Proc. IEEE International Conference on Robotics and Automation, ICRA 2001,
Seoul, Korea, May 21-26, 2001, pp. 2072-2077

Robust Localization Using Context in Omnidirectional Imaging*

Lucas Paletta†, Simone Frintrop††, Joachim Hertzberg††

†Joanneum Research
Institute of Digital Image Processing
Wastiangasse 6, A-8010 Graz, Austria
lucas.paletta@joanneum.ac.at

††GMD – National Research Center for Information Technology
Institute for Autonomous intelligent Systems
Schloss Birlinghoven, D-53754 Sankt Augustin, Germany
frintrop@borneo.gmd.de, joachim.hertzberg@gmd.de

Abstract

This work presents the concept to recover and utilize the visual context in panoramic images. Omnidirectional imaging has become recently an efficient basis for robot navigation. The proposed Bayesian reasoning over local image appearances enables to reject false hypotheses which do not fit the structural constraints in corresponding feature trajectories. The methodology is proved with real image data from an office robot to dramatically increase the localization performance in the presence of severe occlusion effects, particularly in noisy environments, and to recover rotational information on the fly.

1 Introduction

Navigation is a fundamental task for autonomous mobile robots [9]. Particularly, it relies on robust localization methods, e.g. in the presence of occlusion and under noisy and uncertain environment conditions. The interpretation of a single sensor pattern always depends on the ambiguity of a specific view, the imprecision of the environment model, and the uncertainty in the image formation [4, 19]. Instead of generating arbitrarily complex classifiers to solve an a priori ill-posed problem, one may analyze the *context* within the sensor pattern which provides much more efficient cues for localization.

Localization methods are discriminated as *geometric* or *topological*. Geometric approaches require a comparison between the local geometry model extrac- ted from the sensor information with the predetermined model of the entire environment. Typical geometry models are occupancy grids for sonar or laser based localization [15, 21], and models based on local features for visually guided navigation [2]. These models either lack complexity of information or are difficult to be built. In a topological approach, maps are represented by adjacency graphs, with nodes denoting locations and arcs representing transitions between them [22, 23]. Memory-based methods [6, 1, 7, 12, 13] for localization are biologically inspired and do not need an explicit model of the world, but rather match the sensory input with the stored information in memory.

Omnidirectional imaging has become recently an efficient basis for robot navigation [1, 7, 22, 23, 8]. The sensor characteristics enable to provide information not only about a restricted field of view ahead but rather to integrate multi-directional appearances where nearby positions are strongly correlated [12, 8]. Recent advances in robust omnidirectional navigation have been achieved towards illumination invariance [23] and reasonable occlusion tolerance [8], respectively.

The original contribution of this work is to take advantage of the local context in omnidirectional sensing, by reasoning on the structure of ordered 'unidirectional' views in the panoramic image. One may understand panoramic images at certain environment positions in analogy to multi-view captures in object recognition. The proposed method for robot localization thus corresponds to the interpretation of probabilistic dependencies between local structures in 2D [20] and multiple views of 3D [19] objects, respectively. The probabilistic framework enables to take advantage of even partial evidence about a position and to reject false hypotheses which do not fit the structural constraints. These characteristics render it fairly insensitive to severe occlusion and noise in the visual input. Furthermore, as a consequence of the multi-view representation, the robot's orientation is implicitly recovered by localization.

*This work is funded by the European Commission's TMR project VIRGO under grant number ERBFMRXCT960049, and the Austrian Competence Center for Vision Technologies ACV. This support and cooperation is gratefully acknowledged.

Figure 1. Robot *kurt-2* equipped with omnidirectional camera on top.

