Sparse appearance based modeling for robot localization

Olaf Booij, Zoran Zivkovic and Ben Kröse
Informatics Institute, Faculty of Science
University of Amterdam
1098 SM Amsterdam, The Netherlands
Email: obooij@science.uva.nl

***Abstract*— In appearance based robot localization a new image is matched with every image in the database. In this paper we describe how to reduce the number of images in this database with minimal loss of information and thereby increasing the efficiency of localization significantly. First we build an appearance based model consisting of a graph in which the nodes denote images and links denote relations between images. This graph is then pruned using the Connected Dominating Set algorithm. The method is tested on an image set acquired by a mobile robot equipped with an omnidirectional camera driving around in an office environment, as well as sets of images taken while driving through a real furnished home environment.**

I. INTRODUCTION

To effectively navigate from one place to another a mobile robot needs an internal representation, or model, of its environment. Before it can start to plan a path, the robot must determine its present location in this internal representation using its sensor data, the so called localization task. The method used for localization does not only depend on the type of sensors the robot has, but also on the type of representation of the environment.

There are, roughly stated, two types of techniques of building a model using sensor data. Traditionally robots model their environment using a representation in which various properties of the environment are mapped onto world coordinates. Examples range from occupancy grids or polygon models indicating free space to full 3D CAD models [1][2]. Planning a task in this geometrical model is fairly straightforward, because distances in the model correspond directly to distances in world space. Also, controlling the robot is not difficult because lower level robot-controllers are usually based on metric information. However, building a complete 3D model is tedious and scales badly with environments of increasing size.

Recently the use of appearance based models has become more popular [3][4], partly because cameras are becoming cheap sensors. In appearance based, or view based, modeling, the environment is represented by a set of images, which are sometimes complemented with the position from which the images were taken, for example by wheel encoders. Without this geometric knowledge, the only information from the environment are the images themselves. By using similarities of the images we construct a graph, which we call the appearance graph [5]. The nodes in the appearance graph denote the images, while a link between two nodes indicates that the two images are similar. This graph is then used as a topological map of the environment, which still bears the low level sensor readings. Graph clustering techniques can be applied on the appearance graph to find groups of similar looking images, effectively obtaining the parts of the map that resemble the rooms of the environment [5]. Also path planning can be done efficiently using the structure of the graph [6].

Localizing oneself in such a map implies finding the image, and thus the node in the graph, that resembles the image taken from the current position. Of course if the number of images in the map scales up, so does the time to find the best matching image. Thus, it would be advantageous to reduce the number of images in the database. Especially images that are very similar to other images can be discarded without losing much information. In this paper we describe a pruning technique originating from graph theory, called the Connected Dominating Set problem, which will use the structure of the graph to reduce the image set considerably. In this way we end up with a much smaller amount of images, which still gives a good representation of the environment.

The rest of this paper is organized as follows. Section II starts with an overview of other approaches of reducing the number of features in appearance based models and related work. Then, in Section III we describe how the appearance graph is build using images. Section IV presents the graph theoretical pruning technique, which will be used to reduce the size of the appearance graph. We test the proposed methods on an imageset taken in an office environment and in a real home environment, both obtained by a mobile robot, and report the results in Section V. And finally in Section VI we make some conclusions and give directions for future work.

