

Detecting Planar Homographies in an Image Pair

submission 335

Abstract

This paper proposes an algorithm that detects the occurrence of planar homographies in an uncalibrated image pair. It then shows that this plane identification method can be used, to great advantage, as a first step in an image analysis process, when point matching between images is unreliable. The detection is performed using a RANSAC-like scheme based on the linear computation of the homography matrix elements using four points. Results are shown on real image pairs where series of matched points belonging to common planes are automatically and correctly identified.

1 Introduction

Planes abound in man made environments. They are very simple structures for which simplified analysis are available. For this reason, they have been used in several computer vision tasks such as camera auto-calibration [14], metric reconstruction [6], visual measurement [2], and obstacle detection [7]. These works demonstrate that when planes are expected to occur in a given scene, it is always advantageous to exploit their presence. The main simplification brought by planes is due to the fact that under perspective projection, the transformation between the world plane and the image plane is defined by a projective linear relation, called a *homography*. The same kind of relation also holds between two perspective views of a plane, that is a homography relates image points of the plane in one image to the images of the same points in the other image.

But in order to benefit from the presence of planes, these special structures must be detected. Most of the techniques cited above assume that the planar structures have been previously identified. In some cases like in [7], some strategies are proposed to detect the ground plane. But none of them propose techniques to detect multiple planar structures of arbitrary configuration.

We describe here an algorithm based on RANSAC. This technique was used for the detection of a global homography corresponding to a pure rotation motion or a single plane in [12] and [13]. In this latter case, the homography was considered as a degenerate configuration of the fundamental matrix.

By itself, RANSAC can require much computations when only very noisy data is available. We will describe strategies that allow the efficient identification of homographies using a set of point matches that was generated automatically from two images, and where the correct matches represent as few as 5 percent of all matches.

An important aspect of the method presented in this paper, is that we propose to proceed to plane identification as a first step in an image analysis process. This idea has also been exploited in [10], but in this case, only local homographies are found in order to produce a viewpoint invariant measure for stereo matching. Detection will be performed on an uncalibrated pair of images and because planes allow a reliable and precise feature-based matching, their early identification can be useful to subsequent tasks, such as the estimation of the epipolar geometry of the cameras (weak calibration), and the 3D structure estimation.

The rest of this paper is organized as follows : Section 2 reviews some basic concepts in projective geometry. Section 3 presents the proposed algorithm for planar homography detection. Section 4 shows the result of an application of our algorithm. Finally, section 5 studies briefly how this algorithm could be used in the context of weak calibration. Section 6 is a conclusion.

2 Projective Relations

The fundamental matrix is a key concept in stereo vision. It mathematically represents the epipolar geometry of a system of two cameras. For each pair of corresponding points, represented by two 3×1 homogeneous vectors, \mathbf{x} and \mathbf{x}' , the well-known epipolar constraint is expressed as:

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0 \quad (1)$$

where \mathbf{F} is the 3×3 fundamental matrix, a singular matrix with 7 DOF.

A homography is a plane projective transformation represented by a nonsingular 3×3 matrix \mathbf{H} . If \mathbf{x} is the image of a world point belonging to a plane and \mathbf{x}' is the image of the same point in a second view, then

there exists a matrix \mathbf{H} such that:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \quad (2)$$

Since homogeneous coordinates are used here, the equality is only up to an unknown scale factor. When several image points of a plane are available, it becomes possible to compute the value of the elements of \mathbf{H} . For each pair of corresponding points, it is possible to extract two independent linear equations from (2), by rewriting it as:

$$\mathbf{x}' \times \mathbf{H}\mathbf{x} = 0 \quad (3)$$

Since the matrix \mathbf{H} has 8 DOF, the element h_{33} can be arbitrarily sets to 1, giving 8 unknowns. Consequently, a total of 4 point correspondences is required to determine \mathbf{H} .