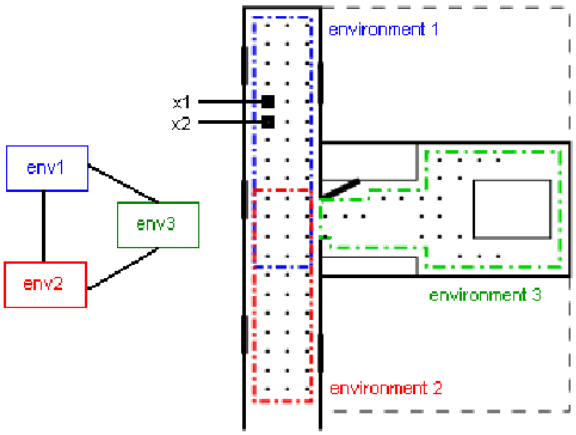

Figure 2. Topological map of the office environment (left) and grid for position measurements (right) in the corresponding navigation modules env1-env3.

Each panoramic image is first partitioned into a fixed number of overlapping uni-directional camera views, i.e., appearance sectors. Distributions of sector images in eigenspace [16, 1] are then represented by mixture of Gaussians [3] to provide a posterior distribution over potential locations. The ambiguity in a local sector interpretation as well as occlusion defects are then resolved by Bayesian reasoning over the spatial context of the current position. Probabilistic navigation modules for localization in small environments are being considered integrated within a global topological navigation concept.

2 Localization by panoramic eigenimages

Complex environments are favourably modelled by topological representations, with nodes representing locations and arcs denoting adjacency of locations. Fig. 2 illustrates a topological map of a small test environment used in the experiments. Within each map module, locations are directly learned from images captured at robot runs. Image based navigation derives a sensor based model of the environment [6, 12], without the load of additional preprocessing and interpretation. Eigenspace analysis of the reference image memory [16, 11] enables significant reduction of the storage requirements and to focus on statistically relevant features. Panoramic eigenimages [1] result then from principal component analysis on omnidirectional sensor images.

The eigenspace of an image training set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ of size N is a low-dimensional representation of the original image space. It captures the maximum variations in the presented data set whereas distances are a measure of image correlation [16]. Recognition is supported by the property that close points in subspace correspond to similar appearances. The eigenspace is spanned by a selection of the ϵ most prominent eigenvectors $\mathbf{e}_i$ of the covariance matrix on $\mathbf{X}\mathbf{X}^T$. The *projection* of image $\mathbf{x}$ into the eigenspace of dimension ϵ is $\mathbf{y}(\mathbf{x}) = (\mathbf{e}_1, \ldots, \mathbf{e}_\epsilon)^T\mathbf{x}$. For localization, samples $\mathbf{x}$ are mapped into eigenspace to $\mathbf{y}(\mathbf{x})$, and assigned the label of the closest position representative $\mathbf{y}_r$.

The robustness in navigation has been extended dealing with illumination variance [23] and occlusion patterns [8], however, the occlusions as in Fig. 9 had not been treated before.

Furthermore, rotational information had either to be incorporated in the representation or to be recovered in the localization process. Local context in the omnidirectional image enables to recover in parallel the rotational and the positional information.

3 Extraction of local appearance sectors

The original contribution of this work is to exploit the spatial relation between uni-directional views in panoramic images. Due to its wide visual field, occlusion of the entire panoramic view becomes very unlikely. Finding evidence for a position from subwindows of the original image, then reasoning over the

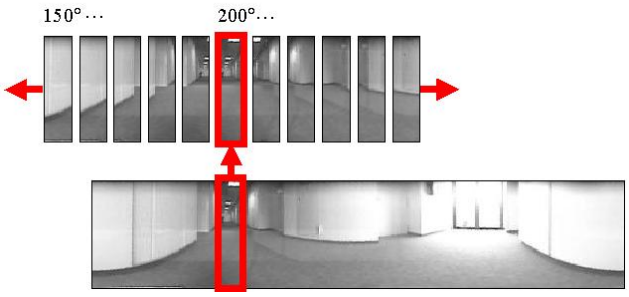

Figure 3. The presented local window approach continually splits the panoramic image (bottom) into overlapping appearance sectors (top).

local decisions, is the method followed in the sequel.

