Junction Matching and Fundamental Matrix Recovery in Widely Separated Views

Etienne Vincent and Robert Laganière
School of Information Technology and Engineering
University of Ottawa
Ottawa, Canada, K1N 6N5
evincent,laganier@site.uottawa.ca

Abstract

A method for finding correspondences between widely separated views is presented. The proposed solution consists in estimating the local perspective distortion between the neighborhoods of junctions. To this end, a formulation is proposed, based on a constrained minimization involving an estimated fundamental matrix. An application is also proposed to fundamental matrix recovery using crude camera pose estimates.

1 Introduction

The correspondence problem is fundamental to computer vision, and a necessary step in many applications. It is however a difficult problem, particularly when correspondences are required between widely separated views, where perspective distortion makes the effective comparison of image areas difficult. This work proposes a method for establishing point correspondences between widely separated views, when the underlying epipolar geometry is known. This situation might present itself when the goal is to interpret a dynamic scene from images taken by fixed calibrated cameras, or in epipolar geometry recovery from an approximate camera pose, as is proposed later in this work.

Essentially, the correspondence problem will be solved by estimating the local perspective distortion between junction point neighborhoods, using the shape of these junctions and the camera system's epipolar geometry. Unlike most proposed solutions for matching in ways that are tolerant to perspective distortion, this formulation is completely invariant to the distortion, instead of an affine approximation.

2 Literature Review

Two important approaches exist to making feature matching robust to perspective distortion. Firstly, [5, 7, 10] use descriptions of feature points that are partially invariant to perspective distortion, and then compare them to establish correspondence. In general, such methods are only invariant to some affine or similarity approximation of perspective distortion. Furthermore, as the the degree of invariance of a description increases, its discriminating power necessarily diminishes.

BMVC 2004 doi:10.5244/C.18.10

The other approach is to estimate the local distortion between pairs of image regions, and then warp the regions before comparing them. This reduces the problem to comparing image regions as they would be seen from the same viewpoint, a situation in which correlation-based similarity measures work well. This approach was used in [8], but requiring the presence of special structures in the images such as sets of coplanar line segments. Similarly, [1] uses an iterative shape adaptation scheme to normalize image patches, and thus eliminate perspective distortion.

Other methods attempt to reach robustness to perspective distortion by using higher level features such as [11], where intensity profiles along line segments joining feature points are used. Finally, [12] suggests using a coarse to fine scheme to iteratively refine point correspondences.

The approach presented in [13] is somewhat similar to the one proposed here, being based on junction points. Image regions, defined by two intersecting image edges, are extracted in an affine-invariant way, and then compared. Their extraction might require a search in a two-dimensional space defined by the lengths along the edges. Finding two corresponding regions extracted in an affine-invariant way amounts to discovering an affine approximation of the homography relating the regions.

3 Epipolar Geometry and Planes

Epipolar geometry is an important and well known tool in computer vision. If $\mathbf{I}$ and $\mathbf{I}'$ are two views of a same scene, and $\mathbf{x}$ and $\mathbf{x}'$ are the projection on $\mathbf{I}$ and $\mathbf{I}'$, of a point $\mathbf{X}$ in space, $\mathbf{x}'$ should be located on $\mathbf{l}'$, the epipolar line of $\mathbf{x}$ in $\mathbf{I}'$. This epipolar line is the projection, in $\mathbf{I}'$, of the line in space passing through $\mathbf{X}$ and the first camera's focal point. The relationship between points and their epipolar lines can be expressed as $\mathbf{l}' = \mathbf{F}\mathbf{x}$, where $\mathbf{F}$ is the 3×3 *fundamental matrix*, $\mathbf{x}$ is expressed in homogeneous coordinates, and $\mathbf{l}'$ is the set of all points $\mathbf{p}$ such that $\mathbf{l}'^{\top}\mathbf{p} = 0$.

The epipolar relationship does not depend on the structure of the scene being viewed. However, more can be said when two views of a planar area are considered. Two views of a plane are related by a projective linear transformation, or *homography*. If $\mathbf{X}$ happens to be located on a planar surface, $\mathbf{x}$ and $\mathbf{x}'$ are related by $\mathbf{x}' = \mathbf{H}\mathbf{x}$, where $\mathbf{H}$ is the 3×3 planar homography matrix.

When such a homography relating two views of a planar area is known, it can be used to remove the perspective distortion through warping. In the case of planar surfaces taken from widely different viewpoints, intensity patterns in untransformed image areas exhibit low correlation, but after warping, can become very similar.

