Matching Disparate Views of Planar Surfaces using Projective Invariants

Manolis I.A. Lourakis, Spyros T. Halkidis and Stelios C. Orphanoudakis
[lourakis—halkidis—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

Feature matching is a prerequisite to a wide variety of vision tasks. This paper presents a method that addresses the problem of matching disparate views of coplanar points and lines in a unified manner. The proposed method employs a randomized search strategy combined with the *two-line two-point* projective invariant 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.

1 Introduction

A fundamental problem in computer vision, that appears in different forms in tasks such as discrete motion estimation, feature-based stereo, object recognition, image registration, etc, is that of determining the correspondence between two sets of image features extracted from a pair of views of the same scene [5, 1, 2, 17]. 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, etc, can have a tremendous impact on the appearance of a scene in different views, thus complicating their matching.

Most approaches to solving the correspondence problem exploit metric information, such as proximity of points, conservation of line orientation, etc. Metric properties, however, are not preserved under perspective projection. This implies that any method for

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

determining correspondence based on metric information, only works for images that have been taken from adjacent viewpoints. Typical approaches that fall in this category can be found in [18, 8, 4]. An alternative approach is to exploit information that remains unchanged under perspective projection, and thus can be used for matching images whose viewpoints differ considerably. In order to derive quantities that are invariant under perspective viewing, one has to make assumptions regarding the structure of the viewed scene. The most common assumption made in the literature is that the features to be matched lie on a single 3D 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. Meer et al [9], 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. Following this, 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 [6] for recognizing planar objects in cluttered scenes. In [6], 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 against a set of models that are known a priori. Starting with a rough initial estimate of the homography, Fornland and Schnörr [3] propose a two-step method for locating the dominant plane present in a scene by iteratively solving both for the plane homography and the stereo point correspondence. Pritchett and Zisserman [13] rely on feature groups to estimate local homographies, which are then used to compensate for viewpoint differences and generate putative point matches. In the sequel, RANSAC is employed to verify consistent matches through the recovery of the epipolar geometry.

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 Preliminaries

In the following, projective (homogeneous) coordinates are employed to represent image points by 3×1 column vectors $\mathbf{p} = (p_x, p_y, 1)^T$. Lines having equations of the form $\mathbf{l}^T \cdot \mathbf{p} = 0$ are also delineated by projective coordinates using the vectors $\mathbf{l}$. 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.

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* ($2\mathcal{L}2\mathcal{P}$) invariant and is given by the following equation [11]:

$$2\mathcal{L}2\mathcal{P}(\mathbf{l_1},\ \mathbf{l_2},\ \mathbf{p_1},\ \mathbf{p_2}) = \frac{\mathbf{l_1} \cdot \mathbf{p_1}}{\mathbf{l_2} \cdot \mathbf{p_1}} \frac{\mathbf{l_2} \cdot \mathbf{p_2}}{\mathbf{l_1} \cdot \mathbf{p_2}}, \tag{1}$$

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 $2\mathcal{L}2\mathcal{P}$ invariant can be interpreted more intuitively as a ratio of distance ratios, i.e. it is an alternative representation of the cross-ratio.

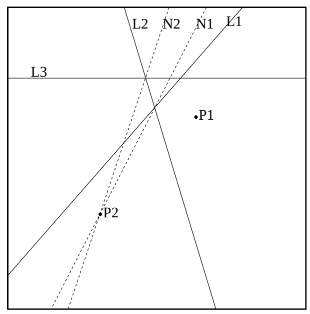

Figure 1: The dashed lines, defined by two $2\mathcal{L}2\mathcal{P}$ invariants, intersect at $\mathbf{p_2}$; see text for explanation.

