Vision Based Topological Markov Localization

Jana Košecká and Fayin Li
Department of Computer Science
George Mason University
Fairfax, VA 22030
{kosecka,fli}@cs.gmu.edu

***Abstract*— In this paper we study the problem of acquiring a topological model of indoors environment by means of visual sensing and subsequent localization given the model. The resulting model consists of a set of locations and neighborhood relationships between them. Each location in the model is represented by a collection of representative views and their associated descriptors selected from a temporally sub-sampled video stream captured by a mobile robot during exploration. We compare the recognition performance using global image histograms as well as local scale-invariant features as image descriptors, demonstrate their strengths and weaknesses and show how to model the spatial relationships between individual locations by a Hidden Markov Model. The quality of the acquired model is tested in the localization stage by means of location recognition: given a new view or a sequence of views, the most likely location where that view came from is determined.**

***Index Terms*— Vision based navigation, localization, mobile robots**

I. Introduction and Related Work

The acquisition of unknown environment models, navigation and pose maintenance belong to the essential capabilities of a mobile robots. The approaches for vision-based model acquisition and localization typically strived to obtain either metric or topological models. The topological models were commonly induced by visibility regions associated with the artificial landmarks. Artificial landmarks simplified the issues of landmark recognition and enabled reliable estimation of the robot's pose with respect to a landmark [1], [2]. In other instances the nodes of the topological model corresponded to segments of trajectories where the set of interest points can be successfully tracked [3]. The techniques which tried to bypass the choice of artificial landmarks have been mainly motivated by approaches used for object recognition. One of the main concerns of these methods is the choice of image representation, which could guarantee some amount of invariance with respect to variations in pose, illumination and scale and be robust to partial occlusion and clutter. The image representations proposed in the past comprised of descriptors computed locally at selected image locations or globally over the entire image. The image locations were selected using various saliency measures and their associated rotationally or affine invariant feature descriptors [4], [5], [6] then enabled effective matching of overlapping and possibly widely separated views. Alternative global descriptors were derived from local responses of filters at different orientations and scales [7] or multi-dimensional histograms [8], [9] computed over the entire image. In case of omni-directional views representations in terms of eigenviews obtained by principal component analysis were applied successfully both for topological and metric model acquisition, thanks to small variations of the image appearance within a location and rotationally invariant image representations [10], [11]. The use of local point features for both metric and topological localization was proposed by [12], [13]. In both of these instances the odometric readings were used in connection with the visual estimates.

The problem of building a metric model and simultaneous localization (SLAM) using solely visual sensing has been demonstrated successfully in case of smaller, single room environments [14] or trinocular stereo [15]. The applicability of these purely vision-based methods to the problems of the scale comparable to those achieved by laser range sensors is very difficult due to often ambiguous nature of visual measurements. In order to enable map building and localization solely by means of visual sensing, suitable representations of the environment at different spatial scales and associated means of localization. The advantages of such representations have been pointed out previously by [16] both from the perspective of model building, localization as well as navigation given the model. These types of hybrid models have been already explored previously using ultrasound sensing [17].

In our approach, the final model will be represented in terms of individual locations, each characterized by a set of representative views. Within the location we will endow the model with a local geometry relative to the set of representative views. In this paper we discuss a method for acquiring the coarse structure of the environment in terms of its topology with the localization being solved by means of location recognition. We compare two different representations of locations in terms of image orientation histograms we proposed previously [18] and local scale invariant features. We report the recognition performance using a single view at the time and demonstrate how to exploit the spatial relationships between locations to improve the classification results. The use of spatial relationships is closely related to recently published work by [19] on using contextual information for place and object recognition. Their approach considered slightly different image representation and used hand labelled data set for learning the observation likelihood of individual locations.

This work is supported by NSF grant IIS-0118732.