The panoramic image is first partitioned into a fixed number of overlapping uni-directional camera views (Fig. 3), i.e., local *appearance sectors* $\mathbf{s}$ at $\phi(\mathbf{x}_i) = \{\mathbf{s}_{i,1}, \ldots, \mathbf{s}_{i,\Sigma}\}$, for a partition into $j = 1..\Sigma$ sector parametrizations κ_j, starting at field angle $360°/\Sigma$ and with predefined width. Eigenspace analysis is then performed on the set of sectors from images captured at all N individual camera positions, i.e., on $\mathbf{X}' = \{\mathbf{s}_{1,1}, \ldots, \mathbf{s}_{N,\Sigma}\}$. To each position ϕ_i, $i = 1..\Phi$, corresponds a parametric set of points in eigenspace which can be interpolated to a closed curve. Curves of neighboring positions in the environment may overlap due to the similar appearance of their respective sector projections (Fig. 4). Since appearance sectors convey less information than the global image itself, one expects a higher degree of ambiguity in this representation, but a higher robustness to occlusion, too.

4 Probabilistic localization under Bayesian context

In uncertain and noisy environments, localization on the basis of a crisp mapping from observations $\mathbf{y}$ to positions ϕ_i becomes unreliable, and soft computing methods are required to quantify the ambiguity by means of beliefs for multiple position hypotheses [7, 21, 10].

In analogy, [14] introduced probabilistic object recognition from single 2D views represented in eigenspace. [21, 7, 10, 17] applied probabilistic localization to robot navigation problems, [5] outlined local appearance recognition of navigation landmarks.

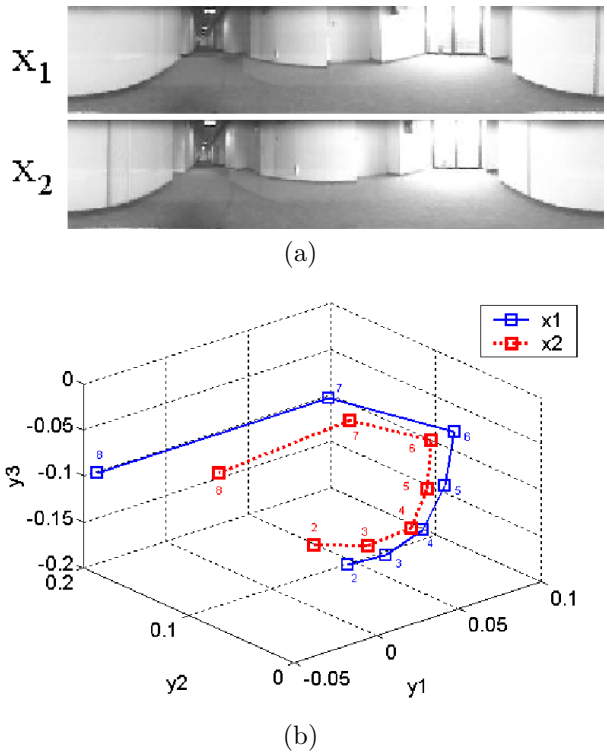

Figure 4. Feature trajectory of sector projections from images $\mathbf{x}_1$, $\mathbf{x}_2$ (Fig. 3) of nearby positions ϕ_1 and ϕ_2, illustrating the ambiguity in their appearance and feature trajectory in eigenspace, respectively.