Obviously, in a context of non-perfect data, many more points should be used. Then, \mathbf{H} would be estimated by a minimization scheme. This is usually done by defining the 9×1 vector \mathbf{h} made of the 9 unknown elements of matrix \mathbf{H} , i.e.:

$$\mathbf{h} = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^T \quad (4)$$

With N point correspondences, it is possible to extract $2N$ linear constraints from (3). This results in a system of the form:

$$\mathbf{B}\mathbf{h} = 0 \quad (5)$$

We then have to solve the following problem

$$\min_{\mathbf{h}} \|\mathbf{B}\mathbf{h}\|^2 \text{ subject to } \|\mathbf{h}\| = 1 \quad (6)$$

The solution being the eigenvector of matrix $\mathbf{B}^T\mathbf{B}$ that corresponds to the smallest eigenvalue. To obtain a more stable linear system, the coordinates of the point correspondence are normalized, as explained in [4], where a similar method is used to compute the fundamental matrix.

The minimization scheme presented above has the disadvantage of only minimizing algebraic quantities rather than meaningful geometric quantities that represent the error in the approximation. It might be preferable to have a method that minimizes the summed Euclidean distances between the measured and mapped points. This could be achieved using a non-linear numerical minimization scheme. Such a scheme would still require an initial solution which would be provided by the method described above [2].

3 The Homography Detection Algorithm

In order to detect planar homographies in an uncalibrated pair of images, an approach based on a RANSAC-like scheme will be used. The idea behind the algorithm proposed here is similar to the one presented in [1], which uses RANSAC to estimate the fundamental matrix (see [8] for more on \mathbf{F} estimation). There is however an important difference : in the case of \mathbf{F} , all correct matches will obey the linear constraint (1), while in the case of a homography, only image points of a common plane will satisfy the constraint (3). A different homography \mathbf{H} must therefore be detected for each visible plane.

3.1 The basic RANSAC Scheme

1. First corners are detected in both images. To this end, we use the corner detector proposed in [5].
2. A variance normalized correlation is applied and all pairs with a sufficiently high correlation score are accepted. These pairs form the set of candidate matches. All pairs are considered, this means that a given corner may have more than one candidate match in the other image. A subset of this set will also be considered, made of the matched pairs for which the correlation score is higher than a very conservative threshold. It is expected that this reduced set of matched points will contain less incorrect matches, that is why this latter set will be used to seed the detection process.
3. Four points are selected from the reduced set of candidate matches. Using equation (3), their corresponding homography is determined.
4. Considering now, the larger set of candidate matches, all pairs that agree with the homography found in step 3 are selected. A pair $(\mathbf{x}, \mathbf{x}')$, will be considered to agree with a homography, \mathbf{H} , if it is such that

$$\text{dist}(\mathbf{H}\mathbf{x}, \mathbf{x}') < \epsilon \quad (7)$$
 for some threshold ϵ . If a sufficient number of pairs are consistent with the homography, we can assume that we have identified a valid homography representing the correspondence between the images of a plane present in the scene.
5. Using all the consistent correspondences found above, the homography is now recomputed by solving the constrained minimization problem (6). This will allow a refinement of our estimate of this homography.

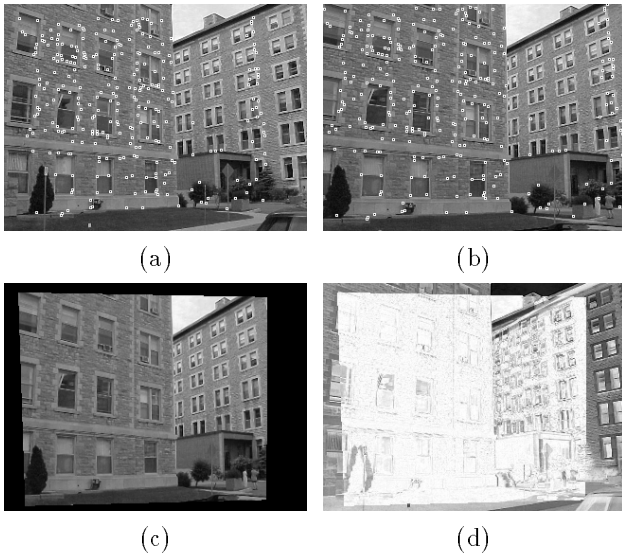


Figure 1: An homography found using our RANSAC-like algorithm. (a) and (b) The pair of images with the point matches that agree with this homography. (c) The result of applying the homography to the right image. (d) The difference between the left image and the image (c).

