Proceedings of the 1998 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Victoria, B.C., Canada • October 1998

Landmark-based Navigation Using Projective Invariants

Vassilios S. Tsonis[1,2], Konstantinos V. Chandrinos[1] and Panos E. Trahanias[1,2]

[1]Institute of Computer Science
Foundation for Research and Technology – Hellas
P.O.Box 1385, Heraklion, 711 10 Crete, Greece

[2]Department of Computer Science
University of Crete
P.O.Box 1470, Heraklion, Crete, Greece

e-mail: {tsonis,kostel,trahania}@ics.forth.gr

Abstract

Landmark-based navigation usually relies on the identification and subsequent recognition of a number of environment objects, that are deemed adequate in describing the workspace structure. This process is inherently difficult in practice, due to the many different poses of an object that may be encountered in navigational trials. To aleviate for that, we propose an approach that employs projective invariants computed on quintuples of points as workspace landmarks. Such quantities remain invariant under different camera positions and provide for effective description of the workspace structure. In order to identify potential corresponding quintuples in image frames, we introduce a simple test based on the covariance matrix estimate of each quintuple. With this test, we effectively by-pass the calculation of point correspondences. Since the above test indicates correspondence between quintuples, and not between their individual points, we subsequently employ a permutation projective invariant for quintuple recognition. Our approach has been extensively evaluated using synthetic as well as real environments. The results obtained verify its robustness along with its applicability in robotic navigation.

1 Introduction

In this paper we present work concerning the ability of robotic platforms to visually recognize landmarks. Landmark recognition is expected to enhance the capabilities of modern robotic platforms and pave the way towards really autonomous robots. Current techniques, entailing accurate measurements and/or knowledge of the environment, confine autonomous navigation to known or engineered environments.

The use of visual landmarks for topological navigation has appeared previously in the literature, but most attempts seem to focus on simple landmark patterns, since there is a certain amount of difficulty that appears when realistic landmarks are to be tackled [1]. Representative approaches include *predesigned landmarks* [2] and *selected landmarks*, where the workspace is known in advance [3]; another approach utilizes *straight lines* [4, 5]. In recent work in our laboratory [6], we have presented a geometric method for visual recognition of landmarks. However, the solution demonstrated relied heavily on the assumption that the workspace consists of parallelepipeds, confining the autonomicity of a robot's navigational task. In this paper we address the same problem and propose a solution for more general workspaces, making use of projective invariants.

Invariants of a variety of features that appear in an image have come of frequent use in computer vision, particularly for tasks of pattern matching and object identification [7]. Their use seems ideal if one wants to ameliorate the computational complexity imposed on a matching search by the construction of a total visual database to reflect the environment a robotic platform has to face. This complexity is mainly induced by the fact that readily observed geometric properties are not invariant under projective transformation. Because of that, either a multiplicity of pose-based transformations of each model will be stored in the database, or pose-dependent variants of each model have to be constructed at run-time, since all these variants need to be tested before any hypothesis can be discarded. With either solution, the matching search is computationally taxed.

2 Method Overview

To tackle these problems our approach adopts projective invariants, particularly 2-D cross-ratios, that are used to recognize and store landmarks during a learning phase. These landmarks are then matched to re-discovered landmarks at navigation time with minimum computational cost. Although invariants provide a fast indexing method into databases of models,

*This work was supported by EC Contract No. ERBFMRX-CT96-0049 (VIRGO http://www.ics.forth.gr/virgo) under the TMR Programme and the General Secretariat for Research and Technology, Greece, under Grant No. 6060

the use of projective invariants as a verifier between features of different frames is often avoided since it usually presupposes solving the correspondence problem. Perhaps the greatest difficulty lies at the fact that most projective invariants are permutation sensitive, i.e. a different ordering of the features usually produces a different value of the projective invariant. In such a case, a value for the invariant has to be stored for any of the possible orders of the features under investigation, i.e for features based on n points, one needs to store $n!$ values. Even in that case, the resolution of the invariant value space would be too fine for a unique matching decision to be made. In our work, we treat quintuples of co-planar points and we by-pass the point-to-point correspondence problem, which is not trivial, by exploiting a permutation insensitive projective invariant. This allows us to store one value for every quintuple, instead of 120, keeping both the search space and matching time to a minimum. The sole assumption of our method is that our environment contains planar surfaces. All conclusions at navigation time are drawn without any knowledge of the environment, pose estimation or calibration of the visual system.