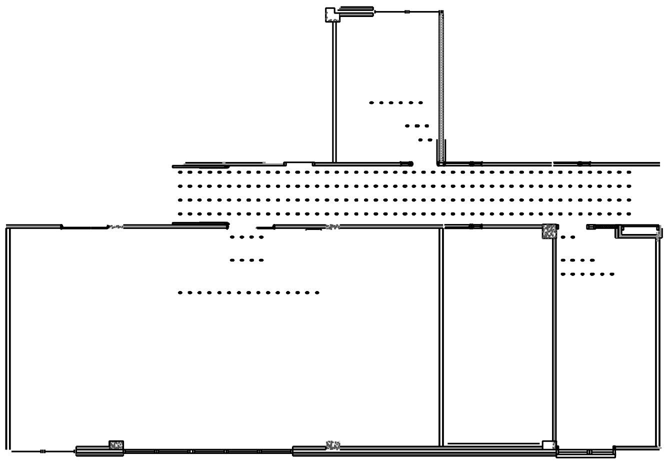
II. RELATED WORK, REDUCING THE NUMBER OF FEATURES IN APPEARANCE BASED SYSTEMS

All applications of appearance based methods have to deal with the problem of selecting which observations to store in its database. This does not only account for robot localization systems, but in general for appearance based vision tasks, such as object recognition [7]. Nevertheless, in most research a crude method is used to create a database of images. A common technique is to make a grid of images, acquiring a number of images proportional to the size of the surface of the environment, as in [8][9]. This is not to be preferred, because the robot has to be placed on the grid points by hand while taking images, rather than letting the robot build a map automatically. Another method is to take one image per unit time, while driving around. In [10] for example images are taken at the rate of 1 Hz. If enough computational resources are available and the environment is small, these simple approaches can be sufficient. In general however the total amount of observed data should be reduced.

In [11] a method is proposed that reduces the set of images using similarities among them. The temporal distance with which the images were taken is used to cluster the images in groups that are approximately from the same location. Then for each group, a number of prototype images is created, which are used for localization. The drawback of grouping images on the basis of temporal distance is that all images taken from the same place but at a different time will be put in a different group, producing probably quite similar prototype images. Indeed one of the assumptions in [11] is that each location is visited only once.

Instead of removing images it is also possible to retain only the most salient image features from the images. In [12] the minimal number of landmarks per region in the environment is calculated so that every image in that region can still be localized. Features that appear in all images of a region are thus regarded as good features. Similar methods do not calculate the minimal number of features, but track features in consecutive images, while calculating their distinctiveness [8][13].

Another technique that is related to our work is that of constructing visibility graphs. The task is to compute the minimal subset of locations in a given geometrical map, so that every point in the map is visible from at least one of these locations [14]. The work presented here is in some way related while we search for the minimal subset of images taken at certain locations, that can be matched with all other images in the map. However in our case the geometrical layout of the environment is unknown. Furthermore the research of visibility graphs is solely based on range detectors, such as sonars, while we use a vision based system.



(a)

(b)

Fig. 1. (a) Bird's eye view of the environment containing a corridor and three rooms. The dots represent the image locations. (b) A matching image pair. The lines indicate matching local features between the two images.

III. Constructing The Appearance Graph

True appearance based approaches only use images as a model of the environment [4][3]. By finding relations among the images we obtain useful information. By matching the image pair-wise we can link similar looking images. These pair-wise relations can be represented as a graph $G = (V, S)$, in which the nodes V denote the images and the links S denote if two images have some similar characteristics. This work was previously presented in [5].

As an example we will build an appearance graph from a set of 234 omnidirectional images taken in an office environment consisting of a corridor and three rooms. In figure 1(a) the positions from where the images are taken is shown. Note that the ground truth positions are used only for presenting the results, they are not used by our methods. For every image a node will be added to the graph, so in this case there will be 234 nodes. For each pair of nodes i, j the value of the link S_{ij} defines the similarity of the nodes to be equal to 1 if and only if it is possible to perform 3D reconstruction of the local space from the two images corresponding to the nodes, as in [15]. We use the standard 3D reconstruction 8-point algorithm [16] and the Scale Invariant Feature Transform (SIFT) features [17] as the automatically detected landmarks in the images (see figure 1(b) for an example of two matching images). If it is not possible to make a 3D reconstruction then there is no link between the nodes and $S_{ij} = 0$. For localization and navigation the robot can use the same 3D reconstruction algorithm. If there is a non-zero link in the graph this also means that if the robot is at one of the connected nodes (corresponding to one image), it can determine the relative location of the other node (corresponding to the other image). If there are no obstacles in between, the robot can directly navigate from one node to the other (e.g. as in [18]). If there are obstacles, one could rely, for example, on an additional reactive algorithm for obstacle avoidance using range sensors. In this way, a graph is constructed from the image set, see figure 2.