Figure 1 is the result of the application of this algorithm to an image pair. The algorithm worked from a set of 12619 candidate matches, 210 of them having a correlation score above the conservative threshold (0.95). The threshold on the number of matches that must agree with a valid homography was set at 250 matches, and a match was considered to agree with a homography if $\text{dist}(\mathbf{H}\mathbf{x}, \mathbf{x}) < 2$. For this image pair, the algorithm required on average 30000 iterations and detected a homography with which 270 matches agreed.

The RANSAC-like scheme described above, allows the detection of planar homographies in a pair of images. However, it is important to note that the set of matches obtained after corner detection and matching will usually contain many false matches. This is especially true for architectural images where features, such as windows, are repeated all over the planar structures we want to detect. The scenes being analyzed usually contain several planes so proportionally very few candidate matches will be good matches lying on the same plane and thus agreeing with the same homography. The basic algorithm, described above, may therefore require many iterations, and needs to be improved in order to perform efficiently. We will now describe some ways to improve the efficiency of

this algorithm.

3.2 Selecting the Four Points

The way in which the four points are selected, can greatly influence the likelihood that they will determine a valid homography. Up to now, the points were selected using a brute force approach where they were taken at random from the set of candidate matches.

A way to screen out degenerate configurations is to look at the area between the four selected points. The areas of the four triangles determined by the points are computed and the points will be considered for the next step only if all the areas are higher than a given threshold (we used 300 pixels²). This eliminates the possibility of the degenerate case where three of the points are colinear. It also avoids situations where the points lie too close to one another, and thus, where the error in the positions of the points become significant in comparison to their relative positions.

Using this area threshold, as described above, we have been able to eliminate around 50% of the selected four point configurations in the image pair shown in Figure 1. These configurations were unlikely to produce a valid homography and not considering them eliminated useless computations.

Another possible strategy would be to select points that lie on the same edges in the image. Points lying on the same edge are very likely to belong to the same plane. However, points lying on the same edge are more likely to be colinear or to lie close to one another, then the method does not perform well. An alternative solution consists in choosing four points on two different edges. This increases the likelihood of the points lying on the same plane while generating less degenerate configurations.

We have also considered selecting points lying on the same straight line. However, we found experimentally that these edge-based strategies often give poor results. They can be advantageous on particular image pairs, but do not generally improve the efficiency of the algorithm. They often require more iterations than the brute force approach. A case where selecting points that lie on two lines becomes advantageous, is when the disparity between the two views becomes very significant. An example of such a situation is given in Figure 4. For this image pair, the point selection heuristic gave, on average, a 37% improvement in the performance of the algorithm over a completely random selection.

3.3 Using the Determinant of the Homography Matrix

Another element that can be considered in order to increase the efficiency of the algorithm is the deter-

minant of the homography matrix. If we can expect the determinant to be within a certain range, then it becomes possible to eliminate several preliminary homographies based on this simple criteria.

If the determinant of a homography is close to zero, it corresponds to a degenerated case and should not be further considered. If the determinant of a matrix is very large, the determinant of its inverse (which corresponds to the inverse homography) will be close to zero, this is again a degenerate case. Consequently, we have chosen to eliminate all homographies with determinants having an absolute value outside of a range $[1/n, n]$ for some value n .