Feasibility and robustness of the proposed method have been proven through experiments both on software simulated realistic indoor environments and on actual indoor environments using TALOS, a robotic platform with an active vision head available to the Computer Vision and Robotics Lab at FORTH.

In the rest of the paper, section 3 offers a formal definition of what constitutes a landmark in our context and details on landmark extraction and section 4 focuses on landmark recognition. Section 5 details our experimental set-up and presents respective results, while section 6 concludes the paper with a brief discussion and outline of future directions.

3 Learning

In this section we describe the learning phase during which landmark patterns are extracted and stored in a model base for future reference. We review the 2D cross-ratio and state the permutation insensitive version we shall use for our calculations in the following sections. We also give the details of our extraction method, since it is, in essence, the same method that we use later, during the recognition phase, to obtain the visual landmarks observed before matching them to the model base.

3.1 Two-dimensional cross-ratios

Computer vision researchers have shown aptitude in using projective invariants for recognition in model based vision, shape descriptors for 3D objects and for producing the projection of a structure from one view to another (transference) [8, 9]. Projective invariants have also been used for the characterization of unknown geometric structure [7]. In this latter use, the non-planarity of five points depicted on an image can be established by calculating a particular invariant, the *cross-ratio*, from two different views without *a priori* scene knowledge or camera calibration. If the values of the invariant calculated on these points from two different views differ significantly then these points are not co-planar. Conversely, if the values coincide we have strong evidence, albeit not proof, that the points are co-planar.

There exist several equivalent definitions of 2-D cross-ratios. In our calculations, we utilized the following definition:

$$\begin{aligned}\mu &= [P_1, P_2, P_3, P_4, P_5] \\ &= \frac{\det(P_1, P_2, P_4)\,\det(P_1, P_3, P_5)}{\det(P_1, P_3, P_4)\,\det(P_1, P_2, P_5)} \\ &= \frac{\begin{vmatrix} x_1 & x_2 & x_4 \\ y_1 & y_2 & y_4 \\ z_1 & z_2 & z_4 \end{vmatrix} \begin{vmatrix} x_1 & x_3 & x_5 \\ y_1 & y_3 & y_5 \\ z_1 & z_3 & z_5 \end{vmatrix}}{\begin{vmatrix} x_1 & x_3 & x_4 \\ y_1 & y_3 & y_4 \\ z_1 & z_3 & z_4 \end{vmatrix} \begin{vmatrix} x_1 & x_2 & x_5 \\ y_1 & y_2 & y_5 \\ z_1 & z_2 & z_5 \end{vmatrix}} \qquad (1)\end{aligned}$$

where P_i are five points, not three of which are colinear, described in homogeneous co-ordinates by (x_i, y_i, z_i). We consider all points to be Cartesian and not ideal, therefore in all our calculations we take z_i equal to 1. Noticing that P_1 is the only point involved in all the determinants we can easily deduce that there are five different cross-ratios defined in the above manner. Straightforward calculations can show that any two of the five different cross-ratios are enough to express the other three. In accordance with [10] we calculate one more cross-ratio,

$$\nu = [P_2, P_1, P_3, P_4, P_5] \qquad (2)$$

Based on these two different cross ratios, one can obtain [10] a permutation insensitive projective invariant $K[\mu, \nu]$, given by

$$\begin{aligned}K[\mu,\nu] &= J[\mu] + J[\nu] + J[\frac{\mu}{\nu}] + \\ &\quad + J[\frac{\nu-1}{\mu-1}] + J[\frac{\mu(\nu-1)}{\nu(\mu-1)}] \qquad (3)\end{aligned}$$

where $J(\lambda)$ is a one dimensional permutation insensitive invariant computed as,

$$J[\lambda] = \frac{2\lambda^6 - 6\lambda^5 + 9\lambda^4 - 8\lambda^3 + 9\lambda^2 - 6\lambda + 2}{\lambda^6 - 3\lambda^5 + 3\lambda^4 - \lambda^3 + 3\lambda^2 - 3\lambda + 1} \qquad (4)$$