Using the $2\mathcal{L}2\mathcal{P}$ invariant, we will prove that points on the plane can be assigned coordinates that remain unchanged under projective transformations, 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 Figure 1. Assume also that $2\mathcal{L}2\mathcal{P}(\mathbf{l_1},\ \mathbf{l_2},\ \mathbf{p_1},\ \mathbf{p_2}) = \alpha$. It is straightforward to show that all points $\mathbf{q}$ such that $2\mathcal{L}2\mathcal{P}(\mathbf{l_1},\ \mathbf{l_2},\ \mathbf{p_1},\ \mathbf{q}) = \alpha$ are constrained to lie on a line $\mathbf{n_1}$ through $\mathbf{p_2}$, which is drawn dashed in Fig. 1. Similarly, if $2\mathcal{L}2\mathcal{P}(\mathbf{l_2},\ \mathbf{l_3},\ \mathbf{p_1},\ \mathbf{p_2}) = \beta$, all points $\mathbf{q}$ such that $2\mathcal{L}2\mathcal{P}(\mathbf{l_2},\ \mathbf{l_3},\ \mathbf{p_1},\ \mathbf{q}) = \beta$ 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 $2\mathcal{L}2\mathcal{P}$ invariants.

Another important concept is the *plane homography* (also known as plane projectivity or plane collineation) $\mathbf{H}$, which relates two uncalibrated views of a plane in three dimensions. Each 3D plane Π defines a nonsingular 3×3 matrix $\mathbf{H}$, which relates two views of the same plane. More specifically, if $\mathbf{p}$ is the projection in one view of a point belonging to Π and $\mathbf{p}'$ is the corresponding projection in a second view, then [11]:

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

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

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

where $\mathbf{H}^{-T}$ denotes the inverse transpose of $\mathbf{H}$. $\mathbf{H}$ has eight degrees of freedom, thus it can be estimated only up to an unknown scale factor. As can be seen from Eq. (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 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 geometrical constraints, to form hypotheses regarding the correspondence of small subsets of S_1 and S_2. The validity of such hypotheses is then checked 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 more 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 number. 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 $2\mathcal{L}2\mathcal{P}$ 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 $2\mathcal{L}2\mathcal{P}$ 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 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 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; otherwise, another basis B_i is tried. 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 **H** derived from the matched bases is available. If required, this initial estimate can be refined as follows. Equations (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*) [15], is applied to determine a new estimate of **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 [7].

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 [12]. 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 the naive, trivial algorithm. Second, since the combination of two $2\mathcal{L}2\mathcal{P}$ 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 [10]. 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 [10]:

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

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. B_1 is then examined for correspondence with all possible bases B_i of S_2.

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. Thus, given a basis 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 $2\mathcal{L}2\mathcal{P}$ invariant is defined in terms of algebraic distances of points from lines. Therefore, the localization errors can be made negligible compared to the algebraic distances by preferring points lying far from the lines defining the invariant.

It could be argued that, instead of attempting to predict 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 subject 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 execution time compared to that needed to calculate the $2\mathcal{L}2\mathcal{P}$ invariants and predict the location of N points in the second view. Since this is an operation that occurs frequently, it would have a significant impact on the running 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 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 is estimated is reduced, while the probability that a random sample of

$N + 4$ features from S_1 contains at least one outlying feature is increased (see Eq. (4)). We have found experimentally that, in terms of running time, a satisfactory compromise between the above two factors can be achieved for $N = 3$. Consequently, the maximum number of iterations found by solving Eq. (4) for m is between 1400 and 1800.

The algorithm described above relies heavily on geometric information, which compared to photometric information such as cross-correlation, is more immune to changes in the illumination resulting from changes in the viewpoint. Furthermore, the use of the $2\mathcal{L}2\mathcal{P}$ invariant enabled the development of an algorithm that handles points and lines in a unified manner by deducing their correspondence simultaneously. Another important feature of the algorithm is that, apart from solving the correspondence problem, it 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), the algorithm cannot encounter a subset R_1 of S_1 that corresponds to some subset of S_2. Finally, the proposed algorithm can also be used for identifying the dominant plane[1] in the case of scenes that are not entirely planar [16, 3, 7]. Features that belong to the dominant plane will be matched, while the remaining features will be treated as outliers.

4 Experimental Results

A set of experiments has been conducted 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 three of these experiments are given in this section.

