Proc. Autonomous Mobile Systems (AMS 2009), Karlsruhe, Germany
pp. 153-160, Informatik Aktuell, Springer 2009

An Improved Sensor Model on Appearance Based SLAM

Jens Keßler[1], Alexander König[1], and Horst-Michael Gross[1]

Neuroinformatics and Cognitive Robotics Lab, Ilmenau University of Technology,
98693 Ilmenau, Germany

Abstract. In our previous work on visual, appearance-based localization and mapping, we presented in [14] a novel SLAM approach to build visually labeled topological maps. The essential contribution of this work was an adaptive sensor model, which is estimated online, and a graph matching scheme to evaluate the likelihood of a given topological map. Both methods enable the combination of an appearance-based, visual localization and mapping concept with a Rao-Blackwellized Particle Filter (RBPF) as state estimator to a real-world suitable, online SLAM approach. In this paper we improve our algorithm by using a novel probability driven approximation of the local similarity function (the sensor model) to deal with dynamic changes of the appearance in the operation area. [2]

1 Introduction

Using mobile robots in everyday life, robust map building and self localization plays a central role while navigating the robot in its environment. In the realm of visual SLAM two types of methods are typically used: landmark-based methods and appearance- or view-based approaches. While landmark-based methods require the extraction and reassignment of distinct visual landmarks, appearance-based methods use an description of the view at a certain point, leading to a more global impression of a scene. Appearance-based approaches compare the appearance of the current view with those of the reference images to estimate the robot's pose ([16],[19]).

Feature/Landmark-based approaches: In many SLAM approaches, the map representation is assumed to be a vector of point-like feature positions (landmarks) [18]. The advantage of feature/landmark-based representations for SLAM lies in their compactness. However, they rely on *a priori* knowledge about the structure of the environment to identify and distinguish potential features or landmarks. Furthermore, a data association problem arises from the need to recognize

[2] The research leading to these results has received funding from the European Community's Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under grant agreement n° 216487 (CompanionAble: http://www.companionable.net/)

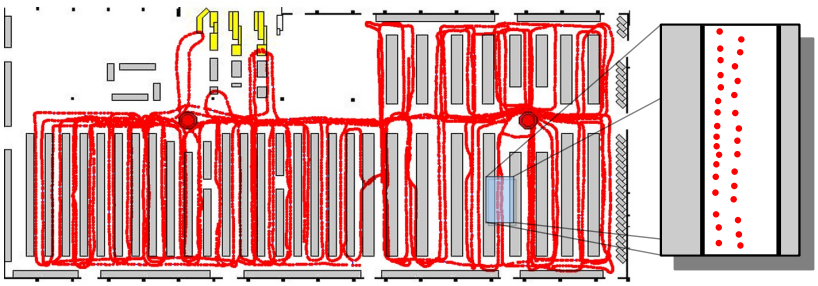

Fig. 1. A manually built map of the operation area, a regularly structured, maze-like home improvement store with a size of $100x50m^2$ (taken from [13]). The dots (nodes of the map) show the positions of stored observations.

the landmarks robustly not only in local vicinities, but also when returning to a position from an extended round-trip. In the field of visual landmark-based SLAM algorithms, Lowe's SIFT-approach [15],[10] has often been used so far. Further important feature/landmark-based approaches are those proposed by Davison using stereo vision [5] or monocular vision [4]. To estimate the landmark positions, popular methods like the Extended Kalmanfilter (EKF) [4] or Rao-Blackwellized Particle Filters (RBPF), [6] like FastSLAM [3], are applied.

Appearance-based SLAM/CML approaches: The Concurrent Map-building and Localization (CML) approach of Porta and Kroese proposed in [17] was one of the first techniques to simultaneously build an appearance-map of the environment and to use this map, still under construction, to improve the localization of the robot. Another way to solve the SLAM-problem was proposed by Andreasson et. al. [1]. Here, a topological map stores nodes with appearance-based features and edges, containing relations between observations and their poses. An essential drawback of this approach is the required offline relaxation phase to correct the nodes' spatial positions by using the found observation matches. The method to estimate the pose difference between images applying the image similarity introduced by Andreasson [1] has been picked up and extended in our SLAM approach. Further approaches that use a topological map representation are described in [2], where a Bayesian inference scheme is used for map building, and in [7], where a fast image collection database is combined with topological maps allowing an online mapping, too.