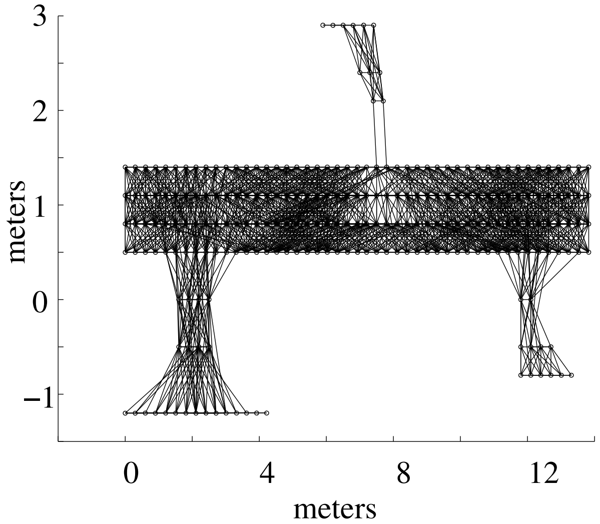


Fig. 2. Graph representing the pair-wise correlation between the images in our database.

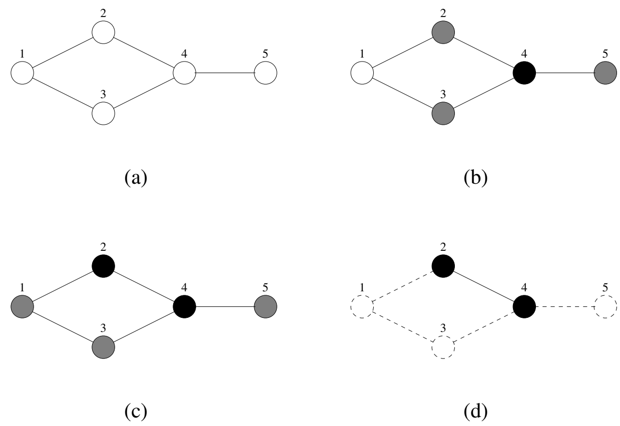


Fig. 3. A simple example describing the approximation algorithm.

Instead of the all-or-nothing method used to construct the graph, i.e. $S_{ij} \in \{0, 1\}$, we could also define a different similarity metric. For example we could give every link a weight that would reflect the quality and robustness of the 3D reconstruction given a pair of images. However this is beyond the scope of this paper.

IV. Creating a sparse graph

Taking a closer look at figure 2 we can see that some parts of the graph are highly connected, especially the part representing the corridor. This means that a lot of images in the corridor are very similar, and removing some of these images is a good idea for making localization more efficient. The problem we have to solve is determining which node of the graph can be removed without causing other nodes to become unreachable. This problem is a known problem in graph theory called the Connected Dominating Graph or Set problem (CDS) and is encountered in a lot of other domains, such as radio broadcasting and computer networking.

A problem when applying the CDS method, is that it tends to break cycles in the graph. This is not to be desired, because a path between two nodes in the original graph can be much smaller than the path between the same two nodes in the Connected Dominating Graph.

In this Section we will give an exact definition of CDS and describe a method to find an approximate solution as well as a method to restore big cycles in the graph.

A. Definition

For a connected graph $G = (V, S)$, a Dominating Graph $G' = (V', S')$ is defined as follows. The set of nodes in the Dominating Graph V' is a proper subset of the original set V, such that every node u in the original set V is either in the Dominating Set V' or is neighboring a node in V':

$$\forall u \in V : u \in V' \vee \exists v \in V' : (u, v) \in S \tag{1}$$

For clarity, a CDS poses no restrictions on the set of links S', except that it is a subset of the links in the original graph. In case of a *Connected* Dominating Graph the subgraph G' is connected. The problem now is to find a connected subgraph with the minimal number of nodes. This task is however known to be NP-complete, but fortunately there are some algorithms that can find a good approximation in polynomial time in the number of nodes [19].