The first experiment refers to the image pair shown in Figures 2(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 Figures 2(c) and (d). Despite the difference in viewpoints, it is clear that almost all features that appear in Fig. 2(a) have matching counterparts in Fig. 2(b). The results of the proposed method are shown in Figures 2(e) and (f), in which matching features are labeled with identical numbers. Running time was 80 seconds on a 180 MHz R5000 SGI O_2. Figure 2(g) shows the result of warping the image in Fig. 2(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. 2(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 Figures 3(a) and (b). Although that the viewed scene is not exactly planar, it is far from the camera and thus it can be considered to be approximately planar. Figures 3(c) and (d) illustrate the extracted features. The output of the proposed method is shown in Figures 3(e) and (f), in which matching features are labeled with identical numbers. Figure 3(g) shows the result of warping the image in Fig. 3(b) according to the plane homography estimated with LMedS. Again, this image has successfully been registered with that in Fig. 3(a). In this particular experiment, a large number of outliers was tolerated. More specifically, about 65% of the features present in Fig. 3(a) do not appear in Fig. 3(b). This clearly demonstrates the robustness of the proposed method. Owing to the large number of outliers, the method required approximately 20 minutes of execution time.

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

The third experiment employs the well-known "pentagon" stereo pair, shown in Figures 4(a) and (b). To make the experiment more challenging, the disparities were increased by rotating the right image 95 degrees counterclockwise. The features extracted from the stereo pair are shown in Figures 4(c) and (d). Figures 4(e) and (f) show the matching features. Running time was around 7 minutes. Figure 4(g) depicts the result of warping the image in Fig. 4(b) using the plane homography estimated with LMedS.

5 Summary and Future Work

In this paper, a method for determining the correspondence of two sets of coplanar features has been presented. This method has several advantages. First, it exploits geometric constraints arising from the structure of a scene, without requiring any calibration information regarding the camera to be known. Second, it is capable of handling disparate views, despite phenomena such as illumination changes, occlusions, perspective foreshortening, etc. Third, it is tolerant to large amounts of outlying features. Fourth, the $2\mathcal{L}2\mathcal{P}$ invariant permits the treatment of the point and line matching problems in a unified manner. 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 the corresponding pair of 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 euclidian distances between predicted points in the second view and actual ones.

Current research efforts focus on techniques to improve the speed of the method by taking into account some photometric information regarding the features to be matched. Namely, referring to the discussion in Section 3, when given a basis B_1 in S_1 that is to be matched with a base B_i in S_2, 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_i ought to be considered. Such a comparison of photometric descriptions 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. Towards this end, steerable filters, as described in [14], might prove useful.

References

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

[2] D.W. Eggert, A. Lorusso, and R. Fisher. Estimating 3-D Rigid-Body Transformations: A Comparison of Four Major Algorithms. *MVA*, 9(5-6):272–290, 1997.

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

[4] R. Horaud and T. Skordas. Stereo Correspondence Through Feature Grouping and Maximal Cliques. *IEEE Trans. on PAMI*, PAMI-11(11):1168–1180, 1989.

[5] 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.

[6] Y. Lamdan, J.T. Schwartz, and H.J. Wolfson. Object Recognition by Affine Invariant Matching. In *Proc. of CVPR'88*, pages 235–344, 1988.

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

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

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

[10] P. Meer, A. Mintz, and A. Rosenfeld. Robust Regression Methods for Computer Vision: A Review. *IJCV*, 6(1):59–70, 1991.

[11] J.L. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, MA, 1992.

[12] F.P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, New York, NY, 1985.

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

[14] R.P.N. Rao and D.H. Ballard. An Active Vision Architecture Based on Iconic Representations. *AI Journal*, 78(1-2):461–505, Oct. 1995.

[15] P.J. Rousseeuw. Least Median of Squares Regression. *Journal of American Statistics Association*, 79:871–880, 1984.

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

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

[18] Z. Zhang. A New and Efficient Iterative Approach to Image Matching. In *Proc. of ICPR'94*, pages 563–565, 1994.

