ELSEVIER

Image and Vision Computing 18 (2000) 673–683

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Matching disparate views of planar surfaces using projective invariants

M.I.A. Lourakis[a,b,*], S.T. Halkidis[a,b], S.C. Orphanoudakis[a,b]

[a]*Institute of Computer Science, Foundation for Research and Technology—Hellas, P.O. Box 1385, Heraklion, Crete, Greece*
[b]*Department of Computer Science, University of Crete, P.O. Box 1470, 714 09, Heraklion, Crete, Greece*

Received 9 December 1998; received in revised form 17 July 1999; accepted 25 October 1999

Abstract

Feature matching is a prerequisite to a wide variety of vision tasks. This paper presents a method that addresses the problem of matching two views of coplanar points and lines in a unified manner. The views to be matched are assumed to have been acquired from disparate, i.e. very different viewpoints. By employing a randomized search strategy combined with the *two-line two-point* projective invariant, the proposed method is able to derive small sets of possibly matching points and lines. These candidate matches are then verified by recovering the associated plane homography, which is further used to predict more matches. The resulting scheme is capable of successfully matching features extracted from views that differ considerably, even in the presence of large numbers of outlying features. Experimental results from the application of the method to indoor and aerial images indicate its effectiveness and robustness.

Keywords: Geometric invariance; Correspondence problem; Wide baseline stereo matching; Planar surfaces; Plane homography

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 between two sets of image features extracted from a pair of views of the same scene [1–9]. The correspondence problem, also known as the matching problem, can be defined as that of identifying features in each set having distinct counterparts in the other set. However, despite efforts by numerous researchers, the 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 can have a tremendous impact on the appearance of a scene in different views, thus complicating their matching. The problem is made even harder by the fact that even the most stable feature extractors might fail to detect corresponding features appearing in images. Most approaches in the literature for solving the correspondence problem rely upon metric information. When matching points, for instance, it is often assumed that corresponding points are in nearby locations between images. Therefore, matching is based on the proximity of points within a prespecified search window. A similar assumption when matching lines is that corresponding lines have similar orientations and/or lengths. Metric properties, however, are not preserved under general perspective projection. This implies that any method for determining correspondence based on metric information, only works for images that have been acquired from adjacent viewpoints. Typical approaches that fall in this category can be found in Refs. [8–12].

In order to facilitate the matching of images whose viewpoints differ considerably, 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. [13] require that the projections of four corresponding coplanar points in general position are known and cast the matching problem into an optimization one. Initially, a 'virtual image' resulting by warping the first image towards the second according to the homography of the known plane, is constructed. Then, the correct matches are found with the aid of a Hough-like mechanism, which relies on the observation that the virtual and second image share the same epipole. Schmid and Zisserman [14] assume that either the epipolar geometry of two views or the trifocal geometry of three views is

* Corresponding author.
E-mail address: lourakis@ics.forth.gr (M.I.A. Lourakis).

0262-8856/00/$ - see front matter
PII: S0262-8856(99)00071-2

known and propose a method for matching lines across images which combines the available geometric constraints with graylevel information. Fornland and Schnörr [15] start with a rough initial estimate of the plane homography and propose a two-step method for locating the dominant plane present in a scene by iteratively solving both for the homography and the stereo point correspondence. Pritchett and Zisserman [16] 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, RANSAC [17] is employed to verify consistent matches through the recovery of the epipolar geometry.

The second alternative approach for determining feature correspondence is to exploit quantities that remain unchanged under perspective projection and can be directly computed from the employed image features. Such quantities that are invariant under perspective viewing are referred to as *projective invariants* [18]. Owing to the lack of general-case view invariants [19], in order to employ projective invariants, one has 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 three-dimensional plane in the scene. Planes are common in aerial images as well as images of man-made environments and impose strong geometric constraints regarding the location of corresponding features [20,21]. Meer et al. [22], 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. Next, a voting scheme coupled with combinatorial search enables them to identify corresponding points in two views. Meer's method shares some similarities with the technique developed by Lamdan et al. [23] for recognizing planar objects in cluttered scenes. In Ref. [23], an affine camera model is assumed and a *geometric hashing* scheme that uses transformation invariant reference frames to index shape information into a hash table is employed. Recognition is achieved by means of a voting mechanism which compares a given object to a set of models that are known a priori. Nagao and Grimson [24] also assume an affine camera model and present an algorithm for recovering the best-fit parameters of the transformation which maps plane features between images. Their approach is based on aligning the centroids of corresponding feature groups, which are preserved under affine transformations [20]. Assuming that the apparent motion between an image pair can be approximated by an affine transformation, Gros and co-workers [25] use affine invariant quantities such as length ratios and affine coordinates with respect to point triplets to describe and match line segments. Rothwell et al. [26] describe a model-based recognition system which employs index functions to derive projective invariant representations of planar objects. The index functions are constructed from projective invariants based on algebraic curves and a canonical projective frame.

In this work, we propose a novel method for determining the correspondence of two sets of coplanar points and lines. The method exploits results from projective geometry and is capable of determining correspondence between images that are related by an arbitrary projective transformation. Moreover, it treats point and line features in a unified manner and is robust to the existence of large amounts of outliers, i.e. features that do not have matching counterparts in either of the two feature sets to be matched. 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. Background

In the following, projective (homogeneous) coordinates are employed to represent an image point (p_x, p_y) by the 3×1 column vector $\mathbf{p} = (p_x, p_y, 1)^{\mathrm{T}}$. A line having equation of the form $\mathbf{l}^{\mathrm{T}} \cdot \mathbf{p} = 0$ 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. Regarding notation, the symbol $\simeq$ will be used to denote equality of vectors up to a scale factor. Vectors and arrays will be written in boldface. For more detailed treatments of the application of projective geometry to computer vision, the interested reader is referred to Refs. [20,27–29].