Experimentally, we have found that most homographies we detected had a determinant relatively close to one. This can be justified by observing that all the detected homographies in several image pairs we considered were almost affine transformations. This is not surprising since an affine transformation can describe any combination of translations, rotations, shearings and scalings, and in the case where the difference in perspective between the two images of plane is small, such a combination of transformation can be a good approximation of the true transformation. This is especially true for stereo pairs. For affine transformation, the determinant corresponds to the ratio of the areas of the projections of the plane on the two image planes. We should therefore require that this ratio be close to one, if it is not the case, then even matching becomes problematic.

By choosing $n = 10$ we were able to discard around 51% of the preliminary homographies found from four points selected randomly in the image pair shown in Figure 1, but satisfying the minimum area requirement previously described. The determinant of the dominant homography in that image pair was found to be approximatively 1.16.

3.4 Relaxing the Distance Threshold

During the homography estimation process, we experimentally found that several of the homographies rejected by the algorithm as it iterates, are relatively close to valid homographies, although not enough matches agree with them. A strategy that could reduce the number of iterations needed to find a homography using our algorithm, would be to devise a way to get a good approximation of a homography from a very crude one. Then the threshold on the number of matches that must agree with a homography for it to be considered valid could deliberately be set low, and the crude homographies found with this soft threshold could then be refined.

We start by using a fairly generous distance thresh-

old (the value of ϵ in equation (7)), such as 3 pixels, with a relatively low threshold for the number of points needed to be compatible with the homography. This allows us to find a homography relatively fast. However, because of the low threshold on the number of points, the homography will only be somewhat close to a valid homography. Now to find the real homography, we iteratively recompute the current homography using all matches compatible with it at that stage.

This strategy works well when the distance threshold is relaxed (we increase it to up to 50 pixels). Since the computed homography is often close to a valid one, most matches compatible with the actual homography should now, because of the very generous distance threshold, be considered compatible with the found homography. Since we can expect that among the putative compatible matches, a majority will be compatible with the real homography, the computed homography should gravitate towards the valid one, as we iteratively recompute it. Once the homography has stabilized near the true homography, the distance threshold can be gradually tighten, down to about one pixel, so as to only include the best matches in its computation.

We found that this technique can decrease by a factor of hundreds the number of iterations in the RANSAC-like algorithm needed to detect an homography. It does so by allowing us to accept crude approximations of true homographies, and improving them to obtain valid ones.

Figure 2 gives an example for which the initial approximation to the homography was found after only 48 iterations of the RANSAC-like algorithm, and 140 matches agreed with it, using a distance threshold of 3. This low number of iterations was typical for the algorithm on this image pair, when only a crude approximation of a homography was sought. After relaxing the distance threshold to 50 and tightening it back to 1, the refined homography now agreed with 374 points. The homography found in this way required less iterations and described more accurately the plane present in the image than the one found in Figure 1.

3.5 Finding Many Homographies

An image pair will often contain more than one plane. When one plane has been identified, we can remove from the set of candidate matches, the matches compatible with the found homography. We also want to remove from the set of candidate matches, the incorrect matches that have at least one of their points lying in a part of an image where the detected plane is visible.

In order to detect which matches, in the candidate

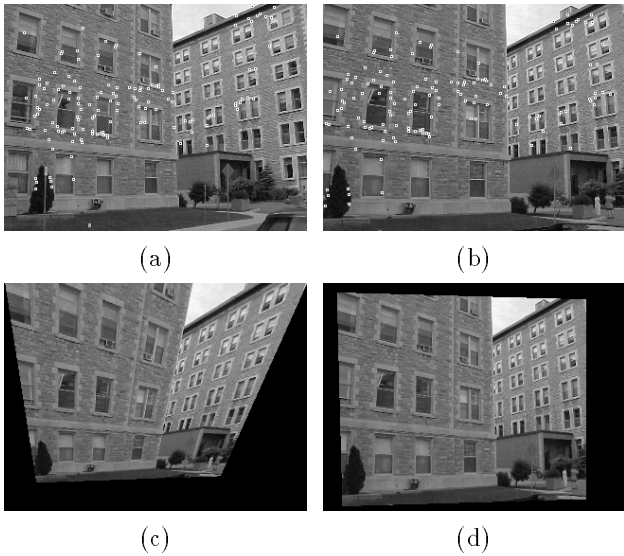


Figure 2: Relaxing the distance threshold. (a) and (b) The pair of images with the point matches that agree with the initial coarse approximation of the homography. (c) The result of applying the coarse approximation of the homography to the right image. (d) The result of applying the final approximation of the homography to the right image.