II. Approach

We propose to represent the large scale structure of the environment in terms of its topology captured by a *location graph*. The nodes of the graph corresponds to individual locations and the transitions represent neighborhood relationships between them. In the presented work we focus on the localization scheme enabled by recognition of locations, which loosely correspond to the regions in the robot's work space which are similar in their appearance. The neighboring locations are typically separated by regions where significant robot navigation decision have to be made; such as hallway intersections, corners and doorways. Initially, the frames of the temporally sub-sampled video sequence obtained in the exploration stage are partitioned and labelled as belonging to different locations. After obtaining a labelled set of views associated with the individual locations, we represent each location in terms of representative feature vectors. In the classification stage we determine given a previously unseen view, what is the location it most likely comes from. Low location likelihoods, which in the presence of thresholds would yield classification errors, are resolved in the second stage by exploiting the temporal context and spatial relationships between neighboring locations modelled in terms of Hidden Markov Model (HMM). We demonstrate the performance of the proposed approach on the model acquisition and localization experiment in indoors environment comprised of 18 locations.

III. Image Features

In order to obtain image representation which captures the essential appearance of the location and is robust to occlusions and changes in image brightness we compare two different image descriptors and their associated distance measure. In the first case we use image histograms integrated over large image subregions and in the second case each image is represented by a set of local scale-invariant features.

A. Image Histograms

The gradient orientation histograms are obtained by first computing the image derivatives $[I_x, I_y]^T = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}]^T$ and assigning orientation to each pixel as atan2(I_y, I_x). The contribution of each pixel to the histogram is weighted by its gradient magnitude $m(x,y) = \sqrt{I_x^2 + I_y^2}$, which has been initially normalized to $[0, 1]$. In order to obtain better discrimination capability of this global representation, we retain some of the spatial information present in the image by computing the histogram for five sub-images (four quadrants and the central region) and stacking them together to form an image descriptor. The most notable characteristic of orientation histogram feature is that it properly reflects the changes in image appearance due to portions of the environment leaving the field of view and reflect presence of corners, doors, and bulletin boards; characteristics which intuitively represent different locations.

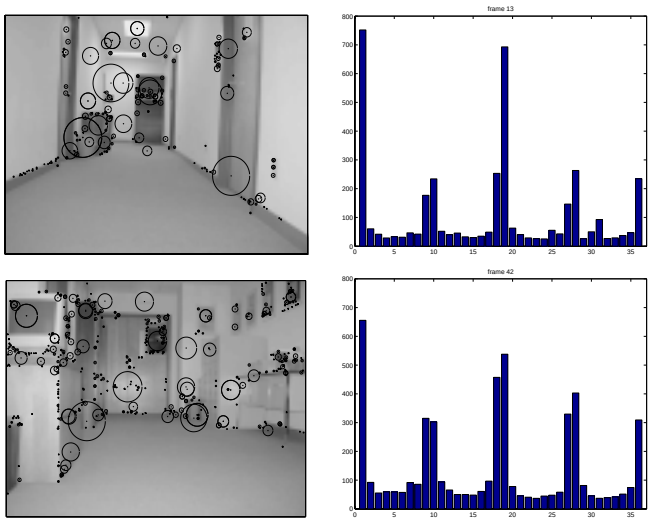

Fig. 1. Locations $l5$ (top) and $l3$ (bottom) of the 4^{th} floor, detected scale invariant features and global gradient orientation histograms (right). The circle center represents the keypoint's location and the radius the keypoint's scale.

B. Scale-Invariant Features

The second descriptor we consider are the scale-invariant (SIFT) features proposed by D. Lowe [20]. The SIFT features correspond to highly distinguishable image locations which can be detected efficiently and have been shown to be stable across wide variations of viewpoint and scale. Such image locations are detected by searching for peaks in the image $D(x, y, \sigma)$ which is obtained by taking a difference of two neighboring images in the scale space

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma). \end{aligned} \quad (1)$$

The image scale space $L(x, y, \sigma)$ is first build by convolving the image with Gaussian kernel with varying σ, such that at particular σ, $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$. Candidate feature locations are obtained by searching for local maxima and minima of $D(x, y, \sigma)$. In the second stage the detected peaks with low contrast or poor localization are discarded. More detailed discussion about enforcing the separation between the features, sampling of the scale space and improvement in feature localization can be found in [20], [21]. The keypoint descriptor is then formed by computing local orientation histograms (with 8 bin resolution) for each element of a 4×4 grid overlayed over 16×16 neighborhood of the point. This yields 128 dimensional feature vector which is normalized to unit length in order to reduce the sensitivity to image contrast and brightness changes in the matching stage. Figure 1 shows the keypoints found in the example images in our environment and their associated global orientation histograms. In our experiments the number of features detected in an image of size 480×640 varies between 10 to 1000. In many instances this relatively low number of keypoints, is due to the fact that in indoor environments many images have small number of textured regions. Note that the detected SIFT features correspond to distinguishable image regions and include both point features as well as regions along line segments.