4 Homography Estimation

Homographies are determined by eight degrees of freedom (DOF). Thus eight independent constraints are needed for their estimation. However, if a homography $\mathbf{H}$ is sought between images related by a fundamental matrix $\mathbf{F}$, then $\mathbf{H}^{\top}\mathbf{F}$ is skew-symmetric:

$$\mathbf{H}^{\top}\mathbf{F} + \mathbf{F}^{\top}\mathbf{H} = \mathbf{0} \tag{1}$$

This results in five independent linear constraints on $\mathbf{H}$; three more are required to fully describe a homography [6].

Homographies describe the relationship between views of corresponding points lying on a planar surface, but also between lines. If $\mathbf{k}$ and $\mathbf{k}'$ are projections of a line in space lying on a plane whose views are related by an homography $\mathbf{H}$, they will be related by:

$$\mathbf{k} = \mathbf{H}^\top \mathbf{k}' \tag{2}$$

Hence a line correspondence $(\mathbf{k}, \mathbf{k}')$ puts two independent linear constraints on $\mathbf{H}$, through Equation (2). And thus, the fundamental matrix and two line correspondences lying on a common planar area determine the homography between the views of that planar area, as they provide $5+2+2$ constraints on the 8 DOF of the homography.

5 Affine Transformations from Junction Points

In this work, junction points will be matched across images. These junction points will be defined as points at the origin of two line segments. When corresponding junction points are compared, the lines defining them also correspond. These line-correspondences can then be used to constrain the estimation of local homographies.

As proposed in [4], these line-correspondences could define a simple affine transformation approximating the homography. But the additional (generally inaccurate) assumption that points along the lines remain at the same distance from the junction points in the different views is also needed to define a unique transformation. This affine approximation of the homography relating the neighborhoods of $\mathbf{x}$ and $\mathbf{x}'$, which are at the intersection of $\mathbf{k}_1$,$\mathbf{k}_2$ and $\mathbf{k}'_1$,$\mathbf{k}'_2$ respectively, is the $\mathbf{H}$ such that:

$$\mathbf{H}\mathbf{x} = \mathbf{x}' \;\text{ and }\; \mathbf{H}\mathbf{x}_i = \mathbf{x}'_i \;\text{ for }\; i \in \{1,2\} \tag{3}$$

where $\mathbf{x}_i$ is an arbitrary point along $\mathbf{k}_i$ and $\mathbf{x}'_i$ is a point on $\mathbf{k}'_i$ chosen such that $\|\mathbf{x}-\mathbf{x}_i\|^2 = \|\mathbf{x}'-\mathbf{x}'_i\|^2$. This approximation of the homography can be used to warp one of the point's neighborhood before comparing it to the other.

6 Local Homographies from Junctions

When $\mathbf{F}$ is known, Equation (1) can be used to constrain an homography, together with the line correspondences. Then, the system is over-constrained, and a minimization is required. If the line-correspondences are assumed to be exact, the best approximation of the homography agrees with them, but minimizes the other constraint. Thus, the homography is defined as the $\mathbf{H}$ which minimizes the Frobenius norm:

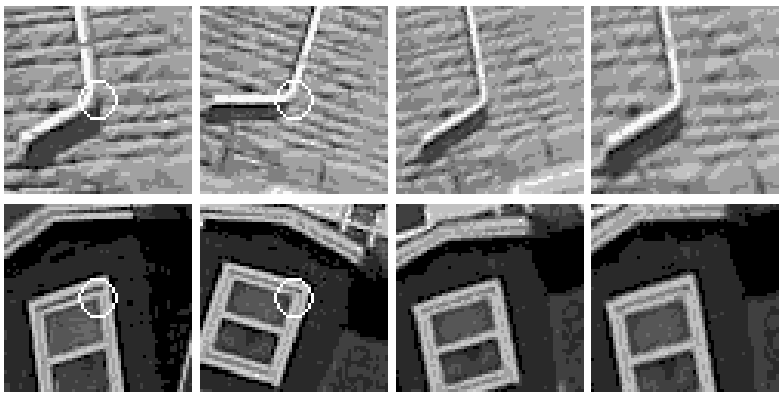
$$\|\mathbf{H}^\top\mathbf{F} + \mathbf{F}^\top\mathbf{H}\|_F \;\text{ subject to }\; \mathbf{k}_i = \mathbf{H}^\top\mathbf{k}'_i \;\text{ for }\; i \in \{1,2\} \tag{4}$$

where $\mathbf{k}_i$ and $\mathbf{k}'_i$ are the lines defining the junctions. A least-squares solution to this constrained system of linear equations can be found directly and efficiently [2]. In fact, only the singular value decomposition of a 9×9 matrix, and another 6×9 matrix are required, with some simple matrix multiplications.

