Appeared in ICVS 2003, Graz, Austria, pp. 460-469, Springer-Verlag, 2003.

A Framework for Robust and Incremental Self-localization of a Mobile Robot

Matjaž Jogan, Matej Artač, Danijel Skočaj, and Aleš Leonardis

Faculty of Computer and Information Science,
University of Ljubljana,
Tržaška 25, Ljubljana, Slovenia,
matjaz.jogan@fri.uni-lj.si,
WWW home page: http://www.lrv.fri.uni-lj.si

Abstract. In this contribution we present a framework for an embodied robotic system that is capable of appearance-based self-localization. Specifically, we concentrate on the issues of robustness, flexibility, and scalability of the system. The framework presented is based on a panoramic eigenspace model of the environment. Its main feature is that it allows for simultaneous localization and map building using an incremental learning algorithm. Further, both the learning and the training processes are designed in a way to achieve robustness and adaptability to changes in the environment.

1 Introduction

With the increase of interest for autonomously navigating mobile robots, which are not any more limited to operate in production halls, but are able to function in the unconstrained environment, several new issues regarding the design and implementation of such machines arose. The need for systems that would be capable of operating in unstructured and dynamically changing environments shifted the focus of robotics research away from the classical artificial intelligence reasoning towards new fields like embodied intelligence, reactive learning, and distributed thinking. The same definition of an intelligent system now means that it is capable of operating in changing conditions, which includes the capability of continuous and unsupervised learning, generalization of knowledge, and robustness against random changes, which can occur in the environment during interaction.

Mobile robot self-localization is an important part of the navigation task (besides defining the goal and planning). Obviously, to estimate the current position, there is a need for a reference model of the environment to which the momentary sensory data has to be compared. The recent decade saw a gradual shift from the systems that use 3D models towards appearance-based approaches. In fact, supported by the findings of psychologists, theories awoke that describe the visual process as a task of recognizing and associating 2D imagery [15, 16, 14, 11]. Instructive collections of work that stresses the relation between the

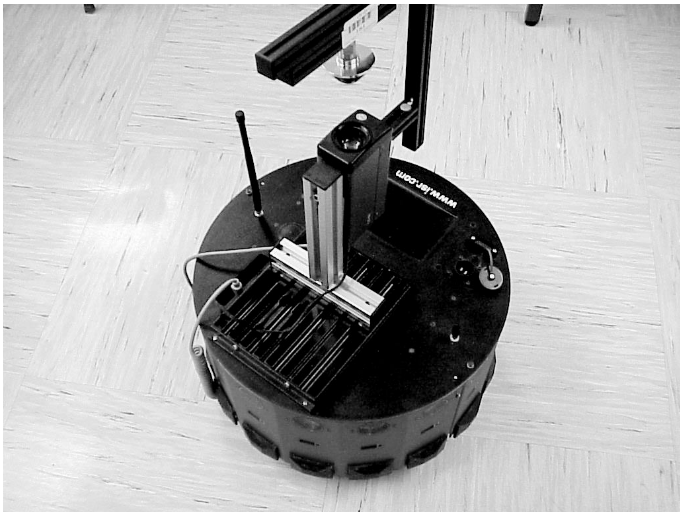

Fig. 1. Mobile robot equipped with a catadioptric panoramic sensor.

3D reconstruction approach and the appearance-based visual recognition can be found in [18] and [1].

In this contribution we present a framework for an embodied robotic system that is capable of appearance-based self-localization. The framework presented is based on a panoramic eigenspace model of the environment. The eigenspace model of appearance is essentially a model of memory that, besides storing compressed imagery, allows also for pictorial retrieval.

Although distributed systems give a solution to limited physical resources, most of the autonomous robots are expected to be embodied, therefore allowing for a limited processing capability and speed. When storing large collections of visual cues, the storage demands can become quickly prohibitive. It is therefore of extreme importance for the model to be optimized in both the terms of descriptive power and compactness. While dealing with these issues, the notions of local vs. global models and modularity arise.