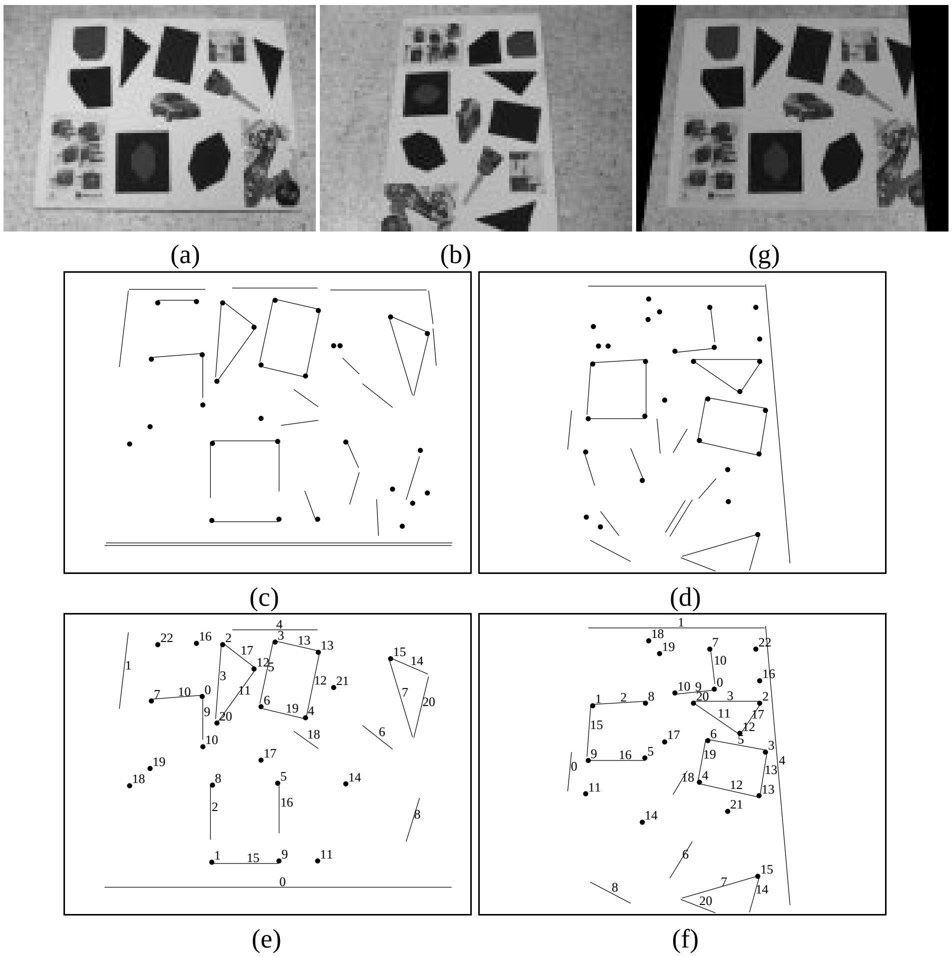

Figure 2: (a, b) two views of a poster, (c, d) the extracted features, (e, f) the computed correspondences and (g) second view warped according to the estimated homography (see text for explanation).