A well-known projective geometry theorem states that four coplanar features, namely a pair of lines $\mathbf{l}_1$, $\mathbf{l}_2$ and a pair of points $\mathbf{p}_1$, $\mathbf{p}_2$, define a quantity that remains unchanged under projective transformations. This quantity is known as the *two-line two-point* ($2\mathscr{L}2\mathscr{P}$) projective invariant and is given by the following equation [20]:

$$2\mathscr{L}2\mathscr{P}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{p}_1, \mathbf{p}_2) = \frac{\mathbf{l}_1 \cdot \mathbf{p}_1}{\mathbf{l}_2 \cdot \mathbf{p}_1} \frac{\mathbf{l}_2 \cdot \mathbf{p}_2}{\mathbf{l}_1 \cdot \mathbf{p}_2}, \tag{1}$$

where $\cdot$ denotes the vector dot product. Noting that $\mathbf{l}_i \cdot \mathbf{p}_j$ is the algebraic distance of point $\mathbf{p}_j$ from the line $\mathbf{l}_i$, the $2\mathscr{L}2\mathscr{P}$ invariant can be interpreted more intuitively as a ratio of distance ratios, i.e. it is an alternative representation of the cross-ratio [18]. This is better understood by noting that the point of intersection $\mathbf{r}$ of $\mathbf{l}_1$ and $\mathbf{l}_2$ determines a line with each of the points $\mathbf{p}_1$ and $\mathbf{p}_2$ (see Fig. 1(a)). Therefore, a pencil of four lines centered at $\mathbf{r}$ is defined, whose cross-ratio is invariant to projective transformations [20].

Using the $2\mathscr{L}2\mathscr{P}$ invariant, it can be shown that points on the plane can be assigned coordinates that remain unchanged under projective transformations. This can be

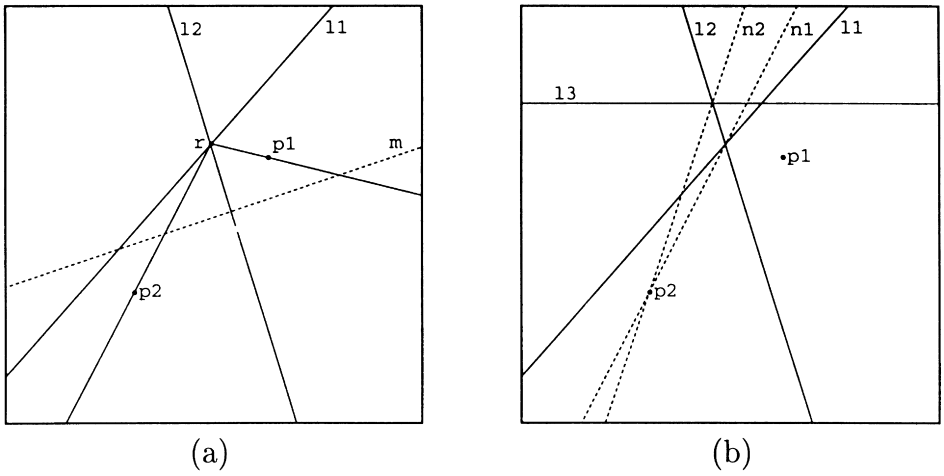

Fig. 1. (a) The pencil centered at **r**, defined by lines $\mathbf{l}_1$ and $\mathbf{l}_2$ and points $\mathbf{p}_1$ and $\mathbf{p}_2$. For any line m intersecting the pencil, the cross-ratio of the pencil is equal to the cross-ratio defined on m by its four points of intersection with the lines of the pencil. (b) The dashed lines, defined by two $2\mathscr{L}2\mathscr{P}$ invariants, intersect at $\mathbf{p}_2$ (see text for explanation).

done as follows. Suppose that $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$ are three lines and $\mathbf{p}_1$, $\mathbf{p}_2$ two points lying on the plane, as shown in Fig. 1(b). Also assume that $2\mathscr{L}2\mathscr{P}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{p}_1, \mathbf{p}_2) = \alpha$. It is straightforward to show that all points $\mathbf{q}$ such that $2\mathscr{L}2\mathscr{P}(\mathbf{l}_1, \mathbf{l}_2, \mathbf{p}_1, \mathbf{q}) = \alpha$ are constrained to lie on a line $\mathbf{n}_1$ through $\mathbf{p}_2$, which is drawn dashed in Fig. 1(b). Similarly, if $2\mathscr{L}2\mathscr{P}(\mathbf{l}_2, \mathbf{l}_3, \mathbf{p}_1, \mathbf{p}_2) = \beta$, all points $\mathbf{q}$ such that $2\mathscr{L}2\mathscr{P}(\mathbf{l}_2, \mathbf{l}_3, \mathbf{p}_1, \mathbf{q}) = \beta$ lie on a line $\mathbf{n}_2$ through $\mathbf{p}_2$. Thus, $\mathbf{p}_2$ is uniquely determined by the intersection of $\mathbf{n}_1$ and $\mathbf{n}_2$. Equivalently, it can be stated that $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$, $\mathbf{p}_1$ form a basis for the projective plane. The coordinates of a point on the plane with respect to this basis are given by a pair of $2\mathscr{L}2\mathscr{P}$ invariants. The fact that four features are required to uniquely determine the location of a point on the projective plane should come as no surprise, since it is known that the cardinality of each basis of the latter is equal to four [20].

Another important concept from projective geometry is the *plane homography* (also known as plane projectivity or plane collineation) **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 **H** with the following properties. Assuming that **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 [20]:

$$\mathbf{p}' \simeq \mathbf{H}\mathbf{p} \tag{2}$$

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

$$\mathbf{l}' \simeq \mathbf{H}^{-\mathrm{T}}\mathbf{l}, \tag{3}$$

where $\mathbf{H}^{-\mathrm{T}}$ denotes the inverse transpose of **H**. Matrix **H** can be estimated only up to an unknown scale factor, thus it has eight degrees of freedom. As can be seen from Eqs. (2) and (3), a single pair of corresponding features provides two constraints regarding **H**, therefore the homography can be recovered using at least four pairs of corresponding coplanar points or lines.

3. Planar feature matching

Suppose that two views of a planar surface are to be matched. Let S_1 and S_2 be the sets of points and lines extracted from the first and second view, respectively. The proposed method employs a randomized search scheme, guided by geometric constraints, to form hypotheses regarding the correspondence of small subsets of S_1 and S_2. 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. In the remainder of this section, the matching algorithm is explained in greater detail.