Further, the robot should be capable to use the model in any time, i.e., the classical divide between the learning and the training stage, typical for the appearance-based recognition methods, should be overcome by introducing an open-ended incremental model capable of dynamic updating and splitting into local representations.

Another important issue of an intelligent embodied system is the fashion in which new knowledge is being acquired. As Rolf Pfeifer argues [17], classical AI models suffer from a number of fundamental problems, such as symbol grounding and dependance on the knowledge of the designer, rather than being environmentally conditioned, knowledge being acquired only through interaction with the environment. The robot should be therefore capable of independent exploration without any intervention of the human operator or other displaced intelligent units. As we show, eigenspace recognition methods comply with the latter requirements, being essentially unsupervised learning schemes.

This paper is organized as follows: in Section 2 we give a general overview of the framework for embodied intelligent self-localization, first describing the hardware and the software used and then concentrating on the overall structure of the system. In Section 3 we first describe the learning stage, which is designed to be a robust procedure. We continue with a description of the recognition part of the system. After emphasizing the importance of constant interaction between the learning and localization scenarios, we introduce the on-line methods, that enable us to use a SLAM (Simultaneous Learning and Map Building) approach. Finally, in Section 4, we give an overview of experimental results and end up with a conclusion and an overview of future work.

2 A Framework for Embodied Intelligent Self-Localization

The robot that we use as a testing platform for the appearance-based self-localization is a Magellan Pro with a Pentium II based onboard computer, manufactured by the iRobot company (Figure 1). The software part of the system is written based on the *Mobility* system, which uses CORBA for distributed computing.

Although the robot comes equipped with a variety of sensors, our goal is to enable localization using only a vision sensor. A schema of the system's framework is depicted in Figure 2. The cognitive part of the system consists of models and routines needed to perform simultaneous localization and map building. Although the learning stage and the localization (recognition) stage are depicted as divided entities, with further development of the system, they will presumably become more and more interconnected, finally becoming a unique cognitive model, using common mechanisms and methods.

As the primary source of external information about the environment, panoramic snapshots enter the system and are encoded in the eigenspace model and/or used for localization. To enhance the accuracy of localization, we further use short range readings from the odometry, which helps us to calculate a probabilistic function for the momentary positional estimate.

The software part of the system is implemented such that it allows for distributed computing. However, since the robot is linked to the network using a slow wireless LAN connection, we tend to run all of the data processing on the robotic platform, so that there is no need to transfer the images over the network.

2.1 The Model of the Environment

As we have already mentioned, we use panoramic images acquired from a catadioptric panoramic camera. Panoramic images are becoming more and more popular in the area of computer vision, their primary advantage being a wide field of view. Further, they provide an efficient representation of the surrounding

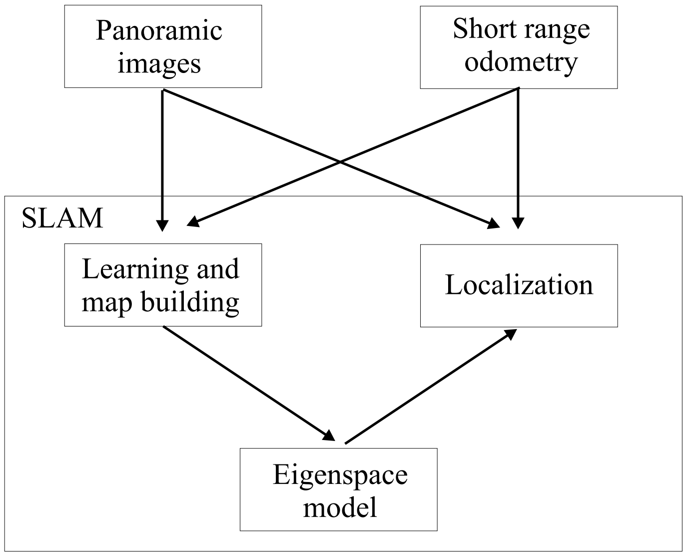

Fig. 2. The framework of the localization system.