4.1 Bayesian interpretation

Given the measurement about position ϕ_i and appearance sector κ_j, the likelihood of obtaining feature vector $\mathbf{y}$ is denoted by $p(\mathbf{y}|\phi_i, \kappa_j)$. The likelihood is estimated from a set of sample images with fixed ϕ_i and κ_j, capturing the inaccuracies such as moderate light or positioning variations. From the learned likelihoods one obtains then via Bayesian inversion

$$P(\phi_i, \kappa_j|\mathbf{y}) = p(\mathbf{y}|\phi_i, \kappa_j)P(\kappa_j|\phi_i)P(\phi_i)/p(\mathbf{y}), \quad (1)$$

and a posterior estimate with respect to the position hypotheses ϕ_i is given by $P(\phi_i|\mathbf{y}) = \sum_j P(\phi_i, \kappa_j|\mathbf{y})$.

In practice, the likelihoods are estimated only at selected values of κ, e.g. at most informative or periodic settings $\kappa_j \equiv k \times \Delta\kappa$ [4, 18]. The posterior estimates for intermediate values κ are determined using a Gaussian mixture model [3], $P(\phi_i, \kappa_j|\mathbf{y}) \approx \sum_g P(g)\ \lambda_{\mu_g,\sigma_g}(\mathbf{y})$, where $P(g)$ are the mixing coefficients and $\lambda_{\mu_g,\sigma_g}(\mathbf{y})$ are Gaussians described by parameters μ_g, σ_g which can be adjusted in a training session.

4.2 Bayesian decision fusion process

Multiple measurements enable to resolve ambiguities in the initial interpretation. In a most general framework on belief integration, the discrimination status is iteratively updated with newly upcoming evidence, fusing the posterior beliefs in the position hypotheses $P(\phi_i|\mathbf{y})$ obtained from each single sector measurement $\mathbf{s}$ with new evidence [4, 18, 19].

Spatial context in the combination of sector evidences corresponds in general to the impact of shift actions $a \equiv \Delta\kappa$ (Fig. 9) on the sector feature trajectory in eigenspace (Fig. 4). Introducing the representation of actions into Bayesian decision fusion [19] leads to

$$P(\phi_i, \kappa_j|\mathbf{y}_1, a, \mathbf{y}_2) = \alpha P(\phi_i, \kappa_j|\mathbf{y}_1, a) p(\mathbf{y}_2|\phi_i, \kappa_j, \mathbf{y}_1, a), \quad (2)$$

where $\alpha = 1/p(\mathbf{y}_2|\mathbf{y}_1, a)$ is the normalizing factor. Spatial context is now exploited using the conditional term $P(\phi_i, \kappa_j|\mathbf{y}_1, a)$: The probability for observing appearance sector (ϕ_i, κ_j) as a consequence of deterministic action $a = \Delta\kappa$ must be identical to the probability of having measured at the action's starting point before, i.e. at appearance sector view $(\phi_i, \kappa_j - \Delta\kappa)$, thus

$$P(\phi_i, \kappa_j|\mathbf{y}_1, a) \equiv P(\phi_i, \kappa_j - \Delta\kappa|\mathbf{y}_1). \quad (3)$$

Furthermore, the probability density of $\mathbf{y}_2$, given the knowledge of view (ϕ_i, κ_j), is conditionally independent of previous observations and actions, and therefore $p(\mathbf{y}_2|\phi_i, \kappa_j, \mathbf{y}_1, a) = p(\mathbf{y}_2|\phi_i, \kappa_j)$. The recursive update rule for M *conditionally dependent* observations accordingly becomes

$$P(\phi_i, \kappa_j|\mathbf{y}_1, a_1, \ldots, a_{M-1}, \mathbf{y}_M) = \alpha p(\mathbf{y}_M|\phi_i, \kappa_j) P(\phi_i, \kappa_j - \Delta\kappa_{M-1}|\mathbf{y}_1, a_1, \ldots, \mathbf{y}_{M-1}), \quad (4)$$

and the posterior distribution on position hypotheses, using $\mathbf{Y}_T^a \equiv \{\mathbf{y}_1, a_1, \ldots, a_{M-1}, \mathbf{y}_M\}$, is then given by

$$P(\phi_i|\mathbf{Y}_M^a) = \sum_j P(\phi_i, \kappa_j|\mathbf{Y}_M^a). \quad (5)$$

In practice, occlusions may induce ambiguity in *single* appearance sectors, but they barely represent sector *trajectories* comparable with the training structures. Analogously, localization at inter-grid points and direction measurements can be resolved by extracting structural constraints from fusion steps.