The matching algorithm starts by randomly selecting a small subset R_1 of S_1, consisting of three lines and $N + 1$ points, where N is an arbitrary positive integer which is further discussed below. It then attempts to match R_1 with a subset of S_2 that contains three lines and $N + 1$ points, as follows. Using the three lines and one of the points in R_1, a basis B_1 for the projective plane is formed. Each of the remaining N points is assigned a pair of $2\mathscr{L}2\mathscr{P}$ invariant values, as explained in Section 2. Following this, all possible subsets B_i of S_2 that contain three lines and one point are examined to determine whether they could match B_1. To achieve this, the bases B_1 and B_i are assumed to be comprised of corresponding features. Then, the two $2\mathscr{L}2\mathscr{P}$ invariants computed in the first view for each of the remaining N points in R_1 are used in the second view to predict the position of their corresponding points. The uniqueness stereo property (i.e. the correspondence between two sets of matching features should be one-to-one) is enforced by taking into account each point in the second

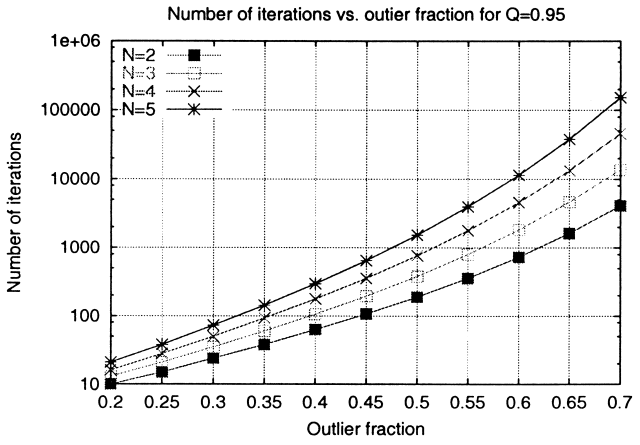

Fig. 2. Plots of the number of iterations m corresponding to different values of N and e. The probability Q is equal to 0.95 in all cases. Note the logarithmic scale in the Y-axis.

view at most once. If all N predicted points in the second view coincide with actual points, then there is evidence that the two bases are indeed corresponding. To verify this assumption, a least squares estimate of the plane homography is computed from R_1 and its corresponding subset in S_2. This estimate along with Eq. (2) is used to bootstrap a process aiming to predict more points in the second view. If a significant fraction of the points in the second view is predicted successfully, then the two bases have been matched and the algorithm stops; otherwise, another basis B_i is chosen. In the case that all possible bases B_i have been considered, a new subset R_1 of S_1 is selected and the process iterates as described above. Upon termination, an estimate of the plane homography $\mathbf{H}$ derived from the matched bases is available. If required, this initial estimate can be refined as follows. Eqs. (2) and (3) are used to predict more points and lines in the second view and a robust estimator, such as the *Least Median of Squares* (*LMedS*) [30], is applied to determine a new estimate of $\mathbf{H}$ that is robust to possible mismatches and errors in the localization of lines and points. More details on the estimation of the homography from matching features can be found in Ref. [31].

Before applying the aforementioned method to real images, a few practical issues need to be resolved. First, noise in the images will cause the location of predicted points in the second view to differ slightly from the location of actual points, even in the case that these points are correct matches. To overcome this problem, we allow for some error by considering a prediction to be correct when its Euclidean distance from the closest actual point is in the order of a few pixels. Since the operation of predicting points in the second view occurs frequently, we speed up the process of locating the actual point closest to a predicted one by precomputing the Voronoi diagram of the points in the second view and employing the slab method to locate the nearest neighbor in planar subdivisions [32]. This technique requires time proportional to $O(\log(m))$ for m points in the second view, a significant improvement compared to the linear time that would be required by a trivial, naive algorithm. Second, since the combination of two $2\mathscr{L}2\mathscr{P}$ projective invariants is not permutation invariant (i.e. the values of the invariants depend on the line pairs used to define them), care should be taken when attempting to predict the locations of the N points in the second view using B_1 and B_i. More specifically, given a labeling of the lines in B_1, all six (i.e. 3!) possible permutations of the labels of lines in B_i should be considered. Finally, in the case that the two views to be matched have very few features in common, the randomized iterative selection of subsets R_1 of S_1 described above might degenerate to an exhaustive, combinatorial search which is very time consuming. In order to avoid this contingency and ensure that the search terminates within reasonable time, a Monte-Carlo type of speedup technique is employed, in which a certain probability of error is tolerated. Assuming that e is the fraction of outliers (i.e. features that do not have a matching counterpart) in the first image, then the probability Q that at least one out of m random samples R_1 of S_1 does not contain any outliers is equal to [33]:

$$Q = 1 - [1 - (1 - e)^{(N+4)}]^m, \tag{4}$$

where $N + 4$ is the cardinality of R_1. Choosing the probability Q around 90–95% and assuming that $e = 60\%$, the solution of Eq. (4) for m gives an upper bound for the number of different sets R_1 that should be tried. Note that Eq. (4) is independent of the cardinality of S_1. For each of the m trials, a basis B_1 is formed from the selected R_1. Basis B_1 is then examined for correspondence with all possible bases B_i of S_2. In Fig. 2, Eq. (4) is solved for m and the solutions are plotted against the outlier fraction e for different values of N, when the probability Q is assumed equal to 0.95.

Having described the matching algorithm, the choice of using more lines than points in the basis sets can now be explained. Real images usually contain less lines than points, therefore the number of different combinations of line sets is smaller than that of point sets of the same size. Thus, given a basis consisting of three lines and one point in the first view, the bases to be considered as candidate matches in the second view are less compared to those that would have to be considered in the case of bases containing more points. Moreover, line segments can be extracted from images more accurately than points, hence calculations involving lines are more tolerant to noise. Lines can also help to reduce the sensitivity to localization errors introduced by the point extractor. Recall from Section 2 that the $2\mathscr{L}2\mathscr{P}$ invariant is defined in terms of algebraic distances of points from lines. Therefore, the localization errors can be made negligible compared to the corresponding algebraic distances by preferring points lying far from the lines defining the invariant.