3.2 Visual Landmarks

For the purposes of this paper we define *visual landmarks* to be sets, VL_i containing *sub-landmarks* $l_{i,j}$. The sub-landmarks are quintuples of co-planar points $P_{i,j,k}$, $k=1..5$ obeying simple assumptions such as *saliency* and *spatial dispersion*. The points that form these quintuples are derived directly from the image using first a robust corner detector [11]. The points derived from the corner-detector indicate potential landmark locations and form a *corner map*. In accordance with recent theories of active and purposive vision we further restrict our set of points by means of a "focus-of-attention" mechanism. Taking advantage of the fact that objects which can be characterised as landmarks should form distinctive enough patterns to be easily extracted from the rest of the environment, we proceed in the construction of a *saliency map* [6, 12]. This map brings out the most distinctive features of an image by assigning to their areas higher values as opposed to smooth regions that receive lower ratings. We compute the saliency map using a number of features that detect in a quantitative manner areas in our images which contain distinctive objects. These features, assuming a window W of our image, include area correlation, image entropy, standard deviation of the intensity histogram over W, as well as standard deviation of pointwise differences in W over successive images.

Out of the points indicated from the intersection of the corner map and the saliency map we pick quintuples l_{ij} of points $P_{i,j,k}$ that are close enough to each other with respect to an image window W_a, but satisfy a threshold for spatial dispersion. We also take consideration so that no three of the points in a quintuple are co-linear, otherwise we cannot apply the 2-D cross-ratio. On these quintuples we calculate the permutation insensitive projective invariant of (4). Prior to commiting l_{ij} to a legal sub-landmark, we need to establish the co-planarity of the five points $P_{i,j,k}$ that constitute l_{ij}. This is based on results from [8] using the fact that the cross-ratio of the points remains invariant over different frames, provided the points used are co-planar.

Since the robotic platform is not assumed to move fast with respect to the frame grabbing, we may safely assume that consecutive frames F, F' do not differ substantially in their structure. This allows for detection of new quintuples l'_{ij} in F', which are potential corresponding candidates for each quintuple l_{ij} in F, by repeating the steps detailed above.

As already mentioned the *invariant criterion*, i.e. the fact that calculation of the invariant should yield the same value over corresponding quintuples in different frames, is rather a strong evidence but not a proof of planarity and coincidence of l_{ij} and l'_{ij}, therefore we need to confine the search in F' to a few (ideally, one) quintuples for which there is indication of correspo-nence. To achieve this, we introduce the *covariance matrix test* which is used as a measure of similarity between l_{ij} and candidate quintuples l'_{ij} by quantifying and comparing the spatial dispersion of each one's points. By employing this test, we explicitly by-pass the point-to-point correspondence problem between the points in l_{ij} and candidate quintuples l'_{ij} solving the far easier spatial distribution correspondence problem between quintuples of points. In essence, the covariance matrix test, serves as a quantitative criterion of similarity between the spatial distributions of the points of two quintuples. Its introduction is justified, since we expect this distribution to change only slightly when the same scene is imaged from different viewpoints.

In order to detect candidate quintuples we establish an area which in the first frame F contains the quintuple l_{ij}. This area is the one we focus on in the consecutive frame F', expanded appropriately to cater for the motion of the robot. To verify that the matching of quintuples in consecutive frames is correct and single out one -or few- corresponding quintuple in F' for each quintuple in F we construct all possible pairs between every quintuple l_{ij} identified in F and its candidate corresponding quintuples l'_{ij} in F'. For each such pair (l_{ij},l'_{ij})we examine the distances of their covariance matrices . We built the covariance matrix for each quintuple according to the unbiased estimator:

$$\hat{\Sigma}_u = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T \qquad (5)$$

where the vector $\mathbf{x}_k = (x_k, y_k), k = 1..5$ is assumed to be the estimated distribution followed by the points $P_{i,j,k}$ in any quintuple and $\boldsymbol{\mu}$ the vector of the *mean*. We then calculate the norm

$$||C|| = max \sum_{q=1}^{2}\sum_{r=1}^{2} \left||\alpha_{q,r}| - |\alpha'_{q,r}|\right| \qquad (6)$$

where $\alpha_{q,r}$ and $\alpha'_{q,r}$ are the respective elements of the two covariance matrices, calculated for each pair of quintuples $(l_{i,j}, l'_{i,j})$ under consideration. The lower the value of this norm, the higher the fidelity on the matching is. As a last step we verify that our originally selected quintuple and the one that the covariance matrix test yields as corresponding give the same invariant value. This is a strong indication that the five points under consideration are co-planar and hence the points $P_{i,j,k}$ originally selected qualify for a sub-landmark, $l_{i,j}$.

3.3 Topological map construction

During a training period the robotic platform identifies and stores landmarks and their respective sub-landmarks in the above manner. Attached to each sub-landmark we also save the value of the covariance matrix norm and the value of the projective invariant. Moreover, we save references to navigational actions selectable at each landmark, building in effect a topological map, which can be used at runtime for decisions concerning the appropriate navigational task.