match set have one of their points lying on the detected plane, we must determine which region of an image agrees with the homography. To do this, we consider the image that is obtained after applying the homography to one of the images in the pair. When this image is compared to the other image in the pair, the region that contains the plane represented by the homography should agree between the two images, while the other structures in the images should not correspond.

To find this region of agreement between the images, the absolute difference of the two images is computed, and thresholded. The thus found binary image is then cleaned up, using some morphological operators. We apply a closing using a small mask to eliminate the isolated pixels in the region corresponding to the plane that are not considered as part of the plane. Then, we use an opening with a larger mask, to eliminate the isolated areas falsely identified as being part of the plane. This procedure will incorrectly identify, as agreeing with the homography, any featureless region of an image. This poses no problems since such a region of the image contains no corners and could therefore be ignored.

Figure 3 illustrates the process of finding a second

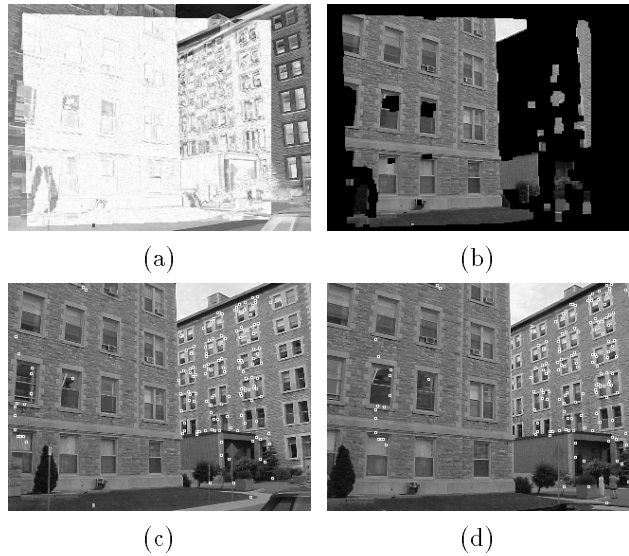


Figure 3: Finding a second homography (a) The difference between the left image and Figure 2 (d). (b) The regions in Figure 2 (c) that were found to agree with the homography. (c) and (d) The pair of images with the point matches that agree with a second homography found after eliminating matches in the regions of (b) from the candidate matches set.

homography in the image pair. In (b), we see the regions of the left image that were found to agree with a first homography. All the candidate matches that lied in that section of the left image were eliminated before our algorithm was used to detect the second homography. The set of candidate corners was thus reduced from 12619 matches to 7270, while its subset of matches with a very high correlation score was reduced from 210 matches to 117. Finding the second homography required 6079 iterations. 102 matches were consistent with this homography, using a distance threshold of 1.

4 Additional Experimental Results

In this section, we show the application of the planar homography detection to two image pairs. The first image contains two views of three planar surfaces which are all detected by the algorithm. This is a case where reliable matching is particularly difficult to obtain because of the repeated elements present all over the plane structures.

Figure 6 shows the feature points belonging to the three homographies that were detected. Notice that the image contained no other features than the three planar surfaces, and that among the 33655 candidate matches, only 520 agreed with a homography with a



Figure 4: An homography found using our RANSAC-like algorithm. (a) and (b) The pair of images with the point matches that agree with this homography.

distance threshold of 1. Thus less than 1.6% of the candidate matches can be considered as exact. Despite this, the three homographies could be identified relatively easily.