It could be argued that, instead of attempting to predict

Table 1
Statistics for the number of iterations, the execution time and the number of verifications resulting from various numbers of predicted points for the image pair in Fig. 3 (see Section 4). The extracted features are 30 points and 36 lines for the first image and 30 points and 28 lines for the second image. The average running time is minimized for $N = 3$

Number of predicted points (N)	Number of iterations		Execution time (seconds)		Number of verifications	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
2	14.2	11.3	283.9	236.7	7957	6627.3
3	24.9	19.3	**108.8**	88.3	358.1	290.4
4	45.6	29.6	179.9	117.6	18.7	12.4
5	57.6	49.4	228.5	197.2	1.5	0.52
6	62.3	50.9	257.7	210.4	1.2	0.4
7	99.5	137.7	386.7	539.0	1	0.0

the location in the second view of N points from R_1 and then recovering the plane homography for verifying that the related bases match, the plane homography could have been used from the beginning. In other words, B_1 could in turn be assumed to match with every subset B_i of S_2 and a homography could be estimated using each such assumption. The reason for not doing this is that the computation of the homography requires more time compared to that needed to calculate the $2\mathscr{L}2\mathscr{P}$ invariants and predict the location of N points in the second view. Since the former operation will occur frequently, it is going to have a significant impact on the execution time of the algorithm. Furthermore, estimating the homography from $N + 4$ matching features yields more accurate estimates compared to those obtained from using just the four features included in two possibly matching bases.

Regarding the choice of a proper value for N, it should be observed that, as N increases, the number of times the plane homography has to be estimated is reduced, while the probability that a random sample of $N + 4$ features in S_1 contains at least one outlying feature is increased (see Eq. (4)). In order to find a value for N which achieves a satisfactory compromise between the above two factors in terms of execution time, several experiments have been carried out. These experiments measured the effect of varying N on the execution time and were performed using features extracted from real images. In most cases, it has been found that the minimum execution time was attained for $N = 3$. The results of one such experiment are summarized in Table 1. For each value of N, 10 runs were made and then the mean and standard deviation of the number of iterations, the execution time[1] and the number of times that a candidate match had to be verified through the recovery of the homography were computed. Table 1 clearly shows that as N increases, the probability that the predicted points are indeed correct matches also increases, resulting in a lower number of unsuccessful verifications using the homography. On the other hand, when N increases, the probability that a random sample of $N + 4$ features contains at least one outlier is also increased, resulting in more iterations, i.e. more samples R_1, that have to be tried for the correct matches to be found. Clearly, the minimum average execution time is attained for $N = 3$. Consequently, the maximum number of iterations, obtained by solving Eq. (4) for m, is between 1400 and 1800.

[1] Notice that, since matching employs randomized search, execution time will vary among runs, even if the input remains the same.

4. Experimental results

A set of experiments have been carried out in order to test the performance of a prototype implementation of the proposed method. Throughout all experiments, the most prominent point and line features were obtained automatically. Representative results from four of these experiments are given in this section. Experimental results in MPEG format are also available online at http://www.ics.forth.gr/proj/cvrl/demos/plmatch/.

The first experiment involves the image pair shown in Fig. 3(a) and (b). The images depict a textured poster lying on the floor, imaged from two considerably different viewpoints. The features extracted from these two images are shown in Fig. 3(c) and (d). Despite the difference in viewpoints, it is clear that almost all features that appear in Fig. 3(a) have matching counterparts in Fig. 3(b). The results of the proposed method are shown in Fig. 3(e) and (f), in which matching features are labeled with identical numbers. Running time was 80 s on a 180 MHz R5000 SGI O_2. Fig. 3(g) shows the result of warping the image in Fig. 3(b) according to a robust estimate of the plane homography computed with LMedS. Note that this image has been registered with respect to that in Fig. 3(a), implying that the estimated homography and thus the underlying corresponding features are correct.

The second experiment is based on a pair of aerial images shown in Fig. 4(a) and (b). Although the viewed scene is not exactly planar, it is far from the cameras and thus it can be considered to be approximately planar. Fig. 4(c) and (d) illustrates the extracted features. The output of the proposed method is shown in Fig. 4(e) and (f), in which matching features are labeled with identical numbers. Fig. 4(g) shows the result of warping the image in Fig. 4(b) according