IV. Environment Model

In the exploration stage the images were taken by a still digital camera about 2 meters apart, with the orientation in the direction of mobile robot heading. The path along which the training sequence was taken visited all locations (some of them twice) and is depicted in Figure 2. In this data set the heading direction was in most cases aligned with the principal directions of the world coordinate frame or perpendicular to it. Along the exploration route the consecutive orientation his-

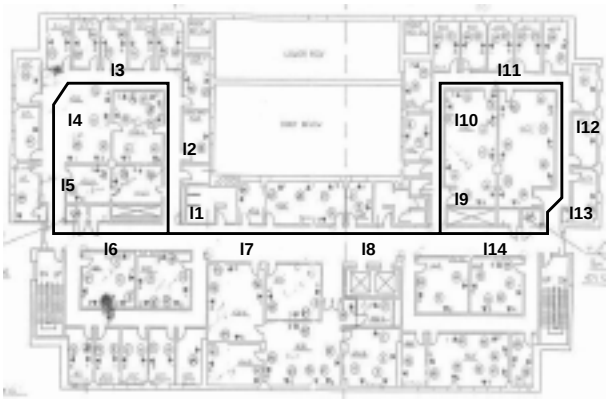

Fig. 2. Floor plan of the 4^{th} floor; exploration route and labels associated with individual locations labelled by hand.

tograms were compared using χ^2 empirical distance measure between two distributions

$$\chi^2(h_i, h_j) = \sum_k \frac{(h_i(k) - h_j(k))^2}{h_i(k) + h_j(k)} \qquad (2)$$

where k is the number of histogram bins. In our case an image descriptor was obtained by stacking five magnitude weighted sub-image orientation histograms. The discrimination capability of the orientation histograms is depicted in Figure 3. The affinity matrices depict all pairwise comparisons between the views using $\chi^2(h_i, h_j)$ and the temporal distance profile measure distances between two consecutive views of the sub-sampled image sequence $\chi^2(h_{t-1}, h_t)$. Note

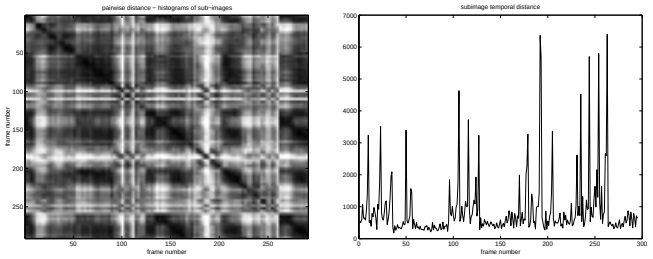

Fig. 3. The pairwise and temporal comparison of orientation histograms of images taken by still digital camera: global histograms image representation (left) and sub-image histograms (right).

that the affinity matrices have clear distinguishing clusters corresponding to the images collected along particular path at locations $l_1, l_2, \ldots, l_N$. The non-diagonal structure of the affinity matrix in Figure 3 also reveals that certain subsequences are similar to each other in spite of the fact that they belong to different locations. This is not surprising since certain locations (e.g. corridors) appear very similar if compared using the orientation histogram descriptor. The clear transitions between the locations are represented by peaks in the temporal histogram comparison plot. These were typically caused by sudden change of robots' heading or more gradual change in the location appearance.

In the case of SIFT features the transitions between individual locations are determined depending on the number of features which can be successfully matched between the successive frames. As long as 4% of features or at least six feature points could be matched successfully between the consecutive views they were assigned as belonging to the same location. More detailed description of the model acquisition stage using SIFT features can be found in [22].