Figure 1 shows two test images, along with epipolar lines. Figure 2 (a) and (b) show closeups of two junction point neighborhoods. It is seen that the corresponding neighborhoods are significantly different. (d) shows (b) after being warped by homographies obtained from Equation (4). These images are obviously very similar to (a).

Figure 1: A pair of images with their corresponding epipolar lines. 2.

.



(a) (b) (c) (d)

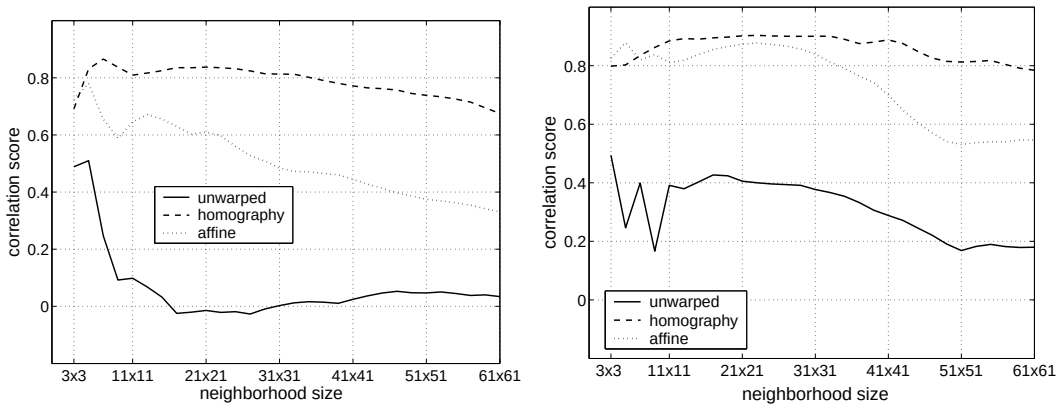
Figure 2: (a) and (b): closeups of image regions centered on junction points from each image in Figure 1. (c): image (b) warped by a transformation computed using Equation (3). (d): image (b) warped by an homography computed using Equation (4).

For comparison, (c) shows the result of warping (b) by the affine transformation computed only from the junction lines using Equation (3), and not the epipolar geometry. It can be seen how this transformation incorrectly preserves the length of the line segments defining the junctions, while they are stretched appropriately in (d). The correlation scores between the warped image regions and the corresponding left image regions are shown in Table 1. It is seen that the proposed method is a clear improvement.

Figure 3 shows the correlation scores between the junction neighborhoods of Figure 2, and their warped counterparts, taken over neighborhoods of different sizes. It is seen that the correlation is improved by the proposed method defined by Equation (4), over the simpler method defined by Equation (3), or over using no warp at all.

7 The Matching Process

The process of finding corresponding junction points between images will now be examined. First, junction points must be detected in both images. This can be achieved with methods based on differential operators, junction model fitting, or the use of edge maps. Here, the method described in [3] was used, but what follows does not depend on this



(a) (b)

Figure 3: Correlation scores between the regions shown in Figure 2, over neighborhoods of varying size around the junction points. (a): gutter image. (b): window image

	unwarped	**affine warp**	**homography warp**
first row of Figure 2	0.034	0.331	0.675
second row of Figure 2	0.180	0.546	0.784

Table 1: Correlation between the two junction neighborhoods shown in Figure 2

choice of specific operator.

As the epipolar geometry is assumed to be available, it can be used to guide matching. Thus, only junction points in the second image that lie close to the epipolar line of a junction point in the first image are considered. For each pair of considered junction points, a homography is computed using Equation (4), and then used to warp the neighborhoods of junctions in the second image, towards a comparison to the first image junction neighborhoods. Normalized correlation is used for the comparison, and only pairs exhibiting high correlation are kept.