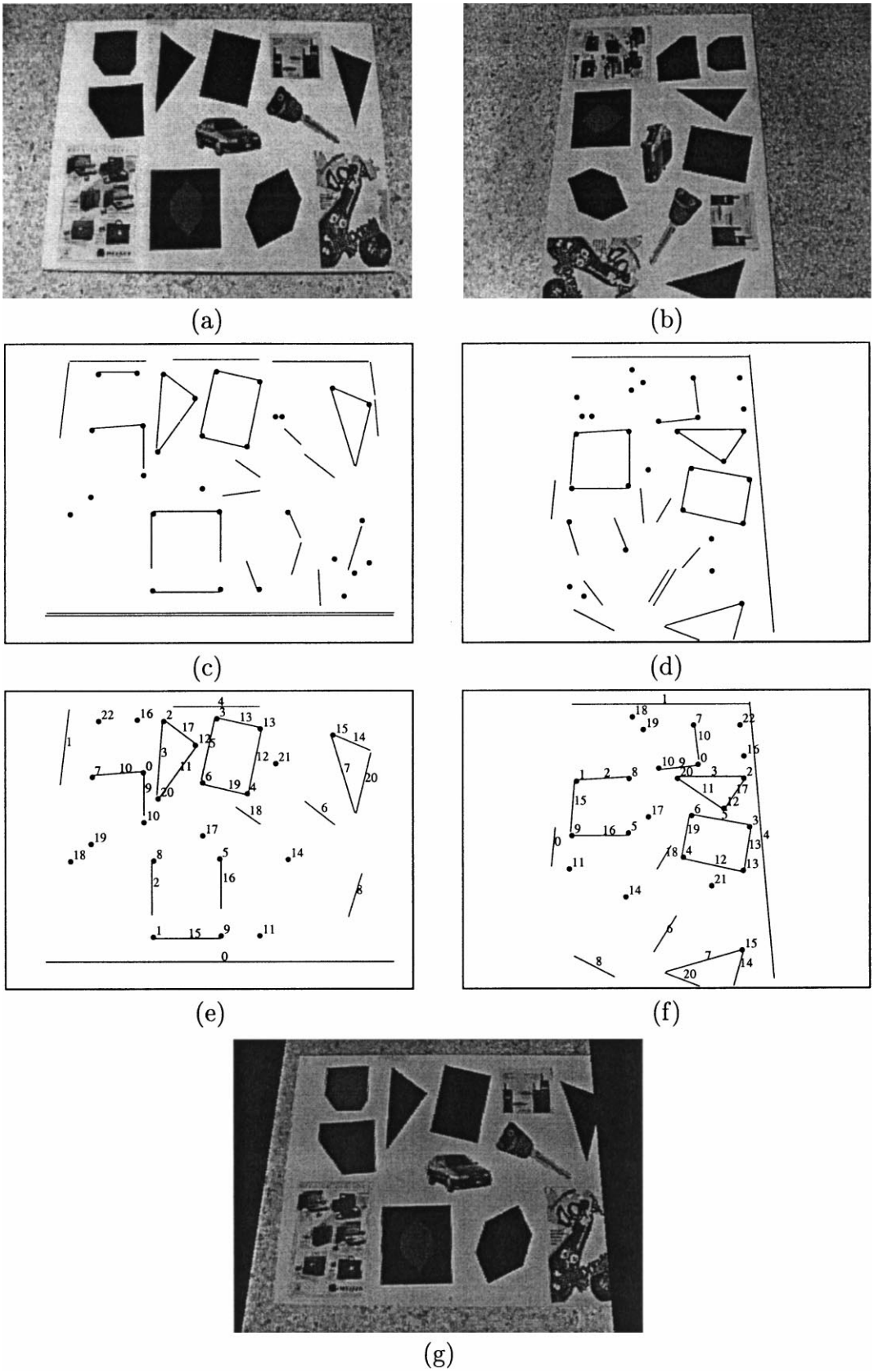

Fig. 3. (a) and (b) Two views of a poster. (c) and (d) The extracted features. (e) and (f) The computed correspondences. (g) Second view warped according to the estimated homography (see text for explanation).

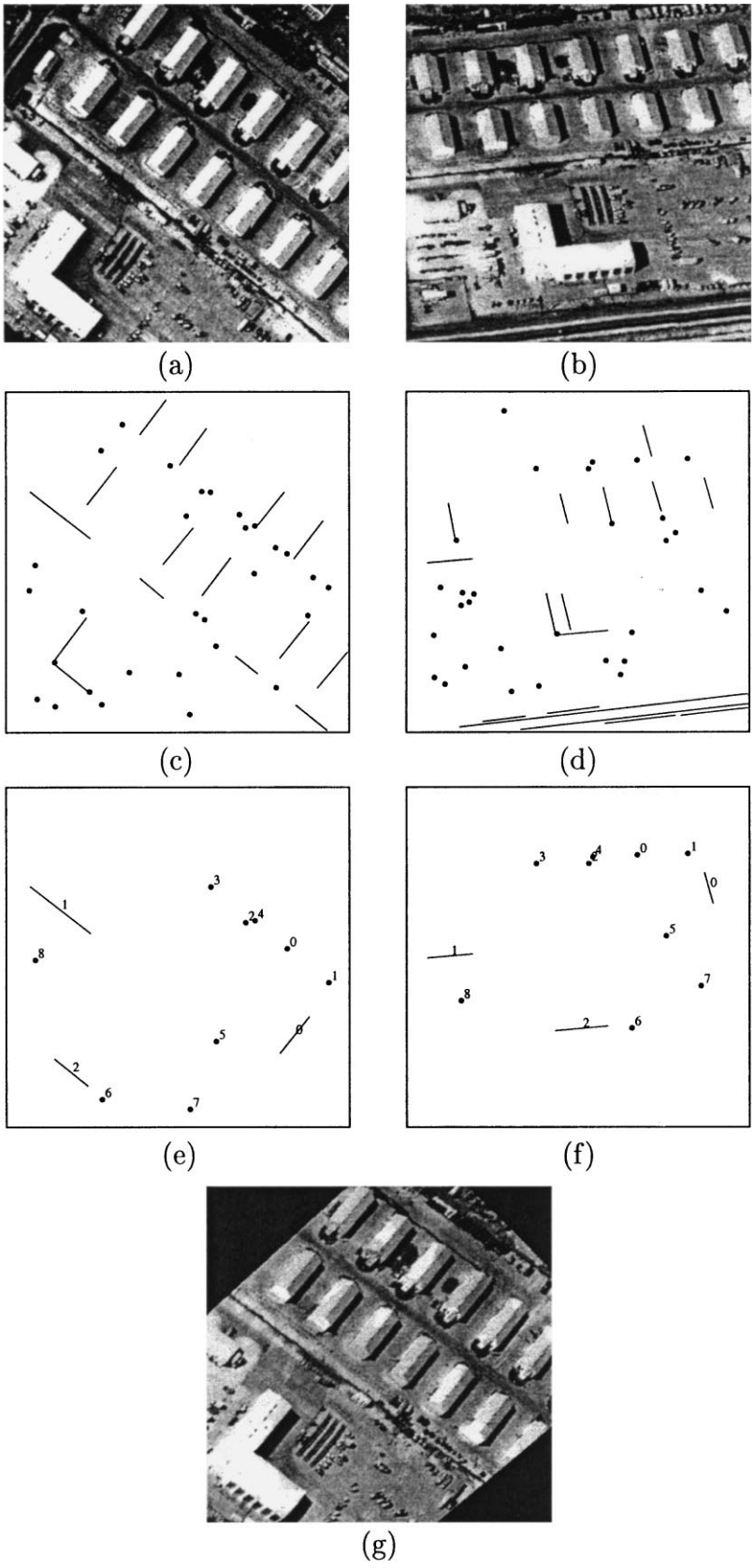

Fig. 4. (a) and (b) Two aerial images (courtesy of P. Meer). (c) and (d) The extracted features. (e) and (f) The computed correspondences. (g) Second image warped according to the estimated homography (see text for explanation).

to the plane homography estimated with LMedS. Again, this image has successfully been registered with that in Fig. 4(a). In this particular experiment, a large number of outliers was tolerated. More specifically, about 65% of the features present in Fig. 4(a) do not appear in Fig. 4(b). This clearly demonstrates the robustness of the proposed method. Because of the large number of outliers, the method required approximately 20 min of execution time.