The assignment of individual views to clusters is in our case induced from the temporal relationships acquired during exploration. We have examined two different methods for initial label assignment; automatic and by hand and obtained comparable recognition results. The automatic location label assignment was obtained by searching for the peaks in the temporal histogram distance profile. First coarse peaks were detected and further refined using an adaptive threshold and the minimum separation distance criterion, yielding a set of dominant peaks. Note that in Figures 3 the dominant peaks are quite distinguishable, clearly separating images associated with the individual locations. In the experiments reported in this paper the location labels are assigned by hand due to the fact that the exploration path contains several cycles. These can be resolved by incorporating odometric estimates as a part of the state estimation.

After temporal clustering of the image sequences obtained in the exploration phase, the sequence was partitioned into 18 locations. Due to the rectilinear structure of indoors environments and presence of large number of corridors, the semantics associated with individual locations corresponds to places in the map approached with some canonical orientations coarsely quantized into four different directions (N, W, S, E). Hence being at the same (corridor) location with two opposite orientations corresponds to being at two different locations in our model. Although at this stage this coarse model is sufficient, in order to enable complete metric localization (e.g. within a room), finer quantization of the orientation space is required.

A. Location Representation

Once the initial sequence was partitioned into invidiual locations, we next obtain representation for each location in terms of a smaller number of prototype image descriptors. In case of orientation histograms we have used Learning Vector Quantization technique (LVQ). LVQ examines the data represented as vectors $\mathbf{x}_i \in \mathbb{R}^n$ and in an iterative fashion builds a set of prototype vectors, called *codebook* vectors, that represent different regions in the n-dimensional feature space. We used the existing implementation of LVQ_PAK package [23] with χ^2 statistics in place of the distance func-

Fig. 4. Examples of representative views of 12 out of 18 locations.

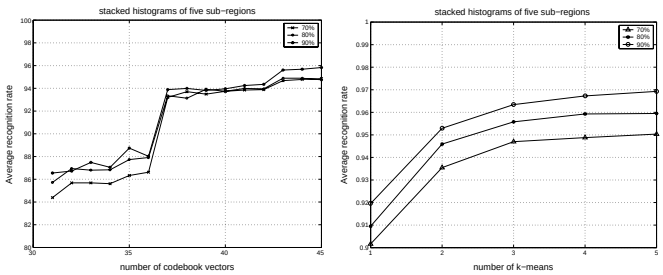

Fig. 5. Recognition rates using nearest neighbor classifier given the representation learned using LVQ method (left) and nearest neighbor classifier given the representation learned using k-means method described above (right).

tion[1]. In the second method we tested, all the views belonging to a particular location were first sampled uniformly, followed by K-means clustering stage. The number of samples varied depending on the location and number of clusters per location varied between 1 to 5.

In case of SIFT feature representation, each location was represented by a number of representative views and their associated SIFT features. The sparsity of the model is directly related to the capability of matching SIFT features in the presence of larger variations in scale. The number of representative views varied between one to four per location and the views were obtained by regular sampling of sub-sequences belonging to individual locations. Examples of representative views associated with individual locations are depicted in Figure 4.

V. Location Recognition

In the first location recognition experiment we have randomly chosen 70%, 80% or 90% of total frames as the training data and the whole sequence is treated as testing data. The recognition experiment was repeated 50 times for K-means and 10 times for representation obtained using LVQ. The recognition rate was recorded each time and averaged over all trials. In both cases we have used nearest neighbor classifier to determine the location which the view came from. The recognition rates of this experiment are in Figures 5 are recorded as a function of total number of prototypes for all locations. The number of prototypes per class depends on differs between locations.

In the case of SIFT keypoints the environment model obtained in the previous section consists of a database of model views [2]. The i-th location in the model, with $i = 1, \dots N$ is represented by n views $I_1^i, \dots, I_n^i$ with $n \in \{1, 2, 3, 4\}$ and each view is represented by a set of SIFT features $\{S_k(I_j^i)\}$, where k is the number of features. In the initial stage we tested the location recognition by using a simple voting scheme. Given a new query image Q and its associated keypoints $\{S_l(Q)\}$ a set of corresponding keypoints between Q and each model view I_j^i, $\{\mathcal{C}(Q, I_j^i)\}$, is first computed. The correspondence is determined by matching each keypoint in $\{S_l(Q)\}$ against the database of $\{S_k(I_j^i)\}$ keypoints and choosing the nearest neighbor based on the Euclidean distance between two descriptors. We only consider point matches with high discrimination capability, whose nearest neighbor is at least 0.6 times closer then the second nearest neighbor. More detailed justification behind the choice of this threshold can be found in [20]. In the subsequent voting scheme we determine the location whose keypoints were most frequently classified as nearest neighbors. The location where the query image Q came from is then determined based on the number of successfully matched points among all model views