Once the appearance sector $P(\phi_i, \kappa_j)$ is recovered from the positional measurement, the rotation $\Delta\kappa$ between a stored training image $\mathbf{x}_s$ and a test sample $\mathbf{x}_t$ is easily determined by comparing the location of sector $\kappa_{j,s}$ within $\mathbf{x}_s$ with the *corresponding* sector $\kappa_{j,t}$ within $\mathbf{x}_t$.

The computational complexity for a single updating step depends crucially on the sector representation of the panoramic image. For Σ sectors, $\Sigma(\Sigma-1)/2$ sector combinations are computed (each with $\Sigma\Phi$ probability updates) to result in $O(\Phi\Sigma^3)$, which enables real-time processing (section 5).

5 Experiments

The experiments were conducted in the hallway of the GMD-AiS department (Fig. 2) with a Neuronics omnidirectional camera installed on top of the robot platform *kurt-2* (Fig. 1). Images were captured from a grid of sensing points each 50 cm apart, and assigned to the corresponding topological structure depicted in Fig. 2. Note that the robot was aligned to only one direction, since rotational information is implicitly encoded within the sequence of sectors.

The images of size 360×200 were split into 36 sectors of size 20×200, each covering $20°$ of the panoramic field of view, and each within a translational shift of $10°$ (Fig. 3). The images were normalized to $\| \mathbf{x} \| = 1$, to account for illumination variations - according to [16]. Principal component analysis was then applied to images of a corresponding topological module (env1-3) to determine the most prominent eigenvectors (Section 2), and each sector image was projected to the 10-dimensional eigenspace.

To test the navigation system, the training images (Fig 5) were occluded by images of an artifical image database (Fig. 6). The impact of occlusion effects was gradually controlled by the percentage of image content covered by artificial images (e.g., Fig. 9).

The likelihood $p(\mathbf{y}|\phi_i, \kappa_j)$ (Eq. 4) of a test sector sample $\mathbf{y}$, given field view (ϕ_i, κ_j), was modelled assuming Gaussian distributed error, with mean $\boldsymbol{\mu}_{\phi_i,\kappa_j}$ and variance $\sigma^2_{\phi_i,\kappa_j}$ indicating the sector specific sample distribution. This distribution can also be estimated from illumination or positioning variances actually experienced at view (ϕ_i, κ_j) [18, 19].

Occlusion tolerance A first experiment compares the navigation performance of the traditional 'global window' approach - classifying the panoramic image - with the one obtained using 'local window' data - classifying the sector image. Both representations were evaluated by a nearest neighbor (NN) classifier, i.e., test samples $\mathbf{y}_t$ were assigned the label of the training representative $\mathbf{y}_s$ which is closest in eigenspace. Fig. 7 depicts the results w.r.t. recognition accuracy and various degrees of occlusion. While the performance of

Figure 5. Panoramic training images of the GMD-AiS department, particularly from env1 (a), env2 (b), and env3 (c).

Figure 6. Images from the artificial image database to model occlusion effects.

the 'global window' approach vigorously decreases with slight occlusion effects, the recognition of single sectors remains *optimal* until 70 % occlusion. Thereafter, performance decreases (with some perturbation due to the test set configuration) since every sector is affected by severe occlusion.

Robustness against noise The second experiment evaluates the robustness of the interpretation against Gaussian noise in the visual input. A probabilistic interpretation of single sectors (section 4.1) is compared with the localization using Bayesian context (section 4.2) between 2 sectors (Fig. 8). With increasing ratio σ_n^2/σ_l^2, where σ_n^2 denotes the noise and σ_l^2 the learned variance respectively, the advantage of the structural evaluation within the context based approach ('2-sector classification') strictly outperforms a single sector evaluation. '1-sector classification' substantially represents a NN classifier which fails with increasing noise.