The third experiment employs the well-known "pentagon" stereo pair, shown in Fig. 5(a) and (b). To make the experiment more challenging, the disparities were increased by rotating the right image 95° counterclockwise. Similar to the previous experiment, the cameras are far from the scene so that the latter can be considered to be planar. Most techniques for image matching that are based on cross-correlation of intensity neighborhoods centered at the detected features would fail for this image pair. This is because, even in the case of corresponding features, image patches differ considerably between images due to rotation. The features extracted from the stereo pair are shown in Fig. 5(c) and (d). Fig. 5(e) and (f) shows the matching features. Execution time was around 7 min. Fig. 5(g) depicts the result of warping the image in Fig. 5(b) using the plane homography estimated with the LMedS estimator.

The last experiment summarizes the results of applying the proposed method on a series of indoor images illustrated in Fig. 6(a)–(i). These images depict a wall of a room and have been captured by a trinocular camera system from three distinct locations in space. The left and right image of each row were in turn matched with the middle one and then the middle images of the first and third row were matched with the middle image of the second row. By employing the estimated homographies, each image was warped towards the middle image of the second row, so as to construct the mosaic [34] shown in Fig. 6(j). Despite some small accumulated errors, this mosaic captures accurately the appearance of the wall from the viewpoint of the image in Fig. 6(e). These errors are due to the fact that the technique used to estimate the homography relies on minimizing the algebraic distance expressed by Eqs. (2) and (3). If greater accuracy is desired, this estimate can serve as the starting point to more elaborate nonlinear methods, that estimate the homography by iteratively minimizing the sum of squared Euclidean distances between points in one view and corresponding transferred points from the other view [35,36].

5. Summary and discussion

In this paper, a method for determining the correspondence of two sets of coplanar features 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 viewpoint and can be computed without any knowledge of camera calibration. Second, it is capable of handling disparate views, despite effects due to illumination changes, occlusions, perspective foreshortening, etc. Therefore, it is applicable in cases where tracking methods assuming small motion between images would fail. Third, it is immune to large numbers of outlying features. Fourth, the algorithm handles points and lines in a unified manner by using the $2\mathscr{L}2\mathscr{P}$ invariant to deduce their correspondence simultaneously. Finally, in order to make a hypothesis regarding the correspondence of two feature sets, the method avoids a direct comparison of the invariants computed in each of the associated views. Instead, the invariants are used to predict the locations of points in the first view in the second one. The correctness of the match is then assessed using intuitively appealing Euclidean distances between predicted points in the second view and actual ones.

Apart from solving the correspondence problem, the proposed method can also detect the case of two sets of planar features that are not matching. This case is reported if, after completing as many iterations as those prescribed by solving Eq. (4) for m, the algorithm cannot encounter a subset R_1 of S_1 that corresponds to some subset of S_2. In addition, the proposed algorithm can also be used for identifying the dominant plane[2] in the case of scenes that are not entirely planar [15,31,37]. Features that belong to the dominant plane will be matched, while the remaining features will be treated as outliers. The most serious drawback of the method is its high cost in terms of execution time, which can be on the order of several minutes when the outlier/inlier ratio is high.

Future research efforts will focus on techniques to improve the speed of the method by taking into account some photometric information regarding the features to be matched. Specifically, referring to the discussion in Section 3, when given a basis B_1 in S_1 that is to be matched with a basis B_i in S_2, a significant reduction in the size of the space to be searched can be made as follows. Instead of taking into account all possible bases B_i formed by features in S_2, only the bases whose component features have photometric descriptions similar to those of the features in B_1 ought to be considered. Such a comparison of photometric description should be made using a loose definition of similarity, so that intensity changes in the vicinity of features do not prevent the algorithm from identifying the correct matches. Various heuristic rules could also be incorporated in the algorithm, so that impossible configurations are efficiently detected and then discarded. For example, parallel lines in the basis feature sets give rise to degenerate configurations and should be avoided. Also, the fact that convexity is preserved by perspective projection [20,22] yields a powerful constraint that can be exploited for reducing the search space size. Finally, it is worth noting that the proposed

[2] Dominant is the plane on which lies the majority of the extracted features.