and their special properties allow for an efficient encoding in the eigenspace [8, 9, 4]. On the top of the robot we mounted a catadioptric panoramic camera, which was the only visual sensor used. We used two different mirrors in the catadioptric camera. The first mirror was of a spherical shape, while the second mirror had a hyperbolical reflective surface of a $r = 1.9$ ratio.[1]

2.2 Learning and Recognition

The images that are to be learned as depictions of the environment the robot is navigating in are represented in a compressed form using the eigenspace representation [21, 15]. The major advantage of the eigenspace method is that it allows to perform image matching in a much lower-dimensional space. As one can deduce from the experimental evidence from previous work, the covariance between two panoramic images drops with the distance between positions where they were taken. Position could be therefore inferred by finding the image in the training set that shows the largest covariance with the momentary panoramic view. However, it can be shown that when projecting an image onto the eigenspace, the L_2 norm and the covariance produce the same result as when calculated on original images, when using the complete set of eigenvectors. It is therefore possible to match images in a much lower-dimensional space by representing all the images by their projections in the eigenspace and then by adopting the L_2 norm or the Mahalanobis distance between the projections as the distance metric. Furthermore, it is possible to densely interpolate the set of points in the eigenspace to

[1] For a review of catadioptric cameras see [5].

obtain a hyperplane that represents an approximation of an arbitrary dense set of images [15].

The model of the environment is therefore represented by the eigenvectors representing the optimized subspace where the trained images lie in and the representations of positions, which are derived from projections of training images or from interpolation between them.

3 The Cognitive Modules

In this Section we briefly describe the modules that provide the routines needed to perform simultaneous localization and map building. As we have already mentioned, the system consists of two separate modules. The learning module implements the learning routines, which are designed in order to allow for an open-ended, incrementally built model of environment. The localization module implements routines needed for robust recognition of panoramic snapshots and robust localization, supported by a probabilistic computation using short-range odometry data. In the future development of the system, both modules will presumably become more and more interconnected, finally becoming a unique cognitive model, using common mechanisms and methods.

3.1 The Learning Module

The standard approach to eigenspace learning is by the eigen-decomposition of the covariance matrix of the training images. Such a method is susceptible to outliers, occlusions, and varying illumination. However, PCA can be considered as a limiting case of a linear Gaussian model, when the noise is infinitesimally small and equal in all directions. From this observation one can derive an algorithm for calculating principal axes, which is based on the EM (expectation-maximization) algorithm [19]. This algorithm consists of two steps, E and M, which are sequentially and iteratively executed:

– **E-step**: Estimate the coefficients using computed eigenvectors
– **M-step**: Compute the new eigenvectors which maximize expected joint likelihood of the estimated coefficients and the training images

Since the EM algorithm can run on subsets of image pixels, our system implements it in order to obtain a consistent subspace representation in the presence of outlying pixels in the training images. By treating the outlying points as missing pixels, we arrive at a robust PCA representation [20].

Another issue in learning is how to represent images that are taken under different orientation of the robot. Such panoramic images have the same pictorial content, since they represent the same view, yet they are rotated for an angle (phase) of ϕ. To solve this problem, our system implements a specific eigenspace representation, called "*eigenspace of spinning–images*" [7, 9], which achieves insensitivity to the rotation of the sensor by integrating multiple rotated versions of a single panoramic image. The representation exploits the fact that a set of

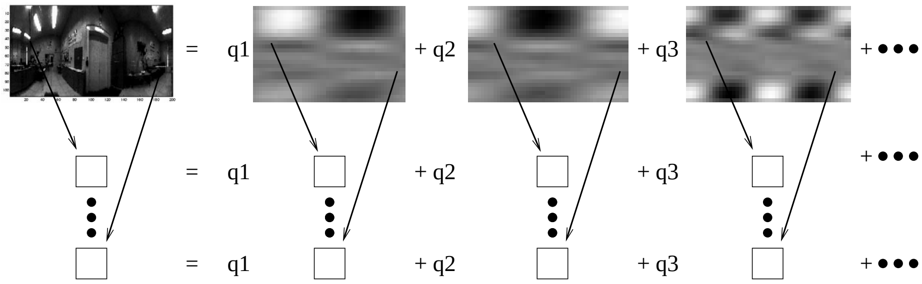

Fig. 3. Calculating the coefficients from a set of linear equations.

rotated templates carrying identical pictorial information can be compressed in the eigenspace in an efficient way. By doing so, we have the advantage of being able to match the incoming image directly to the whole set of rotated images in the recognition phase.