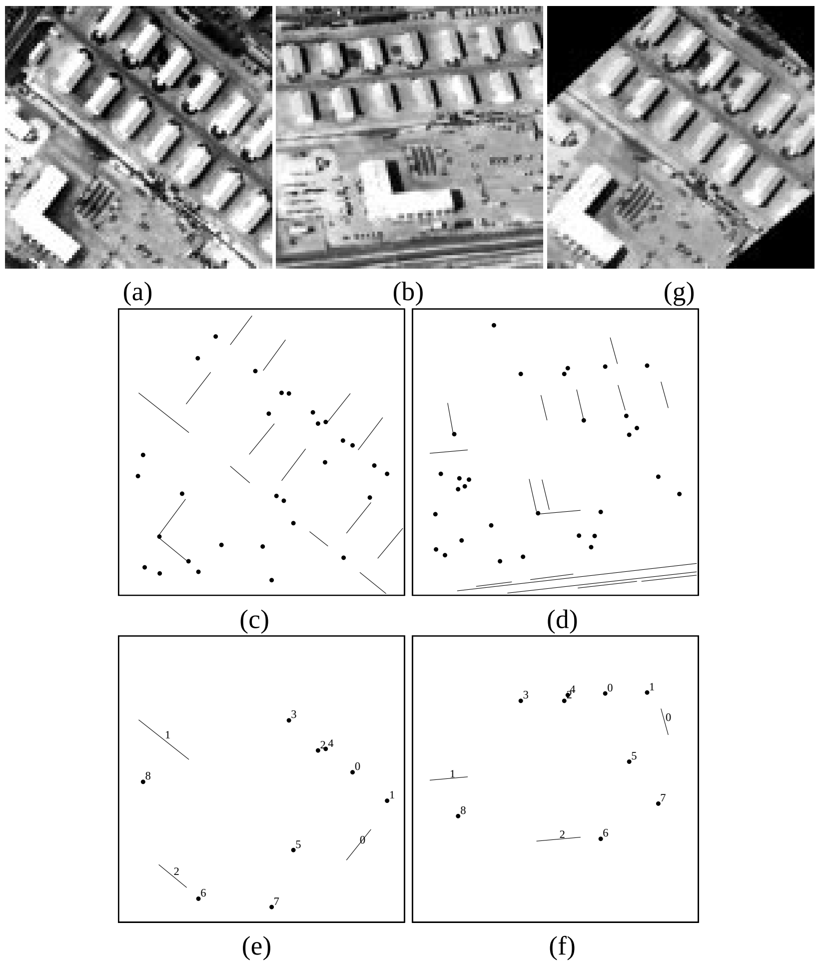

Figure 3: (a, b) two aerial images (courtesy of P. Meer), (c, d) the extracted features, (e, f) the computed correspondences and (g) second image warped according to the estimated homography (see text for explanation).

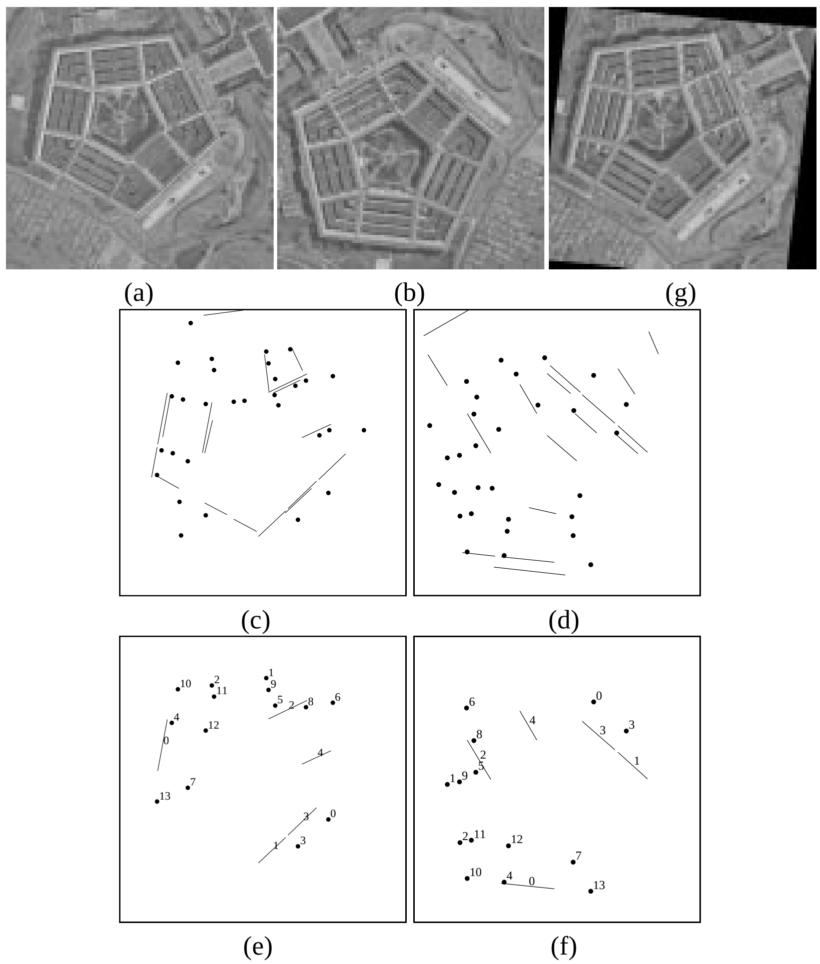

Figure 4: (a, b) the “pentagon” stereo pair, (c, d) the extracted features, (e, f) the computed correspondences and (g) second image warped according to the estimated homography (see text for explanation).