Most of these algorithms will first remove links to make a spanning tree with as many leaves as possible and then remove all the leaves resulting in a smaller tree. The CDS-problem however does not imply anything about the number of links and does not have to be a tree. In our case it is best to conserve all links between the remaining nodes in the CDS, because they indicate that there is a path between the two image locations. So after the tree is pruned of his leaves, the original links should be restored where possible.

B. Approximation Algorithm

A couple of approximation algorithms for the CDS problem are given by Guha in [19]. We use one of these algorithms that produces a Connected Dominating Set and can be implemented to use polynomial time in the order of the number of nodes. This iterative algorithm can be explained as follows, see also figure 3:

1) First color every node of the graph white (figure 3(a)).
2) Initially choose a white node with the biggest number of neighbors.
3) Color this node black and color all white neighboring nodes gray (figure 3(b)).
4) Choose a gray node that has the most links leading to white nodes (figure 3(c)).
5) Goto 3 until there is no white node left.
6) The black nodes now compose the Connected Dominating Set (figure 3(d)).

By connecting the found nodes with links that were in the original graph, we found a Connected Dominating Graph.

C. Restoring cycles

The CDS method has the tendency to break cycles in order to reduce the number of nodes. For localization this will not be a problem, but for path planning based on the pruned graph, this will result in suboptimal paths. By adding a few nodes to the CDS after it has been computed big cycles can be restored. The following iterative procedure finds those nodes that broke a cycle of at least c nodes in lenght.

Pick two nodes i, j that are not in the CDS and that were neighbors in the original graph: $\{i, j\} \notin G' \wedge (i, j) \in S$ (note that i could be equal to j). Search for paths on the CDS connecting these two nodes. If there is a neighbor of i and a neighbor of j which are not connected by a path over the CDS smaller than c than add the i and j to the CDS. Continue examining all pairs i, j using this new CDS.

This brute force algorithm can be implemented efficiently in polynomial time of the number of nodes.

V. Experiments

We conduct two experiments to test the feasibility of using the Connected Dominating Graph for appearance based localization. First we apply it on an image set taken by our Nomad robot in an office environment as already described in the previous sections and compare localization on the bases of the full image set and the reduced set. For the second experiment a more realistic dataset is used taken in a real home environment by the Biron robot [20] equipped with our omnidirectional mirror.

A. Office environment experiment

We will now prune the appearance graph constructed in Section III, see figure 2. The high level of connectedness of the graph indicates that the images in the set are highly correlated. This already implies that a number of omnidirectional images could be left out.

We applied the algorithm to approximate the Connected Dominating Graph on the appearance graph. While the environment did not contain any big cycles, no nodes had to be restored. The resulting subgraph is shown in figure 4. For clarity it must be reminded that the CDS algorithm does not use the positional information of the nodes which is used to visualize the graph. The graph is solely based on information in the images themselves.

As can be seen the graph is strongly reduced; 209 images are removed, while only 25 remain. This reduction to 11% of the images, will cause further processing, such as localization to be sped up by a factor 9. The number of images taken from inside the rooms, that are left in the CDS, is relatively higher than the number of images from the corridor, indicating that the image set correctly represents the environment.

To test the quality of the computed set of images we will use it for a localization task. To localize itself in the appearance based map the robot will take a new image and perform the same matching technique as used for the calculation of the correlation between the images, described in Section III. Hence, we can define the existence of a link between two nodes in the original graph, as the ability to match one of the nodes location given that the other node is in the remaining set. It should be obvious that all the original image locations can be localized in the appearance based map again, because the property of the CDS insists that every node is in the remaining set or can be reached by one link from a node in the remaining set.