When computing a homography between two junctions, it must be decided which line segment corresponds to which one in the other view (which is $\mathbf{k}'_1$, and which is $\mathbf{k}'_2$ in Equation (4)). Fortunately, the clockwise order of the line segments should be preserved. Thus, $\mathbf{k}_1 \times \mathbf{k}_2$ will have the same sign as $\mathbf{k}'_1 \times \mathbf{k}'_2$, and $i, j \in \{1,2\}$ are determined such that $(\mathbf{k}_1 \times \mathbf{k}_2)(\mathbf{k}'_i \times \mathbf{k}'_j) > 0$.

Finally, it is advantageous to require that only the best matches be kept, for given junction points (uniqueness). As well, a junction point's best match in the other image, should also also have the first point as its best match (symmetry) [14].

Figure 4 shows an image on which the matching scheme was applied, using 19×19 correlation windows and different thresholds on the correlation. A similar matching scheme using affine transformations defined by Equation (3), rather then the proposed homographies was also applied. The resulting match sets contained varying quantities of accurate matches, and proportions of mismatches. Each point on the graph represents the result of an experiment with a different threshold on correlation. The vertical axis records the proportion of accurate matches in the result, and the horizontal axis records

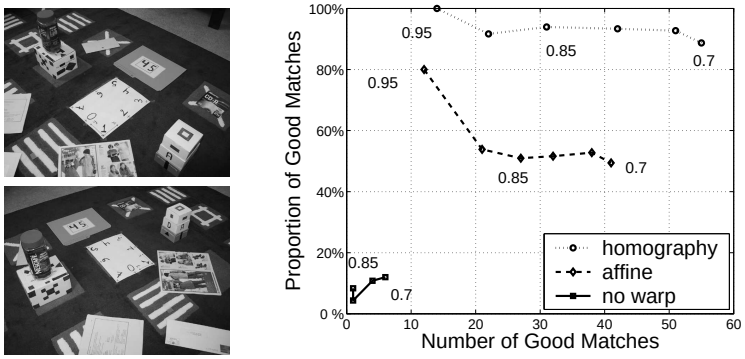


Figure 4: Test images, and the result of matching them with various correlation thresholds. The number and percentage of good matches obtained are shown.

Figure 5: Two image pairs used to test the proposed matching scheme.

their quantity. Matching is successful when it results in many junction pairs, among which few are mismatches (a situation corresponding to the upper-right part of the graph). It is seen that the proposed homography estimation method compares favorably with the use of the simpler approximation of perspective distortion.

For comparison, similar experiments were conducted where correlation was applied without any warp, with disastrous results. Small image patches undergo less perspective distortion, so 3×3 neighborhoods were used for the unwarped experiments of Figure 4, but the results were nevertheless far inferior to those obtained from the other methods.

Table 2 shows the result of applying the proposed matching scheme to the four widely separated image pairs shown in Figures 1, 4, 5, with a correlation threshold of 0.7 and using 19×19 neighborhoods. The results of the same scheme, but with the proposed homography estimates replaced by the affine transformations described by Equation (3) are also recorded for comparison. We still see a significant improvement from the use of the proposed method. The results of other experiments, where correlation was applied without prior warping are also shown in Table 2. For these experiments, the correlation window size was reduced to 3×3, but still produced far inferior results.

8 Epipolar Geometry Refinement

Fundamental matrix estimation is a key step in many computer vision applications. Unfortunately, it is very difficult to estimate a fundamental matrix between widely separated views. It could sometimes be computed from camera pose, and internal parameter esti-

	unwarped	affine warp	homography warp
Figure 1	10.5%	57.7%	75.9%
Figure 4	12.0%	49.4%	88.7%
Figure 5 (left)	25.0%	65.6%	88.9%
Figure 5 (right)	30.9%	46.3%	61.9%

Table 2: Proportion of good matches found in widely separated views

(a) (b)

Figure 6: (a): Crude epipolar geometry estimate. (b): Improved with proposed approach.

mates, but such a direct approach is very sensitive to inaccuracies in measurements.

The solution will first estimate camera pose with positional sensors, yielding a crude estimate of the fundamental matrix. This estimate will then be used, by the previously described matching scheme, to find point matches between the views. These matches can, in turn, be used to produce a refined estimate of the sought fundamental matrix.