Figure 4 shows a homography that was identified in an image pair with significant disparities, and where the difference in point of views is such that the perspective differences between the two images is significant. The four points were selected so that two pairs of points lied on two different lines. The lines where extracted and matched automatically. A set of 14743 candidate matches was used. The algorithm looked for a homography agreeing with 100 matches with a distance threshold of 3. An homography agreeing with 133 matches was found in 7638 iterations, 1337 of which satisfied the minimum area and determinant requirements. After relaxation and tightening, 74 matches agreed with the homography with a distance threshold of 1.

5 Estimation of the fundamental matrix

The result of the proposed algorithm is an estimation of the homography matrices corresponding to planes visible in a pair of images, together with a set of point correspondences located on these identified planes. All this information can then be used in the tasks to follow, within a given vision process. We propose, here, to demonstrate why these results could be used for camera weak calibration (the estimation of \mathbf{F}). In particular we will show that in a case where the set of matches used is very noisy, it would be very advantageous to use homography detection as a first step towards weak calibration.

The fundamental matrix can be estimated using at least two homographies. The estimation can be performed using the matches compatible with these ho-

mographies, or directly from the homography matrices [9]. In this latter case, the following relation between \mathbf{F} and \mathbf{H} is used:

$$\mathbf{H}^T \mathbf{F} + \mathbf{F}^T \mathbf{H} = 0 \quad (8)$$

If in some case, it was much easier to find two such homographies than to perform calibration directly from the set of matches, then it would be advantageous to first find the homographies in order to perform the calibration. Even if the calibration obtained using the two homographies was less accurate than the one that could be obtained directly, it could serve as a starting point for a more accurate calibration process.

We have found that with architectural images, the set of matches found by automatic corner detection and variance normalized correlation can be very noisy. In fact less than ten percent of the matches found in this way were usually valid.

Let us now compare the problem of identifying a planar homography in a scene using RANSAC, with the problem of the fundamental matrix estimation with RANSAC, in the context of noisy data.

The theoretical expression of the probability of selecting a good subset of matches in RANSAC [11] is given by:

$$1 - (1 - (1 - \epsilon_{out})^p)^N \quad (9)$$

where p is the number of matches necessary to compute a solution (4 for \mathbf{H} , 8 for \mathbf{F}), ϵ_{out} is the proportion of the matches that are not compatible with the homography or fundamental matrix, and N the number of iterations performed. In the case of homography detection, ϵ_{out} is larger since not all good matches lie on the same plane. Here, we will make the reasonable assumption that half of the good matches are compatible with some homography. Using this expression, we can compute the ratio between the number of iterations that would be necessary to perform the two tasks with the same probability of success. We obtain

$$\frac{N_{\mathbf{H}}}{N_{\mathbf{F}}} = \frac{\log(1 - r^8)}{\log(1 - (\frac{1+r}{2})^4)} \quad (10)$$

where $\frac{N_{\mathbf{H}}}{N_{\mathbf{F}}}$ is the ratio of the number of trials needed to compute \mathbf{H} to the number of trials needed to compute \mathbf{F} , and r is the proportion of valid matches in the data set.

Figure 5 shows the theoretical ratio of the number of iterations needed to find the main homography in a scene using RANSAC, to the number of iterations needed for weak calibration using RANSAC, as a function of the proportion of the matches that are good matches. We notice immediately that it is far

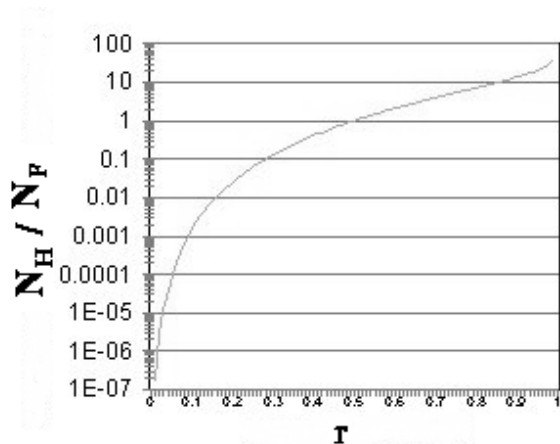


Figure 5: The number of trials required to find a planar homography in a scene compared to the number of trials required for weak calibration, both using RANSAC, as a function of the reliability of the match set used.

easier to find a planar homography than to perform calibration, when the set of matches is very noisy. For example, estimating \mathbf{H} is 10 times faster in the case of $r = 30\%$.