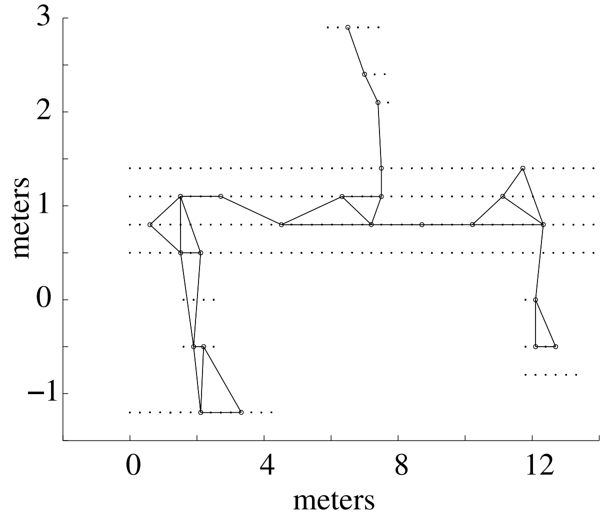


Fig. 4. The pruned graph obtained by applying the CDS-algorithm on the complete correlation graph.

The localization of new images given the pruned graph is simulated with the leave-one-out method in the following manner. First we leave one node with its links out of the original graph, which is regarded as a new image that needs to be localized in the graph. Then a CDS is computed from the remaining nodes. We can now test if the new image can be matched with a remaining image by checking if there is a link in the original graph to one of the nodes in the computed CDS.

We repeated this scheme for all the images in the data set and found that 226 of the 234 (97%) could be localized in the appropriate graph, see figure 5. Also, most of the images that could not be localized in the appearance map were taken at the geometrical border of the robots domain. Of course a robot is not expected to be able to localize itself when positioned outside its map.

B. Home environment experiment

The CDS algorithm is also tested on images taken from a real home environment under less controlled circumstances. The home consisted of 2 rooms connected by a hallway (see figure 6 for a sketch of the house). The robot made two runs through the house while taking images. The first set of images, consisting of 253 images, is used as a training set to build the appearance based representation and the second set, consisting of 300 images, is used to test it. In figure 7 example images are shown. As opposed to the data set taken in the office environment the images taken in the home were not taken

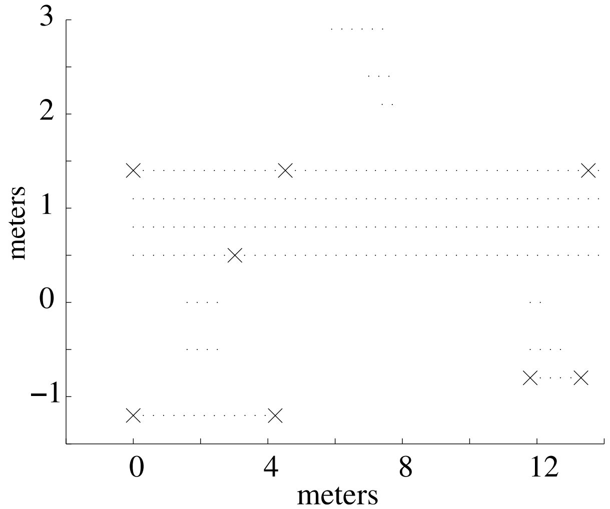


Fig. 5. Result of the test to match new images with a pruned appearance graph. An X indicates an image that could not be matched with the CDS. All other images denoted by the dots can be correlated.

from grid positions, but with a frequency of 1 image per 2 seconds while the robot drove through the environment. The image quality was not as good as the ones taken from the office environment, due to bad lighting. The robot was stopped a few times by a human guide to give the robot new orders, while the camera kept taking images. Therefor, large subsets of the data set contain images that were taken from the same position. It would be advantageous to discard a lot of these images using our pruning technique.

Again an appearance based graph was constructed using similarities in the images and no nodes had to be added to restore cycles. It is difficult to visualize this graph while not having the actual robot positions when taking the images. In figure 8 the connectivity matrix is shown that depicts the graph by showing which images-pairs matched (similar to the visualization used in [11]). As can be seen by the blocks in

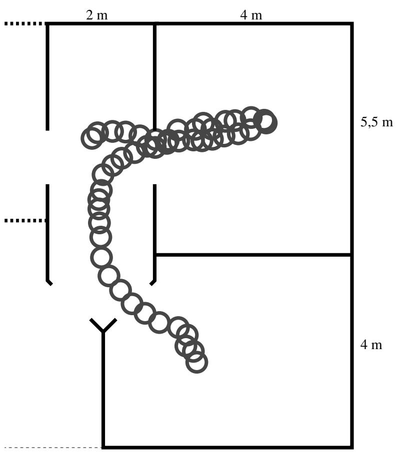


Fig. 6. Sketched map of the home environment and the approximate route of the robot, while taking the training set.