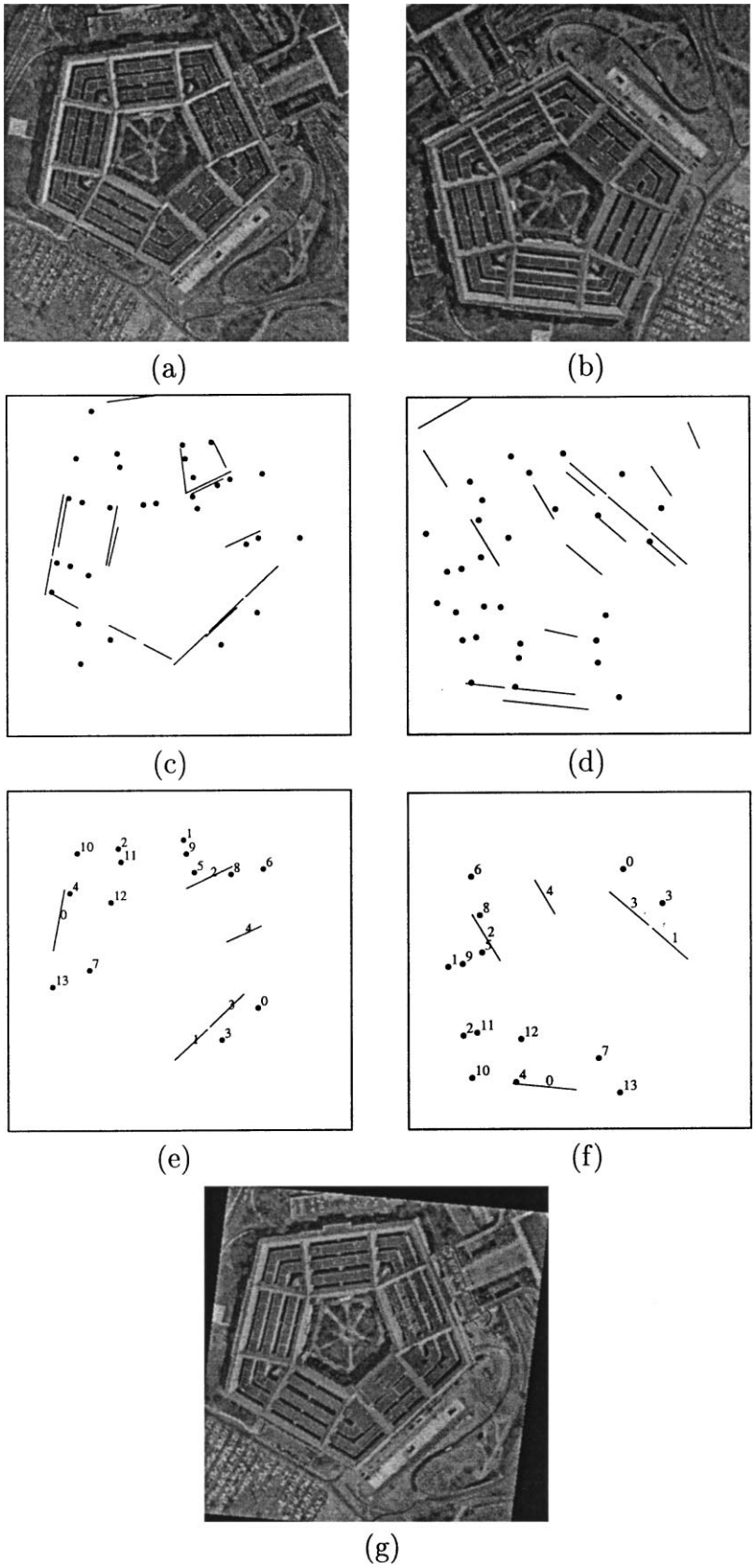

Fig. 5. (a) and (b) The “pentagon” stereo pair. (c) and (d) The extracted features. (e) and (f) The computed correspondences. (g) Second image warped according to the estimated homography (see text for explanation).

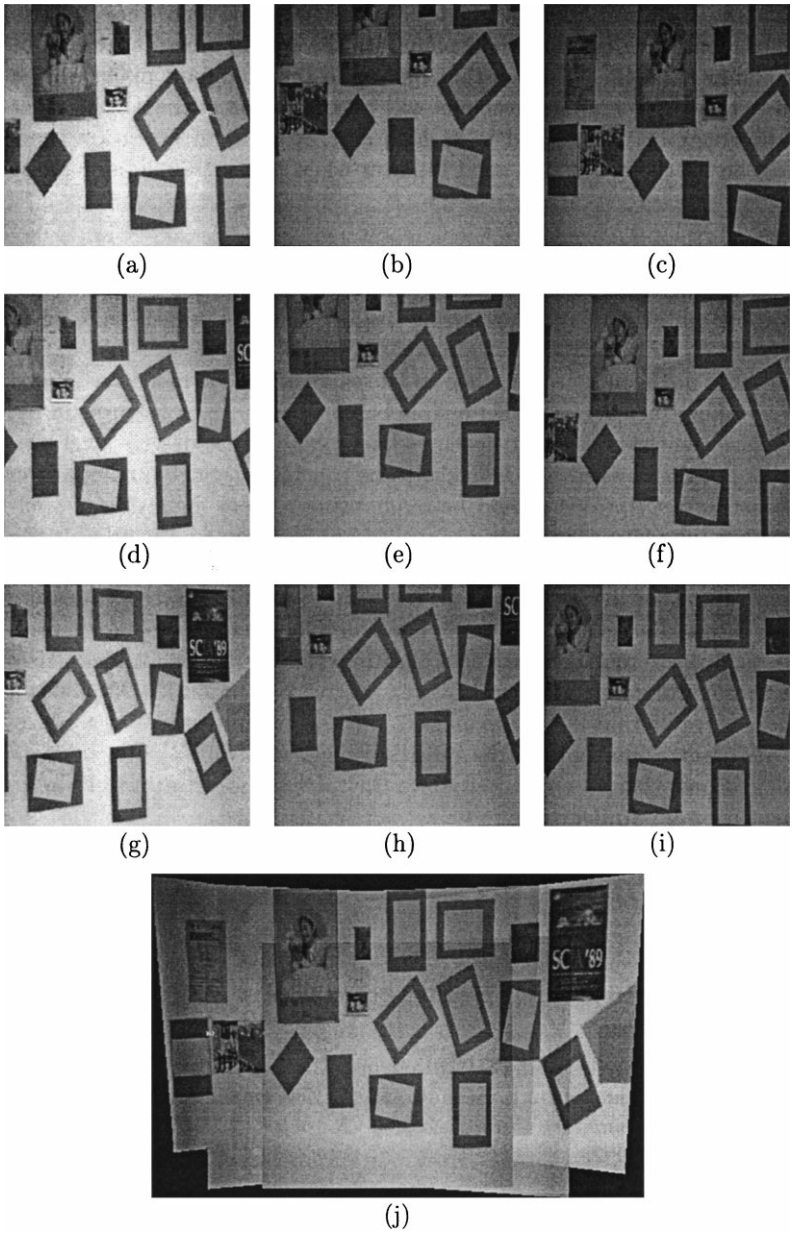

Fig. 6. (a)–(i) A series of images of a wall (courtesy of the INRIA RobotVis group). (j) The mosaic constructed after matching the images and warping them towards image (e). Each of the first three rows depicts the images acquired by a trinocular system from a single location. The brightness differences in patches of the mosaic are caused by the fact that the intensities of the original images have not been normalized prior to warping.

method involves data independent calculations, therefore an alternative way of speeding it up would be to implement it on a parallel computer.