4 Landmark Recognition

Once our platform is considered trained we can have our robot move around and recognize landmarks. In doing this, we follow a procedure similar to the one described earlier for extracting landmarks in the first place. That is, we first apply the corner detector and then the procedures that build the saliency map. Having built the saliency map, we consider the intersection of the saliency map and the corner map and we begin testing hypotheses by looking in the model base at areas close to the areas dictated by our landmarks. The intersected saliency and corner map for the image we get at navigation time may introduce more than one candidate quintuples in the respective area, therefore the test of the covariance matrix is used again to narrow down the possibilities of a match. Finally, applying the projective invariant and comparing it with the stored value of the original landmark in the model-base indicates first of all that the new quintuple under consideration contains co-planar points and resolves any ties between quintuples that might have risen from the covariance matrix test. This whole procedure results at one quintuple being identified to be the same as the stored one, therefore *a sub-landmark has been recognized* and according actions for the current navigation task may take effect. Extensive experiments documented in the following section have proven the feasibility of this approach. Note again at this point that using a permutation insensitive invariant frees us from solving the problem of one-to-one correspondence between the points that form the quintuples.

5 Experimental Results

To quantitatively evaluate our method we followed a dual approach. First, we set up a number of experiments with simulated environments where the camera could be placed at different view points and register what was seen. We used these simulated world tests, where the ground truth is known, to evaluate the covariance matrix test, i.e. to verify that the lowest values of the norm described in Eq. (6) appear when a quintuple is tested against its trully respective quintuple in different frames. We also verified that the value of the norm is sensitive to noise, which is essential in our case, as it increases drastically when one or more false correspondence points are introduced . The simulated environment tests which we describe below allowed us to define a threshold of acceptable test values that behaved satisfactorily during landmark extraction and recognition in real environments.

To set up our experiments we used *POVRAY 3.0* on a Solaris 2.6 platform, to create a sequence of 50 frames. The frames depicted a world constructed by 5 surfaces, i.e. three walls, ceiling and floor. We tested our algorithms on the frames derived in this way and we saw that the results were valid except for the statistical error. On every frame we randomly selected 8 quintuples, 2 per surface, excluding the far-end wall. For every possible pair of frames we apply the covariance matrix test on all 8 corresponding quintuples, which resulted to 9,800 experiments for correct matching of sub-landmarks. In Fig. 1 we present histograms depicting the values of the covariance matrix test against the number of tests that returned each value. As can be observed, in Fig. 1a where we run the test on the actual corresponding points for each quintuple, the covariance matrix test reported values well confined in a certain interval, clearly marked by a threshold. In Fig. 1b, where only four points of each quintuple are in true correspondence, the values spread out in a much larger interval, and only a small percentage of them satisfies the threshold appearing in Fig. 1a. In Fig. 1c to Fig. 1e we relax the true correspondence one point at a time with obvious value shift on behalf of the covariance matrix test, until no corresponding points exist in the quintuples under consideration (Fig. 1f). This way it has been verified that the covariance matrix test is a valid quantitative criterion to express the proximity in space of points belonging to quintuples registered from different viewpoints.

Next, we implemented the above mentioned algorithms and verified them using TALOS, the mobile robotic platform available at the Computer Vision and Robotics Lab, at FORTH. A number of experiments have been conducted that demonstrated the robustness and applicability of the method. In Fig. 2 we present a sample result illustrating the steps both during the learning phase (left column) and the recognition phase (right column). Fig. 2a shows a frame encountered in the learning phase with the points detected as candidates for quintuples superimposed on the raw data. The actual sub-landmarks extracted in the manner described in section 2 are outlined by their *Minimum Bounding Rectangle* (Fig. 2c). The covariance matrix test for these sub-landmarks (computed as described in Section 3.2) has resulted in the values 747 (left sub-landmark) and 504 (right sub-landmark) respectively, which are in accordance with the imposed threshold. The permutation insensitive invariant value for establishing the co-planarity of the points constituting the two sub-landmarks has shown a difference of 3% and 4% respectively. The same scene viewed from a different vantage point has been

