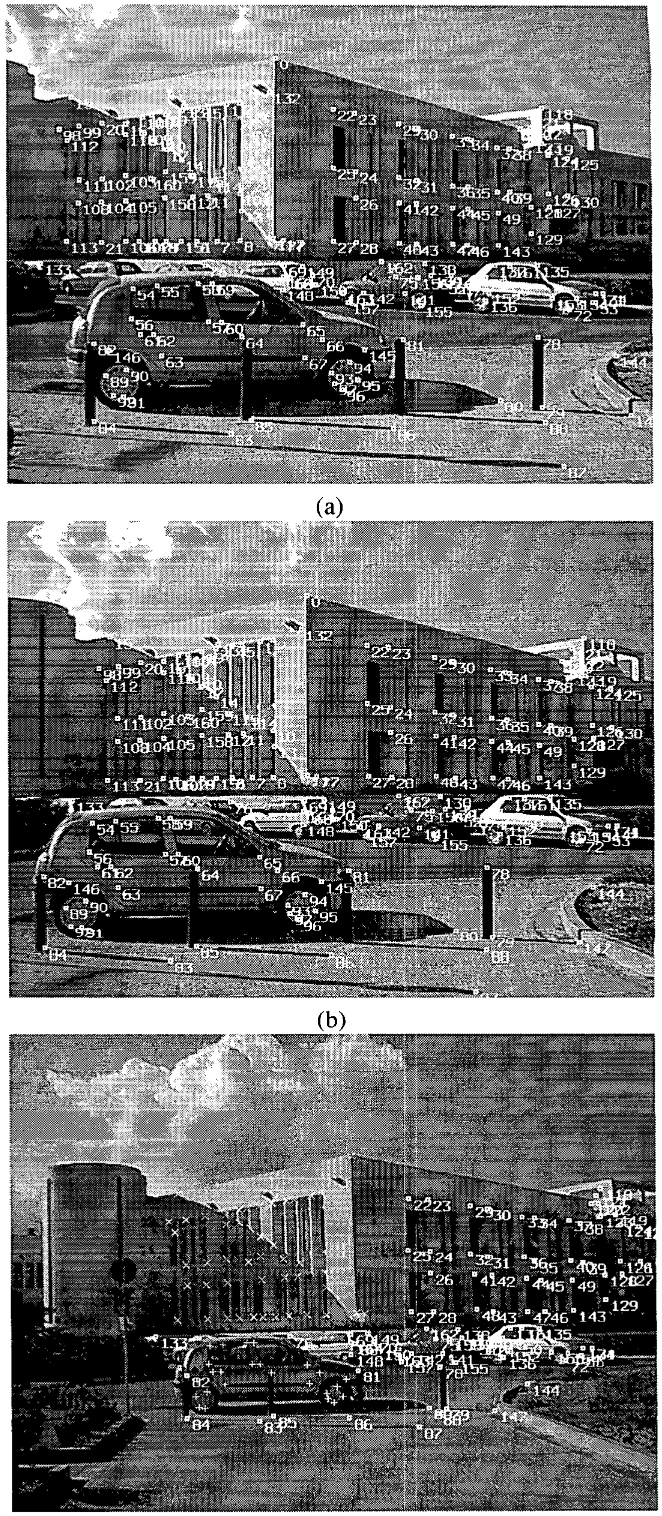
[9] P. Meer, R. Lenz, and S. Ramakrishna. Efficient Invariant Representations. *IJCV*, 26(2):137–152, Feb. 1998.

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

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

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

[13] 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.



(a)

(b)

(c)

Figure 2. Outdoor scene experiment: (a)-(b) first stereo pair, (c) an image from the second stereo pair.