Acknowledgements

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

References

[1] L.G. Brown, A survey of image registration techniques, ACM Comp. Surv. 24 (4) (1992) 325–376.
[2] U.R. Dhond, J.K. Aggarwal, Structure from stereo: a review, IEEE Trans. SMC 19 (6) (1989) 1489–1510.
[3] D.W. Eggert, A. Lorusso, R. Fisher, Estimating 3-D rigid-body transformations: a comparison of four major algorithms, MVA 9 (5/6) (1997) 272–290.
[4] T.S. Huang, A.N. Netravali, Motion and structure from feature correspondences: a review, Proc. IEEE 82 (2) (1994) 252–268.
[5] Q.-T. Luong, O. Faugeras, Self-calibration of a moving camera from

point correspondences and fundamental matrices, IJCV 22 (3) (1997) 261–289.
[6] S. Noronha, R. Nevatia, Detection and description of buildings from multiple aerial images, in: Proceedings of CVPR'97, June 1997, pp. 588–594.
[7] I. Weiss, Geometric invariants and object recognition, IJCV 10 (3) (1993) 207–231.
[8] M.I.A. Lourakis, R. Deriche, Camera self-calibration using the singular value decomposition of the fundamental matrix: from point correspondences to 3D measurements. RR-3748, INRIA Sophia-Antipolis, August 1999. Available at http://www.inria.fr/rapports/sophia/RR-3748.html.
[9] S.B. Kang, A survey of image-based rendering techniques, in: Videometrics VI (SPIE International Symposium on Electronic Imaging: Science and Technology), San Jose, CA, 1999, pp. 2–16.
[10] J.H. McIntosh, K.M. Mutch, Matching straight lines, CVGIP 43 (3) (1988) 386–408.
[11] C. Medioni, R. Nevatia, Matching images using linear features, IEEE Trans. PAMI PAMI-6 (6) (1984) 675–686.
[12] Z. Zhang, A new and efficient iterative approach to image matching, in: Proceedings of ICPR'94, Jerusalem, Israel, 1994, pp. 563–565.
[13] N. Georgis, M. Petrou, J. Kittler, On the correspondence problem for wide angular separation of non-coplanar points, IVC 16 (1998) 35–41.
[14] C. Schmid, A. Zisserman, Automatic line matching across views, in: Proceedings of CVPR'97, 1997, pp. 666–671.
[15] P. Fornland, C. Schnörr, A robust and convergent iterative approach for determining the dominant plane from two views without correspondence and calibration, Proceedings of CVPR'97, Puerto Rico, June 1997 pp. 508–513.
[16] P. Pritchett, A. Zisserman, Wide baseline stereo matching, in: Proceedings of ICCV'98, Bombay, India, January 1998, pp. 754–760.
[17] M.A. Fischler, R.C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–385.
[18] K. Kanatani, Computational cross ratio for computer vision, CVGIP 60 (3) (1994) 371–381.
[19] J.B. Burns, R.S. Weiss, E.M. Riseman, The non-existence of general-case view invariants, in: J.L. Mundy, A. Zisserman (Eds.), Geometric Invariance in Computer Vision, MIT Press, Cambridge, MA, 1992, pp. 120–134 chap. 6.
[20] J.L. Mundy, A. Zisserman, Geometric Invariance in Computer Vision, MIT Press, Cambridge, MA, 1992.
[21] A. Shashua, Projective structure from uncalibrated images: structure-from-motion and recognition, IEEE Trans. PAMI 16 (8) (1994) 778–790.
[22] P. Meer, R. Lenz, S. Ramakrishna, Efficient invariant representations, IJCV 26 (2) (1998) 137–152.
[23] Y. Lamdan, J.T. Schwartz, H.J. Wolfson, Object recognition by affine invariant matching, in: Proceedings of CVPR'88, 1988, pp. 235–344.
[24] J. Nagao, W.E.L. Grimson, Affine matching of planar sets, CVIU 70 (1) (1998) 1–22.
[25] P. Gros, O. Bournez, E. Boyer, Using local planar geometric invariants to match and model images of line segments, CVIU 69 (2) (1998) 135–155.
[26] C.A. Rothwell, A. Zisserman, D.A. Forsyth, J.L. Mundy, Planar object recognition using projective shape representation, IJCV 16 (1) (1995) 57–99.
[27] O. Faugeras, Three-dimensional Computer Vision: a Geometric Viewpoint, MIT Press, Cambridge, MA, 1993.
[28] K.I. Kanatani, Computational projective geometry, CVGIP 54 (3) (1991) 333–348.
[29] R. Mohr, Projective geometry and computer vision, in: C.H. Chen, L.F. Pau, P.S.P. Wang (Eds.), Handbook of Pattern Recognition and Computer Vision, World Scientific, River Edge, NJ, 1993, pp. 369–393 chap. 2.4.
[30] P.J. Rousseeuw, Least median of squares regression, J. Am. Statist. Ass. 79 (1984) 871–880.
[31] M.I.A. Lourakis, S.C. Orphanoudakis, Visual detection of obstacles assuming a locally planar ground, Proceedings of ACCV'98, LNCS No. 1352, Hong Kong, China, January, Vol. 2, 1998, pp. 527–534.
[32] F.P. Preparata, M.I. Shamos, Computational Geometry: an Introduction, Springer, Berlin, 1985 chap. 2.
[33] P. Meer, A. Mintz, A. Rosenfeld, Robust regression methods for computer vision: a review, Int. J. Comp. Vis. 6 (1) (1991) 59–70.
[34] I. Zoghlami, O. Faugeras, R. Deriche, Using geometric corners to build a 2D mosaic from a set of images, in: Proceedings of CVPR'97, Puerto Rico, June 1997, pp. 420–425.
[35] Q.-T. Luong, O.D. Faugeras, Determining the fundamental matrix with planes: instability and new algorithms, in: Proceedings of CVPR'93, 1993, pp. 489–494.
[36] P.H.S. Torr, A. Zisserman, Robust computation and parametrization of multiple view relations, in: Proceedings of ICCV'98, Bombay, India, 1998, pp. 727–732.
[37] D. Sinclair, A. Blake, Quantitative planar region detection, IJCV 18 (1) (1996) 77–91.