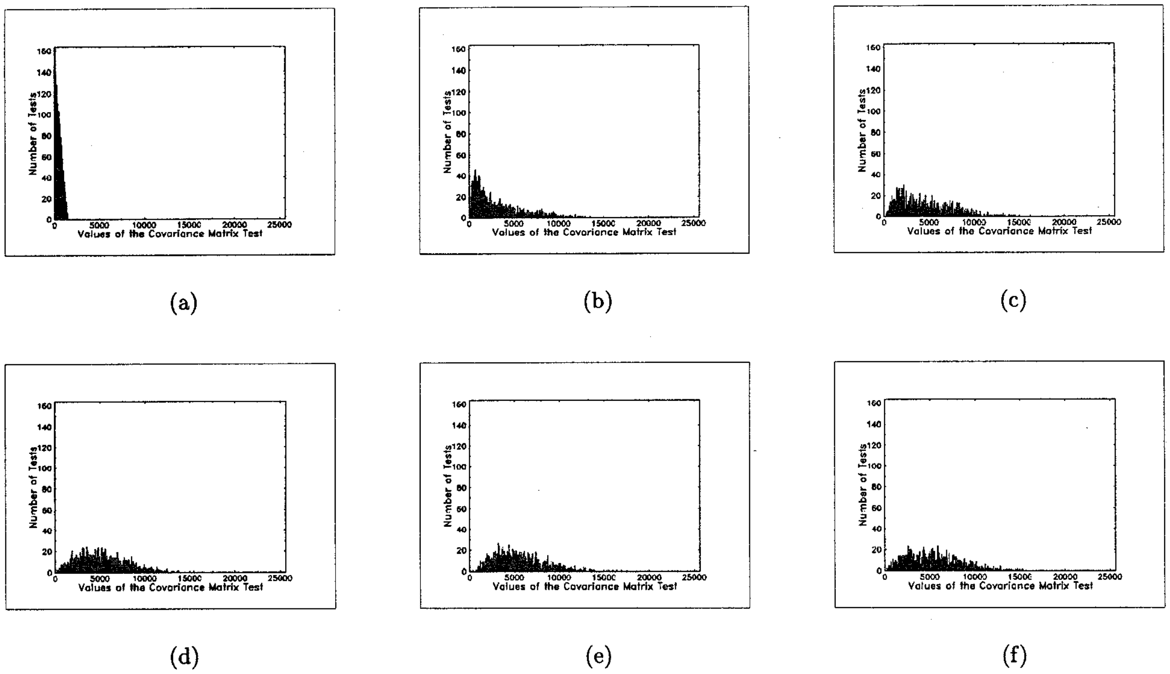


Figure 1: Results of the covariance matrix test (see text for explanation).

observed during a navigation trial and is illustrated in Fig. 2b. As can be verified from Fig. 2d the same sub-landmarks have been succesfully recognized and they are outlined by the expanded window which the algorithm used to detect them. The covariance matrix test has produced the values of 783 and 562, respectively; similarly, the invariant values used for the recognition of these sub-landmarks have shown a difference of 3.3% and 4.5%, respectively. Therefore, it is evident that this example verifies the applicability of our method for robust extraction and recognition of landmarks with the sole assumption that the workspace contains planar surfaces.

6 Summary and Conclusions

In this paper, we have presented an approach for automated landmark extraction during an initial learning period and a subsequent recognition capability during navigation tasks. This approach supports topological navigation, utilising permutation insensitive projective invariants and relies only on the existence of planar regions in the workspace. Experimental results have shown that the method facilitates accurate recognition, avoiding confinements posed by geometric methods or exceptionally taxing computations involved in solving point-to-point correspondence problems. We intend to enhance the real time implementation of our method by improving on the search at recognition time, possibly by utilising the values of the invariant as an *indexing* mechanism for the model base. Additionally, we intend to study uncertainty treatment in our implementation, derived by partial or total occlusion of sub-landmarks. To tackle this we intend to introduce confidence measures and sub-landmark expectation during the recognition phase, based on the topological map.