Implicit rotational information With the decision for a most likely sector in the training image set, one is able to recover the rotation of the robot w.r.t. the training image. Fig. 9 depicts the test image of 70% occlusion (top) and the recovered training image

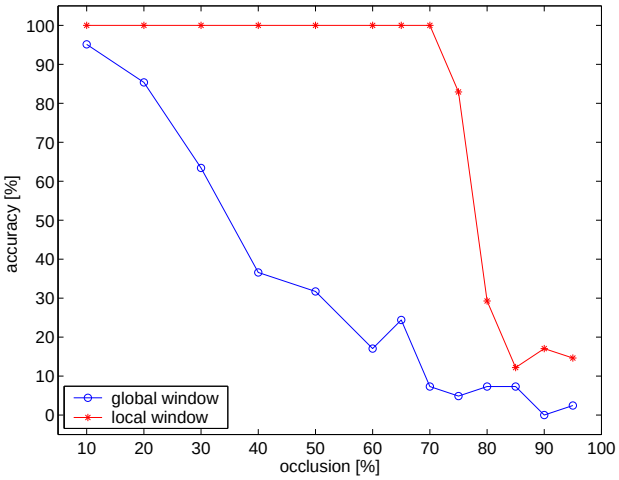

Figure 7. Comparison of occlusion sensitivity between the global and the local window approach.

(bottom). The Bayesian context classifier determined sectors κ_1 and κ_2 as most discriminating sectors, which enables - under knowledge of the corresponding sectors (κ_1',κ_2') in the train image - to recover the rotation $\Delta\kappa$ between these images.

The local sector navigation system enables real-time localization. The Bayesian update procedure for $\Sigma = 36$ sectors and $\Phi = 41$ positions required $\Phi\Sigma^2(\Sigma - 1)/2 \approx 9.3 \times 10^5$ computations, which represented a tractable load for a 300 Mhz PC.

6 Conclusions

Spatial context is an important cue for robot localization which is naturally derived from a sequence of appearance sectors in the panoramic image. The Bayesian framework for the decision fusion process enables to quantify the uncertainty in the discrimination of each position.

The results from navigation experiments using the office robot demonstrate that the Bayesian reasoning allows highly occlusion and noise tolerant localization, vigorously improving the navigation performances in comparison to previous approaches.

The localization system is considered to enable accurate visual navigation of autonomous robots even at crowded places such as offices, factories, and urban environments. The detection of the occluding image regions may then provide a starting point to apply object recognition for further interpretation.

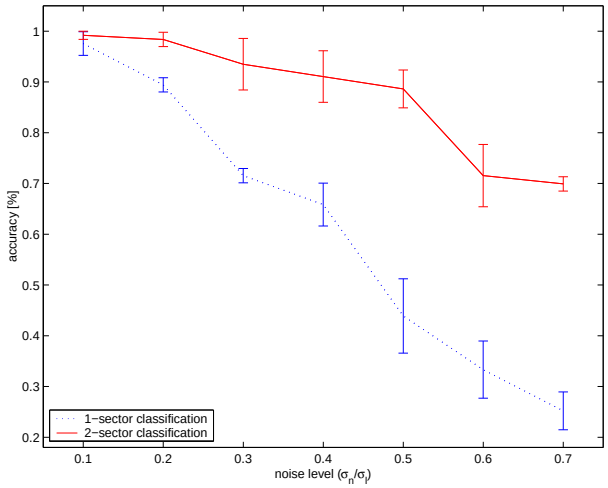

Figure 8. Comparison of noise tolerance between a 1-sector probabilistic interpretation and a 2-sector approach utilizing Bayesian context.