(a)

(b)

(c)

Fig. 7. A few images taken at different locations in the environment. (a) and (c) are from the two rooms, while (b) is from the hallway connecting them.

the figure, there are large subsets of images that look alike. The images that were taken from the hallway (image 1 to 25 and 160 to 180) did not match very well with each other or the images neighboring them in the rooms. This was caused by the lack of features found in the area (see figure 7(b)). Localizing new images from the hallway will be more difficult than images taken in the relatively feature rich rooms.

We again applied the Connected Dominating Graph algorithm on this appearance based graph in order to prune the image set. Only 11 images were preserved, while 242 images were left out. Of these 11 image, 7 were taken from the hallway and 2 from both rooms. The hallway is thus represented with more images than the rooms, which is caused by the lower degree of connectivity due to the lack of features in that area. This is a good result, because now there is a higher change of matching new images taken in the hallway.

To test the quality of the pruned graph with respect to localization, all 300 images from the second run were used. We first tried to localize these images using all images 253 of the original representation. One of the test images in the hallway did not match any image and could thus not be localized. Then we tried to localize the test images using the pruned graph consisting of 11 images. Of the 300 images 33 could not be localized, while 89% was successfully matched. Most of the errornous images were taken from the hallway and some from one of the rooms.

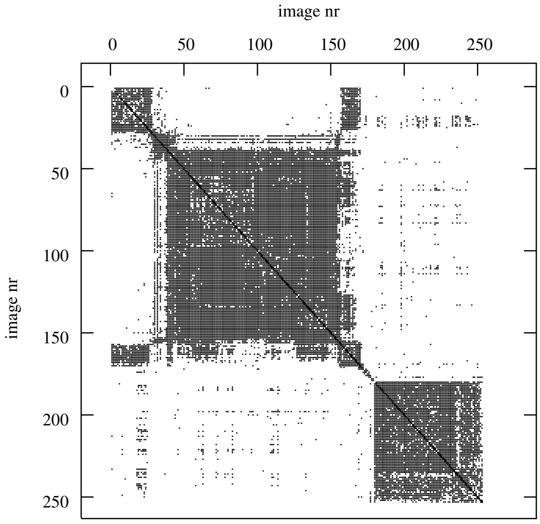


Fig. 8. Visualization of the connectivity matrix. A black dot denotes that the two images match. The large black blocks in the matrix, indicate that large subsets of images look alike.

VI. Conclusion

In this paper we described the application of a Connected Dominating Set approximation to reduce the size of an appearance based model. Using this technique the performance of on-line localization can be improved significantly, by reducing the number of images in the dataset, while trying to preserve the total information in the set. We have successfully applied the method on a set of omnidirectional images taken by our robot. It could however also be used to enhance other appearance based applications, such as object-recognition systems. However it has yet to be shown that the method can also be applied to a set of perspective (not omnidirectional) images.

In the experiment in the real home environment the number of images was reduced to a very small number of images, perhaps too small. It would be better to be able to control this reduction so to keep the number of images in the pruned set at a fixed size. One way of controlling it is by using a different similarity measure that would reflect the robustness of the matches and put this as weights on the edges of the graph. There are variants of the Connected Dominating Graph algorithm that can be applied on such a weighted graph, which have a parameter setting the robustness of the resulting pruned graph.

The proposed algorithm is meant for off-line use, that is: the complete set of images of the whole environment is provided to the algorithm at once. It would be useful to also have an on-line algorithm for rejecting images, in which the image dataset is composed incrementally by an exploring robot. This is considered to be future work.

Acknowledgment

We would like to thank the Applied Computer Science group of Bielefeld University for there work on making the real home environment dataset.