6 Conclusion

We have proposed in this paper, an algorithm for the detection of planar homographies in image pairs. The method is based on RANSAC, and we found that by using appropriate strategies, the algorithm can effectively detect the visible planar structures even if the available match set is unreliable. In fact, we were able to successfully identify homographies, even when this set contains as few as 5% of correct matches.

This robustness makes our algorithm potentially attractive to be used as a first step in a give vision process. In particular, we believe that the algorithm presented for planar homography detection could be used with great advantage towards weak calibration in the situation were the set of matches is noisy and when planes are expected to be present in the scene.

References

[1] P. Beardsley, P. Torr, A. Zisserman, 3D Model Acquisition from Extended Image Sequences, *Technical report*, Dept of Eng. Science, University of Oxford, 1996.

[2] A. Criminisi, I. Reid, A. Zisserman, A Plane Measuring Device, *Image and Vision Computing*, Vol. 17, 625-634, 1999.

[3] O. Faugeras, Q. Luong, S. Maybank, Camera self-calibration: Theory and experiments, *Proc. 2nd European Conference on Computer Vision*, Santa Margherita, pp. 321-334, 1992.

[4] R. Hartley. In Defense of the Eight-Point Algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 133-135, 1997.

[5] R. Laganière, A Morphological Operator for Corner Detection, *Pattern Recognition*, Vol. 31, pp. 1643-1652, 1998.

[6] D. Liebowitz, A. Criminisi, A. Zisserman, Creating Architectural Models from Images, *Proc. EuroGraphics*, 1999.

[7] M. Lourakis, S. Orphanoudakis, Visual Detection of Obstacles Assuming a Locally Planar Ground, *Proc. 3rd Asian Conference on Computer Vision*, Hong-Kong, pp. 527-534 1998.

[8] Q.-T. Luong, R. Deriche, O. Faugeras, T. Papadopoulo, On determining the Fundamental matrix: analysis of different methods and experimental results, *Technical Report*, INRIA, 1993.

[9] Q.-T. Luong, O. Faugeras, Determining the Fundamental matrix with planes: unstability and new algorithms, *Proc. Conference on Computer Vision and Pattern Recognition*, New-York, pp. 489-494, 1993.

[10] P. Pritchett, A. Zisserman, Wide Baseline Stereo Matching, *Proc. 6th International Conference on Computer Vision*, Bombay, pp. 754-760, 1998.

[11] A. Shashua, Trilinearity in visual recognition by alignment, *Proc. 3rd European Conference on Computer Vision*, pp. 479-484, 1994.

[12] P. Torr, A. Zisserman, Robust Computation and Parameterization of Multiple View Relations, *Proc. 6th International Conference on Computer Vision*, Bombay, pp. 727-732, 1998.

[13] P. Torr, A. Zisserman, S. Maybank, Robust Detection of Degenerate Configurations for the Fundamental Matrix, *Computer Vision and Image Understanding*, Vol. 71, pp. 312-333, 1998.

[14] B. Triggs, Autocalibration from Planar Scenes, *Proc. European Conference on Computer Vision*, Freiburg, pp. 89-105, 1998.



(a)



(b)



(c)



(d)



(e)



(f)

Figure 6: Three homographies found using our RANSAC-like algorithm on an image pair. (a) and (b) The point matches that agree with a first homography. (c) and (d) The point matches that agree with a second homography. (e) and (f) The point matches that agree with a third homography.