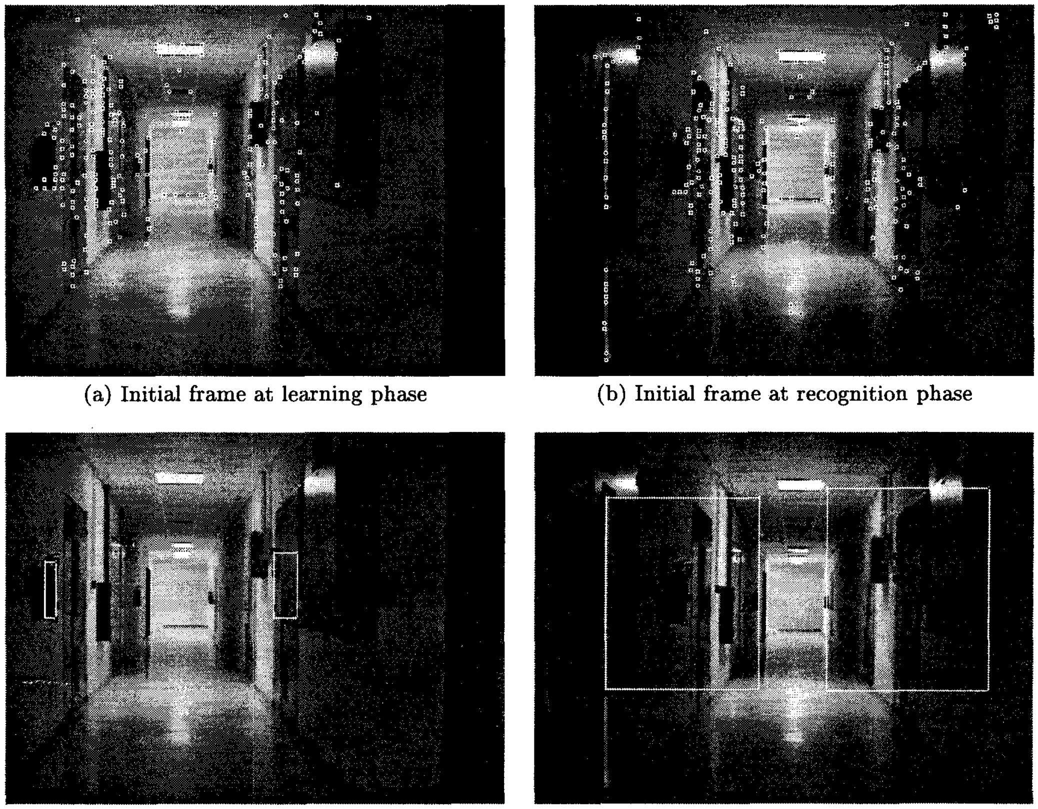
References

[1] S. Shah and J. K. Aggarwal. Modeling structured environments using robot vision. In *1995 Asian Conf. on Computer Vision*, Dec. 1995.

[2] C. J. Taylor and D. J. Kriegman. Vision-based motion planning and exploration algorithms for mobile robots. In *Workshop on the Algorithmic Foundations of Robotics*, 1994.

[3] H. Nasr and B. Bhanu. Landmark recognition for autonomous mobile robots. In *IEEE Intl. Conf. on Robotics and Autom.*, pages 1218–1223, 1988.

Learning Phase **Recognition Phase**



(a) Initial frame at learning phase

(b) Initial frame at recognition phase

(c) Sub-landmarks extracted and stored

(d) Sub-landmarks rediscovered and recognized

Figure 2: Landmark extraction and recognition in a real workspace.

[4] E. Yeh and D. J. Kriegman. Toward selecting and recognizing natural landmarks. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS'95)*, Pittsburgh, PA, Aug. 1995.

[5] A. Kosaka and J. Pan. Purdue experiments in model-based vision for hallway navigation. In *Workshop on Vision for Robots in IROS'95*, pages 87–96, Aug. 1995.

[6] P. E. Trahanias, S. Velissaris, and T. Garavelos. Visual landmark extraction and recognition for autonomous robot navigation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS'97*, Grenoble, France, Sep. 7-11 1997.

[7] Joseph L. Mundy and Andrew Zisserman, editors. *Geometric Invariance in Computer Vision.* The MIT Press, 1992.

[8] C. A. Rothwell et al. Planar object recognition using projective shape representation. *International Journal of Computer Vision*, 16:57–99, 1995.

[9] D. Sinclair and A. Blake. Quantitative planar region detection. *International Journal of Computer Vision*, 18(1):77–91, 1996.

[10] R. Lenz and P. Meer. Projection and permutation invariant object recognition. In *Proceedings 8th Scandinavian Conference on Image Analysis*, pages 389–396, 1993.

[11] S.M. Smith and J.M. Brady. SUSAN - a new approach to low level image processing. *Int. Journal of Computer Vision*, 1997.

[12] J. Clark and N. Ferrier. Attentive Visual Servoing. In A. Yuille A. Blake, editor, *Active Vision*, Artificial Intelligence, chapter 9, pages 137–154. MIT Press, Cambridge, MA, 1993.