$$C(i) = \max_j |\{\mathcal{C}(Q, I_j^i)\}| \text{ and } [l, num] = \max_i C(i)$$

where l is the index of location with maximum number num of matched keypoints. Table I shows the location recognition

sequence (# of views)	NO.1 (250)	NO.2 (134)	NO.3 (130)
one view	84%	46%	44%
two views	97.6%	68%	66%
four views	100%	82%	83%

TABLE I
Recognition rate in % of correctly classified views.

results for SIFT features as a function of number of representative views per location on the training sequence of 250 views and two test sequences of 134 and 130 images each. The two additional test sequences were taken at different days and times of day, exhibiting larger deviations from the path traversed during the training. Despite a large number of representative views per location relatively poor performance on the second and third test sequence was due to several changes in the environment between the training and testing stage. In 5 out of 18 locations several objects were moved or misplaced. Some misclassification examples are shown in Figure 6. Note that in examples a) and b) are the misclassification which occurred using orientation histogram representation. These location are quite similar in their appearance, but can be easily disambiguated using more discriminative image representation such as SIFT features. On the other hand in Figure 6c the location was misclassified due to the dynamic change of the environment between training and testing stage and

[1]In spite of the fact that χ^2 statistics is not a metric (triangle inequality does not hold), we chose to use it as our distance measure due to its good discrimination characteristics [24].

[2]It is our intention to attain a representation of location in terms of views (as opposed to some abstract features) in order to facilitate relative positioning tasks in the subsequent metric localization stage.

a) b) c)

Fig. 6. Examples of test images which were misclassified in the recognition stage: the first row are the test images and the second row are the images which are closest to the nearest neighbor class center. Changes in the appearance of location L_4 and L_6 between the training and testing. In the left image pair the bookshelve was replaced by a table and couch and in the right pair recycling bins were removed.

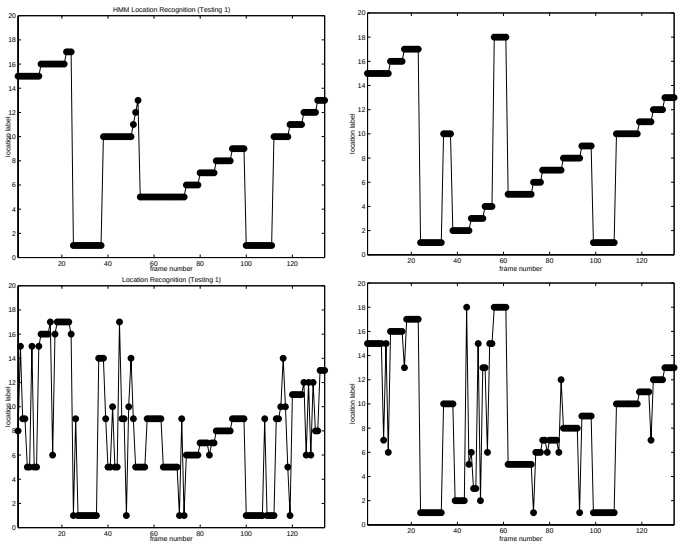

Fig. 7. Test sequence classification results for orientation histograms (left column) and SIFT features (right column) representations. The top row shows the location label assignments for each frame of the test sequence while taking into account the spatial relationships modelled by HMM. The bottom row show the same experiment with HMM turned off.

neither of the two representations could successfully classify this instance. In the next section we demonstrate how can the use of spatial relationships between locations improve the location recognition accuracy, while still retaining these relatively simple image representation.