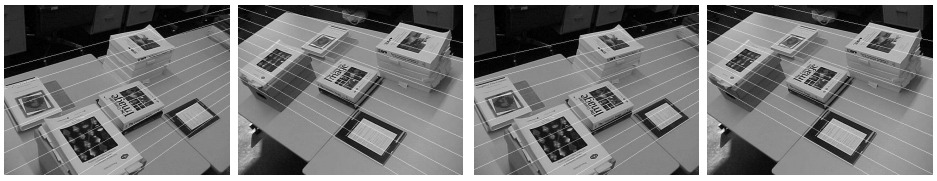
Figure 6 (a) shows an initial estimate of a pair's epipolar geometry, obtained from approximate measurements of the pose. Matches were then sought to improve the estimate. Four iterations were used, where junction points were iteratively matched and a fundamental matrix estimated from the result. This was done using the RANSAC-based software described in [9]. A set of 35 pairs containing 10 mismatches was obtained and used to compute the final epipolar geometry shown in Figure 6 (b). It can be seen that the refined estimate is much better than the original, as the drawn epipolar lines now appear to correspond more accurately.

Figure 7 shows the result of a similar experiment. This time, the image pair is the result of a small displacement of one of the cameras from a previously calibrated image pair. Thus, the previous epipolar geometry can be a starting point in estimating the new one. Note that it would be very difficult to extract enough correct matches automatically, without warping, as these views are very widely separated. From the 47 junction point pairs obtained, 16 agreed with the refined epipolar geometry shown in Figure 7 (b).

9 Tolerance Analysis

If the presented fundamental matrix recovery scheme is to be used in practice, the matching scheme on which it relies must tolerate errors in the original estimate of the fundamental matrix. To demonstrate this tolerance, calibrated test images were used. They are shown in Figure 8 (a), with their correct epipolar geometry. Experiments were conducted where the camera pose used for matching was perturbed, and the effect on the number of correct matches found was recorded.

Some results of these trials are recorded in Figure 9. Each point in these graphs



(a) (b)

Figure 7: (a): Crude epipolar geometry estimate. (b): Improved with proposed approach.

(a) (b)

Figure 8: Test images, with an example of the epipolar geometry computed from perturbed pose. Points used to evaluate the fundamental matrix deformation are also shown.

represents an experiment with a particular set of perturbed parameters. Each experiment was conducted with a fundamental matrix computed from a camera pose with position randomly selected in a cube of 30cm diameter centered on its true location, and each of its three orientation components within 3° of its true value.

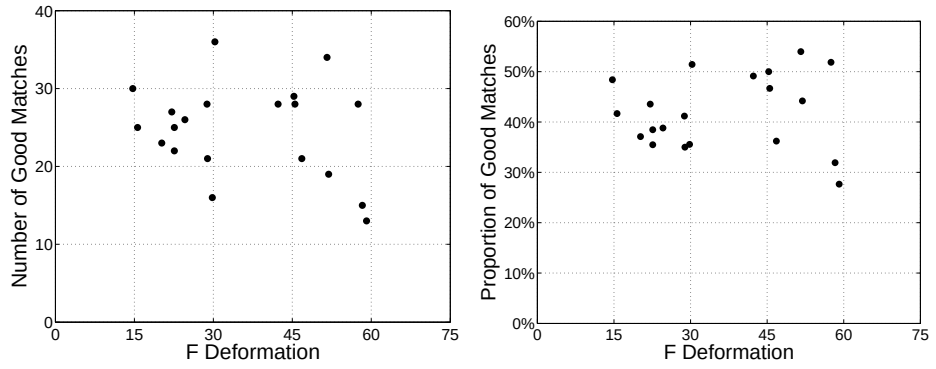
For each experiment, a measure of the deformation in the perturbed fundamental matrix was recorded as the maximal distance between a left-image point's epipolar line, and its corresponding point in the right image. Figure 8 (b) shows an example of a perturbed epipolar geometry, also illustrating the computation of the deformation measure. It is seen that the feature point shown on the right image, has an epipolar line in the left image which runs 29.8 pixels from its expected position on the corresponding point. Since this is the largest such distance, it is taken as the measure of fundamental matrix deformation.

In each experiment, the accurate matches found by the proposed matching scheme were counted through visual inspection. The points in Figure 9 (a), have their x-coordinate as the given measure of deformation, and their y-coordinate as the number of accurate matches that were found. In Figure 9 (b), the y-coordinates represent the proportion of accurate matches in the computed match sets.

For each experiments, it is seen that an adequate number of matches were found (between 14 and 36). These numbers are relatively low, but sufficient for the reliable fundamental matrix estimation using a RANSAC-based scheme. Also, the number of mismatches remained reasonable, as it varied between 24 and 41, resulting in a proportion of good matches that was between 28.9% and 52.9%.