3.2 Adaptable Robust Recognition Module

To localize the robot we have to find the coefficients of the input image and then search for the nearest point on the interpolated hyperplane, which represents the model built on the basis of the images in the training set. The standard method to recover the coefficients is to project the image vector onto the eigenspace [15]. However, this way of calculation of parameters is non-robust[2] and thus not accurate in the case when the input image locally deviates from the image approximated in the environment map.

To overcome the erroneous calculation of image parameters when the visual content deviates from the learning examples, we use the robust approach [11], that, instead of using the image vectors as a whole, generates and evaluates a set of hypotheses $\mathbf{r}$ as subsets of k image points $\mathbf{r} = (\mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_k})$. In fact, the coefficients can be retrieved by solving a set of linear equations on $k = n$ points, where n denotes the dimensionality of the eigenspace.
The principle of such computation is clearly illustrated in Figure 3, where q_i are the image parameters.

By selecting only $p,\ p \leq n$ eigenimages as our basis we have to solve an over-constrained system in a robust way. We solve the system on $k, k > p$ points, where k is significantly smaller than the total number of image points. After the robust solving of the set of equations, we first perform an α–trimming step, in order to allow only the points on which the error is arbitrary small to contribute to the further computation of the parameters. To further increase the probability of avoiding points that are noise or represent occlusion, several different subsets of points are generated, resulting in multiple hypotheses. A hypothesis consists of

[2] Robustness is defined as the extent of the ability of a method to give expected results despite the deviations of the input data.

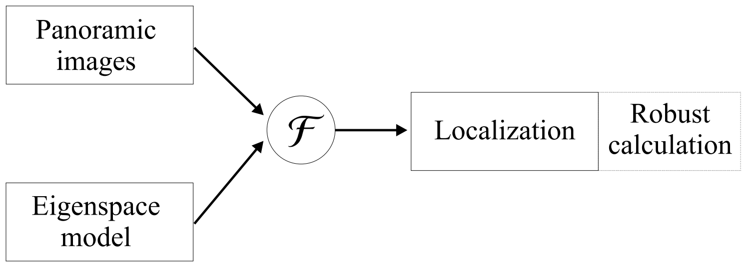

Fig. 4. Schema of the robust localization subsystem.

a set of parameters, an error vector $\boldsymbol{\epsilon}$ calculated as the squared difference between the original image and its reconstruction, and the domain of compatible points that satisfy an error margin constraint. These hypotheses are then subject to a selection procedure, based on the *Minimal Description Length* principle, as described in [11].

Illumination artifacts are another problem that influence the overall robustness of the system. A straightforward approach to the problem of illumination in appearance-based learning and recognition is to learn the appearance under all of the possible light conditions [13]. However, an object in the environment can produce so many different images that it is not clear how to sample all of them.

Instead of this, we employ an extended schema of robust calculation of parameters: by convolving the eigenimages with linear filters, we significantly remove the illumination artifacts [10]. As the eigensystem is a linear equation, both the eigenspace and the input image can be convolved with linear filters without changing the results of computation. The important notion here is that the model itself does not depend on filters at all - filtering can be done just in the process of localization, using an arbitrary number and type of filters, or even using none of them, when the illumination conditions are equal to those encoded in the eigenspace.

The overall schema of the robustified localization system is depicted in Figure 4. $\mathcal{F}$ denotes the bank of linear filters applied. Note that the also robust calculation can be turned on or off, according to the momentary conditions.