VI. Markov Localization

The recognition rates reported in the previous section were based solely on the single view and did not exploit the neighborhood relationships between the views. The spatial relationships between individual locations determined by temporal context are modelled by a Hidden Markov Model (HMM). The use of temporal context is motivated by the work of [19] which addresses the place recognition problem in the context of wearable computing application. In our model the states correspond to individual locations and the transition function determines the probability of transition from one state to another. Since the states (locations) cannot be observed directly each location is characterized by its associated observation likelihood $P(L_t = l_i|o_{1:t})$ denoting the conditional probability of being at time t and location l_i given the available observations up to time t. The problem of localization can then be formulated as a problem of estimating most likely location given all available measurements up to time t. The location likelihood can then be estimated recursively using the following formula

$$P(L_t = l_i|o_{1...t}) \propto p(o_t|L_t = l_i)P(L_t = l_i|o_{1:t-1}) \quad (3)$$

where $p(o_t|L_t = l_i)$ is the observation likelihood, characterizing how likely is the observation o_t at time t to come from location l_i .

a) Histogram observation likelihood: In case of orientation histograms, the probability that the observation comes from a particular location $p(o_t|L_t = l_i)$ is obtained by first finding the closest cluster center among all classes based on Bayes rule. The chosen nearest cluster is then approximated with a spherical Gaussian distribution with the cluster center as the mean. The probability of the test image belonging to this cluster center then becomes the probability of the test image belonging to the location. Alternative representation of individual locations in terms of Gaussian mixtures has been proposed in [19]. We have found this soft assignment to be less effective in our environment.

b) SIFT observation likelihood: In the case of SIFT features the conditional probability $p(o_t|L_t = l_i)$ that a query image Q_t at time t characterized by an observation $o_t = \{S_l(Q_t)\}$ came from certain location, is directly related to the cardinality of the correspondence set $C(i)$, normalized by the total number of matched points across all locations

$$p(o_t|L_t = l_i) = \frac{C(i)}{\sum_j C(j)}.$$

The second term of Equation (3) can be further decomposed

$$P(L_t = l_i|o_{1:t-1}) = \sum_j^N A(l_i, l_j)P(L_{t-1} = l_j|o_{1:t-1}) \quad (4)$$

where N is the total number of locations and $A(l_i, l_j) = P(L_t = l_i|L_t = l_j)$ is the probability of two locations being adjacent. All the transition probabilities between individual locations were assigned non-zeros values despite the fact that the transitions between certain locations did not exist. In case of orientation histograms, in the presence of a transition between two locations the corresponding entry was assigned value p_1 and in the absence of the transition it was assigned value p_0. In the final stage all the rows of the matrix were normalized. The performance reported in the following experiments used the ratio of $p_1/p_2 = 1.5$. The ratio of values p_1 and p_0 affected the final recognition rate. In case of SIFT features the presence of a transition between two locations the corresponding entry of A was assigned a unit value and in the final stage all the rows of the matrix were normalized. We have tested the improvements in the recognition rate for both image descriptors on training data and new test sequences. Not

surprisingly in both cases the employment of HMM improved the recognition rate compared to single view recognition. Although the recognition rate for training data we on average 98%, we found the orientation histograms to be inferior to SIFT features on new test sequences. This was primarily due to the larger deviations of the path from the original exploration path and some dynamic changes in the environment. The results of location recognition on new test sequence are in Figure 7. The recognition performance using HMM enabled us to eliminate most of the previous classification errors and achieve classification rate around 99%. Although some of the individual views were misclassified, the order of locations visited during the test sequence was determined correctly by SIFT features Hidden Markov Model in Figure 7 upper right plot. In the case of orientation histograms frames 38 to 55 were misclassified with the use of HMM, yielding 90% recognition rate. Turning the HMM off by making all transitions equally likely decreased the overall recognition rate for both image descriptors.

VII. Conclusions

We have demonstrated an approach for vision-based topological localization by means of place recognition. While in the single view recognition case we have observed several classification errors, those were successfully eliminated using the spatial relationships modelled by Hidden Markov Model. We also compared two different image descriptors, and showed the SIFT features to be superior to orientation histograms due to their higher discrimination capabilities and better invariance properties with respect to viewpoint changes. We are currently in the process of carrying out more extensive experiments and fully automating the model acquisition stage. The presented work only deals with capturing the coarse spatial structure of the indoor environment. In parallel we are developing methods to enabling precise relative positioning within individual locations, using geometric pose estimation techniques. This step is essential for enabling simultaneous model acquisition and localization by means of purely visual sensing without relying on the odometry.