10 Comparison with Plane Reconstruction

In the previous section, experiments were conducted where an approximation of the camera pose was available. Under such circumstances, and when comparing two views of



(a) (b)

Figure 9: Tolerance of the fundamental matrix recovery scheme. Each point corresponds to an experiment with a given set of perturbed parameters. (a) Total number of good matches versus **F** deformation. (b) Proportion of good matches versus **F** deformation.

a planar surface, the camera pose could have alternatively been used to reconstruct the planar surfaces in space. These surfaces would then yield local homographies between the views. Such an approach will now be compared with the previously described one, of estimating the homographies using a fundamental matrix only.

Planar surfaces in space, can be determined by the two lines intersecting at a junction. These lines, in space, are easily computed from corresponding lines in the two images. However, because of inaccuracies in the estimates, it is unlikely that the two computed lines will actually intersect in space. For this reason, the surfaces are approximated as those parallel to both lines and located at the midpoint of the minimal distance between them. But due to this approximation, the computed homographies now do not necessarily transfer the junction points exactly on their counterparts. This can be fixed by rectifying, for each considered junction points, but there is no obvious best way of doing it (should the correction be in rotation or in translation?). An alternative solution was to compound a translation with the initial homography estimate, to correct for the shift in position.

The use of camera pose information did not greatly improve the estimated homographies. When the junction points in the images of Figure 8 were matched using this approach, 24 correct matches were found (and 9 mismatches), versus 25 correct matches (and 6 mismatches) when only the fundamental matrices was used. With the perturbed sets of parameters from the experiments presented in Figure 9, both methods again performed similarly. The use of pose information yielded on average, 26.6 correct matches, constituting 42.0% of the resulting match set, while the use of the fundamental matrix yielded 24.7 correct matches on average, accounting for 41.9% of the match set. Furthermore, each method produced superior results in about half of the trials.

Thus, the use of pose information is not greatly advantageous in the described approach to matching. In addition, it is computationally more expensive, and only appropriate in a narrower range of applications. Therefore, the proposed method, which requires only a fundamental matrix, should generally be preferred.

11 Conclusion

An effective solution was proposed for the correspondence problem between widely separated views. Local homographies were defined using the shape of junctions, and estimates of the epipolar geometry. It was shown experimentally that when camera pose can be estimated, an exact fundamental matrix can be recovered in situations where other schemes would fail. This approach to fundamental matrix recovery would be useful in the difficult situation of widely separated cameras where the usual correlation schemes fail.

References

[1] A. Baumberg. Reliable feature matching across widely seperated views. In *CVPR*, volume 1, pages 774–781, 2000.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University, 2000.

[3] R. Laganière and R. Elias. The detection of junction features in images. In *ICASSP*, 2004.

[4] R. Laganière and E. Vincent. Wedge-based corner model for widely separated views matching. In *ICPR*, pages 672–675, 2002.

[5] D. Lowe. Object recognition from local scale-invariant features. In *ICCV*, volume 2, pages 1150–1157, 1999.

[6] Q.-T. Luong and T. Viéville. Canonical representations for the geometries of multiple projective views. *Comp. Vision and Image Underst.*, 64(2):193–229, 1996.

[7] P. Montesinos, V. Gouet, R. Deriche, and D. Pelé. Matching color uncalibrated images using differential invariants. *IVC*, 18(9):659–671, 2000.

[8] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *ICCV*, pages 754–760, 1998.

[9] G. Roth and A. Whitehead. Using projective vision to find camera positions in an image sequence. In *Vision Interface*, pages 225–232, 2000.

[10] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *PAMI*, 19(5):530–535, 1997.

[11] D. Tell and S. Carlsson. Wide baseline point matching using affine invariants computed from intensity profiles. In *ECCV*, pages 814–828, 2000.

[12] P. Torr and C. Davidson. Impsac: Synthesis of importance sampling and random sample consensus. In *ECCV*, pages 819–833, 2000.

[13] T. Tuytelaars, L. Van Gool, L. D'haene, and R. Koch. Affinely invariant regions for visual servoing. In *Int. Conf. on Robot. and Autom.*, pages 1601–1606, 1999.

[14] E. Vincent and R. Laganière. Matching feature points in stereo pairs. *Machine Graphics & Vision*, 10(3):237–259, 2001.