3.3 The SLAM Approach

The eigenspace learning and recognition method applied in the standard way has its drawbacks. One of them is the fact that the localization stage is strictly separated from the learning stage. In the learning stage we capture all images first, and only then can we construct the model. The model built in this way can not be modified unless we keep the original images. To update the model with new images, we have to construct a new one from the scratch. Therefore, standard approaches are not optimal for performing simultaneous learning (environment exploration) and localization (SLAM [3]).

To overcome these problems, we developed an incremental method for building the subspace. Incremental computation of eigenvectors has been considered before [6]. However, for a method to be completely on-line, we have to simultaneously update both the eigenvectors and the low-dimensional representations of images. In this way, we can discard the original images immediately after the updating of the subspace. One has to be aware, however, that the low-dimensional representations of the images are only approximations of the originals.

With our approach [2] we are able to perform simultaneous exploration and localization, which means that from the very first moment of the exploration (learning) phase, the robot can use the momentary model of the environment, as it is built incrementally. By collecting new pictorial evidence, the model grows, and by applying a multiple eigenspace growing procedure [12], it can be segmented into logical submodels.

4 Experimental Results

In this section we give the experimental results that show how our self-localization system performs in navigation. In an incremental training phase, the robot explored an indoor environment, storing snapshots at positions that are depicted as empty squares on Figure 5. Then, we positioned the robot on an arbitrary position and sent an order to navigate to a position which was determined by a previously acquired panoramic image. In order to navigate to the goal, the robot performs the following steps: first he estimates his momentary position and orientation, then he estimates the goal's position and determines the vector which points in its directions. He then moves for 70cm in the direction of the goal and again performs localization. According to the new estimate, he recalculates the homing vector and moves for another 70cm.

We present the results as the odometry of the path that the robot followed during his navigation to the goal. In order to demonstrate robustness, we performed tests in both static and dynamic conditions, both with normal and robust techniques.

The leftmost map on Figure 5 illustrates the path of the robot in the case when there are no occlusions in the environment. The filled squares denote the estimated training positions, while the circles denote the positions where the robot stopped to perform self-localization. One can associate the estimated and the actual positions by their numbering. Please note that the information on the odometry accumulated a large amount of error. In truth, the actual ending position differed from the expected one for less than 5cm.

The map in the center illustrates the path of the robot in the presence of significant occlusions and change in illumination. The navigation was in this case performed without employing any of the robust features of the localization system. The rightmost map illustrates, how the robust features of the system improve the performance in harsh conditions.

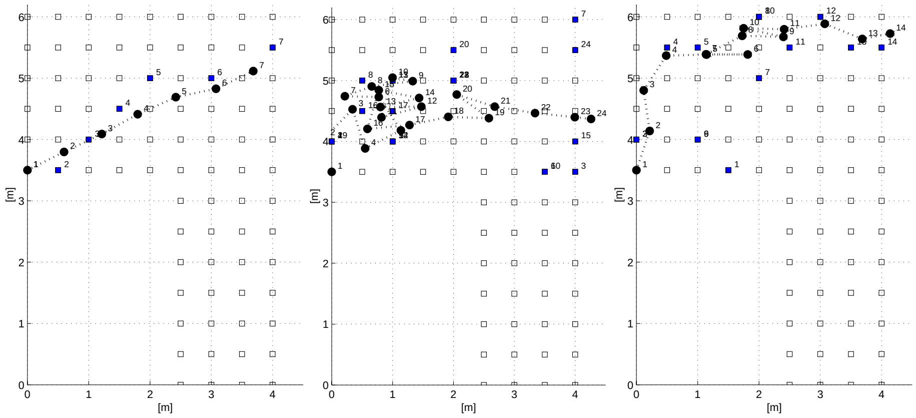

Fig. 5. Localization results in point-to-point navigation.

5 Conclusions

In this contribution we presented a framework for an embodied robotic system that is capable of appearance-based self-localization using a parametric eigenspace model built from panoramic snapshots of the environment.

The eigenspace approach for appearance-based learning and recognition proved itself to be a suitable core for the self-localization system. It provides an efficient representation of the environment, which can be stored in the robot's onboard memory. Further, it allows for a true on-line learning process, enabling simultaneous localization and map building. As we showed in the description of our system, the eigenspace method can be extended with mechanisms that provide robustness both in the learning and in the localization process.