Acknowledgements

We would like to thank D. Lowe for making the code for detection of SIFT features available.

References

[1] A. Briggs, D. Scharstein, and S. Abbott, "Reliable mobile robot navigation from unreliable visual cues." in *In Fourth International Workshop on Algorithmic Foundations of Robotics, New Hampshire*, 2000.

[2] C. J. Taylor and D. Kriegman, "Vision-based motion planning and exploration algorithms for mobile robots," *IEEE Transaction on Robotics and Automation*, vol. 14, no. 3, pp. 417–427, 1998.

[3] G. Bianco, A. Zelinsky, and M. Lehrer, "Visual landmark learning," in *IROS*, Japan, October 2000.

[4] R. Sims and G. Dudek, "Learning envirmental features for pose estimation," *Image and Vision Computing*, vol. 19, no. 11, pp. 733–739, 2001.

[5] A. Pope and D.Lowe, "Probabilistic models of appearance for object recognition," *Internatinal Journal of Computer Vision*, vol. 40, no. 2, pp. 149–167, 2000.

[6] J. Wolf, W. Burgard, and H. Burkhardt, "Robust vision based localization for mobile robots using image-based retrieval system based on invariant features," in *IEEE International Conference on Robotics and Automation*, 2003.

[7] A. Torralba and P. Sinha, "Indoors scene recognition," in *AI Memo 2001-015*, 2001.

[8] B. Schiele and J. L. Crowley, "Object recognition using multidimensional receptive field histograms." *International Journal of Computer Vision*, 2000.

[9] I. Ulrich and I. Nourbakhsh, "Appearance based place recognition for topological localization," in *IEEE Conference on Robotics and Automation*, November 2000, pp. 1023–1029.

[10] M. Artac, M. Jogan, and A. Leonardis, "Mobile robot localization using an incremental eigenspace model," in *IEEE Conference of Robotics and Automation*, 2002, pp. 1025 – 1030.

[11] J. Gaspar, N. Winters, and J. Santos-Victor, "Vision-based navigation and environmental representations with an omnidirectional camera," *IEEE Transactions on Robotics and Automation*, pp. 777–789, December 2000.

[12] M. G. P. Rybski, S. Roumeliotis and N. Papanikopoulous, "Appearance-based minimalistic metric SLAM," in *Intl. Conference on Intelligent Robots and Systems IEEE/RSJ*, October 2003, pp. 194–199.

[13] J.-F. L. O. M. M. G. P. Rybski, F. Zacharias and N. Papanikopoulous, "Using visual features to build topological maps of indoor environments," in *Intl. Conference on Intelligent Robots and Systems IEEE/RSJ*, October 2003, pp. 194–199.

[14] A. Davidson and D. Murrray, "Simultaneous localization and map building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 865–880, 2002.

[15] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale invariant visual landmarks," *International Journal of Robotics Research*, 2002.

[16] B. Kuipers and Y. T. Byun, "A robot exploration and mapping strategy based on a semantic hierarch of spatial representations," *Journal of Robotics and Autonomous Systems*, no. 8, pp. 47–63, 1991.

[17] N. Tomatis, I. Nourbakhsh, and R. Siegwart, "Hybrid simultaneous localization and map building: a natural integration of topological and metric," *Robotics and Autonomous Systems*, 2002.

[18] J. Košecká, L. Zhou, P. Barber, and Z. Duric, "Qualitative image based localization," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[19] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-based vision system for place and object recognition," in *International Conference on Computer Vision*, 2003.

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, p. to appear, 2004.

[21] M. Brown and D. Lowe, "Invariant features from interest point groups," in *In Proceedings of BMVC, Cardiff, Wales*, 2002, pp. 656–665.

[22] Disguised, "Experiments in location recognition using scale-invariant sift features," George Mason University, Tech. Rep. TR A30, 2004.

[23] T. Kohonen, J. Hynninen, J. Kangas, J. Laaksonen, and K. Torkkola, "LVQ_PAK - the learning vector quantization program package," Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland, Tech. Rep. TR A30, 1996.

[24] P. Barber, "Image-based localization for mobile robot navigation," Master's thesis, George Mason University, Department of Computer Science, 2002.