The work described in this paper was conducted within the EU Integrated Project COGNIRON (”The Cognitive Companion”) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

References

[1] J. Borenstein, H. R. Everett, and L. Feng, *Navigating Mobile Robots: Sensors and Techniques*. Wellesley, MA: A. K. Peters, Ltd., 1996.

[2] S. Thrun, “Robotic mapping: A survey,” in *Exploring Artificial Intelligence in the New Millenium*, G. Lakemeyer and B. Nebel, Eds. Morgan Kaufmann, 2002, pp. 1–35.

[3] B. Kröse, R. Bunschoten, S. Hagen, B. Terwijn, and N. Vlassis, “Household robots look and learn,” *Robotics and Automation Magazine*, vol. 11, no. 4, pp. 45– 52, December 2004.

[4] S. D. Jones and J. L. Crowley, “Appearance based processes for visual navigation,” *IEEE International Conference on Intelligent Robots and Systems, France*, 1997.

[5] Z. Zivkovic, B. Bakker, and B. Kröse, “Hierarchical map building using visual landmarks and geometric constraints,” in *Intl. Conf. on Intelligent Robotics and Systems*. Edmundton, Canada: IEEE/JRS, August 2005.

[6] B. Bakker, Z. Zivkovic, and B. Krose, “Hierarchical dynamic programming for robot path planning,” *Technical Report, University of Amsterdam (Submitted to IROS 2005)*, 2005.

[7] G. Dudek and D. Jugessur, “Robust place recognition using local appearance based methods,” in *IEEE Conf. on Robotics and Automation*, San Francisco, April 2000, pp. 1030–1035.

[8] R. Sim and G. Dudek, “Learning and evaluating visual features for pose estimation,” *In Proc. International Conference Computer Vision*, 1999.

[9] M. Jogan and A. Leonardis, “Robust localization using an omnidirectional appearance-based subspace model of environment,” *Robotics and Autonomous Systems, Elsevier Science*, vol. 45, no. 1, pp. 51—72, 2003.

[10] I. Ulrich and I. Nourbakhsh, “Appearance-based place recognition for topological localization,” in *Proceedings of ICRA 2000*, vol. 2, April 2000, pp. 1023 – 1029.

[11] J. Kosecká, L. Zhou, P. Barber, and Z. Duric, “Qualitative image based localization in indoors environments.” in *CVPR (2)*. IEEE Computer Society, 2003, pp. 3–10.

[12] P. L. Sala, R. Sim, A. Shokoufandeh, and S. J. Dickinson, “Landmark selection for vision-based navigation,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, IEEE/RSJ. Sendai, Japan: IEEE Press, September 2004, pp. 3131–3138.

[13] S. Se, D. Lowe, and J. Little, “Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks,” *International Journal of Robotics Research*, vol. 8, no. 21, pp. 735–758, 2002.

[14] L. J. Guibas, R. Motwani, and P. Raghavan, “The robot localization problem,” in *Algorithmic Foundations of Robotics, The 1994 Workshop on the Algorithmic Foundations of Robotics, A. K. Peters*, Goldberg, Halperin, Latombe, and Wilson, Eds., 1995. [Online]. Available: citeseer.ist.psu.edu/guibas95robot.html

[15] F. Schaffalitzky and A. Zisserman, “Multi-view matching for unordered image sets, or “How do I organize my holiday snaps?”,” in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, vol. 1. Springer-Verlag, 2002, pp. 414–431.

[16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision, secon edition*. Cambridge University Press, 2003.

[17] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] R. Basri, E. Rivlin, and I. Shimshoni, “Visual homing: Surfing on the epipoles,” *International Journal of Computer Vision*, vol. 33, no. 2, pp. 117–137, 1999.

[19] S. Guha and S. Khuller, “Approximation algorithms for connected dominating sets,” *Algorithmica*, vol. 20, no. 4, April 1998.

[20] J. Fritsch, M. Kleinehagenbrock, A. Haasch, S. Wrede, and G. Sagerer, “A flexible infrastructure for the development of a robot companion with extensible HRI-capabilities,” in *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, Spain, April 2005, pp. 3419–3425.