We concluded the paper with a set of experiments which demonstrate the effectiveness of the system. In future work we intend to enhance the accuracy of localization by introducing more knowledge on short range odometry, implement the probabilistic framework and enhance the process of incremental learning with the support for building modular local representations. Further, we are investigating how to build reliable maps with a fully unsupervised procedure in order to achieve a genuine SLAM approach.

References

1. *IEEE Computer Society Workshop on Models versus Exemplars in Computer Vision,* Kauai, Hawaii. IEEE Computer Society, December 2001.
2. Matej Artač, Matjaž Jogan, and Aleš Leonardis. Mobile robot localization using an incremental eigenspace model. In *ICRA 2002*, pages 1025–1030. IEEE Robotics and Automation Society, 2002.
3. A. J. Davison and N. Kita. Sequential localisation and map-building for real-time computer vision and robotics. *Robotics and Autonomous Systems*, 36(4):171–183, 2001.

4. José Gaspar, Niall Winters, and José Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transaction on Robotics and Automation*, 16(6), December 2000.
5. Chris Geyer and Kostas Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, (43):223–243, 2001.
6. P. Hall, D. Marshall, and R. Martin. Incremental eigenanalysis for classification. In *British Machine Vision Conference*, volume 1, pages 286–295, September 1998.
7. Matjaž Jogan and Aleš Leonardis. Robust localization using eigenspace of spinning-images. In *IEEE Workshop on Omnidirectional Vision*, pages 37–44. IEEE Computer Society, June 2000.
8. Matjaž Jogan and Aleš Leonardis. Robust localization using panoramic view-based recognition. In *15th International Conference on Pattern Recognition*, volume 4, pages 136–139. IEEE Computer Society, September 2000.
9. Matjaž Jogan and Aleš Leonardis. Robust localization using an omnidirectional appearance-based subspace model of environment. *Accepted for publication in Robotics and Autonomous Systems*, 2002.
10. Matjaž Jogan, Aleš Leonardis, Horst Wildenauer, and Horst Bischof. Mobile robot localization under varying illumination. In *ICPR 2002*, volume II, pages 741–744. IEEE Computer Society, 2002.
11. Aleš Leonardis and Horst Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding - Special Issue on Robust Statistical Techniques in Image Understanding*, 78(1):99–118, 2000.
12. Aleš Leonardis, Horst Bischof, and Jasna Maver. Multiple eigenspaces. *Pattern Recognition*, 35(11):2613–2627, 2002.
13. Hiroshi Murase and Shree K. Nayar. Illumination planning for object recognition using parametric eigenspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, (12):1219–1227, 1994.
14. Hiroshi Murase and Shree K. Nayar. Detection of 3D objects in cluttered scenes using hirearchical eigenspace. *Pattern recognition letters*, (18):375–384, 1997.
15. S. K. Nayar, S. A. Nene, and H. Murase. Subspace methods for robot vision. *IEEE Trans. on Robotics and Automation*, 12(5):750–758, October 1996.
16. S.K. Nayar and T. Poggio, editors. *Early visual learning*. Oxford University Press, 1996.
17. Rolf Pfeifer and Christian Scheier. *Understanding Intelligence*. MIT Press, Cambridge, 2001.
18. J. Ponce, A. Zisserman, and M. Hebert, editors. *Object Representation in Computer Vision II, ECCV'96 Int. Workshop, Cambridge*, volume 1114 of *Lecture Notes in Computer Science*. Springer, 1996.
19. Sam Roweis. EM algorithms for PCA and SPCA. In *Neural Information Processing Systems 10 (NIPS'97)*, pages 626–632, 1997.
20. Danijel Skočaj, Horst Bischof, and Aleš Leonardis. A robust PCA algorithm for building representations from panoramic images. In *Proc. ECCV02*, volume IV, pages 761–775. Springer, 2002.
21. M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. Computer Vision and Pattern Recognition, CVPR-91*, pages 586–591, 1991.