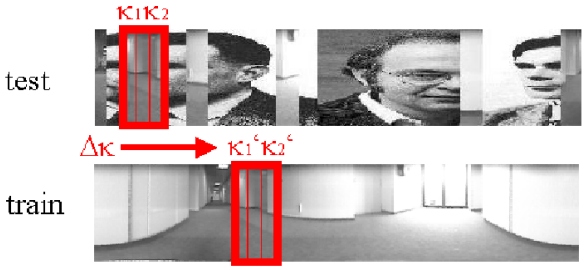

Figure 9. Test image at position ϕ_2 with 70% occlusion (top) and recovered train image (bottom) plus corresponding rotation component $\Delta\kappa$.

References

[1] N. Aihara, H. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localization using omnidirectional images. In *Proc. International Conference on Pattern Recognition*, pages 1799–1803, 1998.

[2] N. Ayache and O. Faugeras. Maintaining representations of the environment of a mobile robot. *IEEE Transactions on Robotics and Automation*, 5(6):804–819, 1989.

[3] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[4] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz. Appearance-based active object recognition. *Image and Vision Computing*, 18(9):715–727, 2000.

[5] V. C. de Verdiére and J. L. Crowley. Local appearance space for recognition of navigation landmarks. In *Proc. Symposium on Intelligent Robotics Systems*, pages 261–269, 1998.

[6] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. Riseman. Image-Based homing. In *Proc. International Conference on Robotics and Automation*, pages 620–625, 1991.

[7] M. Ishikawa, S. Kawashima, and N. Homma. Memory-based location estimation and navigation using Bayesian estimation. In *Proc. International Conference on Neural Information Processing*, pages 112–117, 1998.

[8] M. Jogan and A. Leonardis. Robust localization using panoramic view-based recogniton. In *Proc. International Conference on Pattern Recognition*, pages 136–139, 2000.

[9] D. Kortenkamp, R. P. Bonasso, and R. Murphy, editors. *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. The MIT Press, Cambridge, MA, 1998.

[10] B.J.A. Kroese and R. Bunschoten. Probabilistic localization by appearance models and active vision. In *Proc. International Conference on Robotics and Automation*, pages 2255–2260, 1999.

[11] S. Maeda, Y. Kuno, and Y. Shirai. Active navigation vision based on eigenspace analysis. In *Proc. International Conference on Intelligent Robots and Systems*, pages 1018–1023. Grenoble, France, 1997.

[12] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Visual navigation using omnidirectional view sequence. In *Proc. International Conference on Intelligent Robots and Systems*, pages 317–322, 1999.

[13] R. Moeller, D. Lambrinos, T. Roggendorf, R. Pfeifer, and R. Wehner. Insect strategies of visual homing in mobile robots. In T. Consi and B. Webb, editors, *Biorobotics*. AAAI Press, Menlo Park, CA, 2000.

[14] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[15] H.P. Moravec. Sensor fusion in certainty grids for mobile robots. *AI Magazine*, pages 61–74, 1988.

[16] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.

[17] C.F. Olson. Probabilistic self-localization for mobile robots. *IEEE Transactions on Robotics and Automation*, 16(1):55–66, 2000.

[18] L. Paletta and A. Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1-2):71–86, 2000.

[19] L. Paletta, M. Prantl, and A. Pinz. Learning temporal context in active object recognition using Bayesian analysis. In *Proc. International Conference on Pattern Recognition*, pages 695–699, 2000.

[20] B. Schiele and A. Pentland. Probabilistic object recognition and localization. In *Proc. International Conference on Computer Vision*, 1999.

[21] S. Thrun, W. Burgard, and D. Fox. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning*, 31(1-3):29–53, 1998.

[22] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. International Conference on Robotics and Automation*, pages 1023–1029, 2000.

[23] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor. Omnidirectional vision for robot navigation. In *Proc. IEEE Workshop on Omnidirectional Vision*, 2000.