Contribution of this paper: In our previous approaches ([12],[13]), dealing with an appearance-based Monte Carlo Localization, a static topological model of the environment was developed (see Fig. 1). The nodes of this environment model were labeled with appearance features extracted from an omni directional image. The essential contribution of our approach presented in [14] was the combination of the appearance-based, visual localization concept with a RBPF as state estimator to a visual SLAM approach, to estimate a topological map of the environment. Instead of a single observation, typically used in the field of

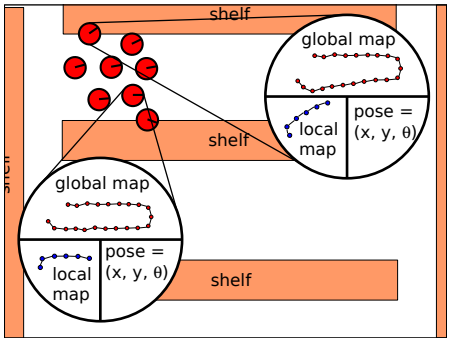

Fig. 2. The map representation of the particles in our approach: Each particle models its current pose estimation, an estimation of the complete global map, and a local map. Some maps are more likely to be correct than others.

appearance-based localization and mapping, another key idea of our approach was to utilize local graphs to perform the evaluation step. These graphs containing the last n observations in n nodes, representing a kind of short-term memory of the very latest observations and pose estimations (see Fig. 2). Another novel idea consisted in online estimating an environment-depending sensor model to evaluate the likelihood of each map. In continuation of this work, we developed a method to further reduce the demand on memory of the environment model and prevent the algorithm from collecting new observations infinitely. Furthermore, we are introducing a method to deal with dynamical changing environments to improve the reliability of the sensor model.

2 Appearance-based SLAM approach with RBPF

In this section the basic idea of our algorithm presented in [14] is explained briefly. Particularly, the graph matching process to determine the likelihood of the map to be correct will be described more precisely. Furthermore, the adaptive sensor model is discussed.

2.1 RBPF with local and global graph models

Our appearance-based SLAM approach utilizes the standard Rao-Blackwellized Particle Filter approach to solve the SLAM problem, where each particle contains a pose estimate $\boldsymbol{x}_i$ (position x, y and heading direction φ) as well as a map estimate (see Fig. 2). The environment model (map) used in our appearance-based approach is a topological graph representation, where each node i, representing a place $\boldsymbol{x}_i$ in the environment, is labeled with appearance-based features $\boldsymbol{z}_i$ of one or more omni directional impressions at that node. To solve the SLAM problem, the RBPF has to determine the likelihood of the graph-based maps in the particles to be correct. Therefore, our approach uses two different types of maps: a *global map* $\boldsymbol{m}^G = \langle \boldsymbol{x}_{1:(l-1)}, \boldsymbol{z}_{1:(l-1)} \rangle$, which represents the already known environment model learned so far and a *local map* $\boldsymbol{m}^L = \langle \boldsymbol{x}_{l:t}, \boldsymbol{z}_{l:t} \rangle$ representing the n latest observations and the local path between them (e.g. the last two meters of the robot's trajectory). To prevent the filter from under-/oversampling of the estimated probability distribution we use the KLD sampling technique (see [11]) to adjust the particle count as needed.

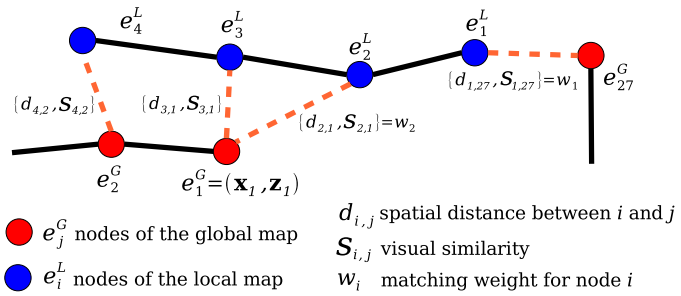

Fig. 3. Basic idea of our map matching algorithm: the likelihood of a given map configuration is determined by using the spatial distances d_{ij} and visual similarities S_{ij} for comparison between each pair of nodes i (in the local graph) and j (in the global graph).

2.2 Graph matching

In the context of RBPF, the probability distribution of the sensor model is determined directly by comparing the local and global map, giving the importance weight $w \approx p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{m})$ of a particle. For this purpose, corresponding pairs of nodes $\langle e_i^L, e_j^G \rangle$ in both maps are selected by a simple nearest neighbor search in the position space, where each node e_i^L of the local map is related to the nearest neighbor node e_j^G of the global map. To keep the computational complexity for the nearest neighbor search as small as possible, we use a quad-tree-like structure for indexing the nodes, so the search does not depend on the total count of nodes. The relation between each selected pair of corresponding nodes $\langle e_i^L, e_j^G \rangle$ provides two pieces of information, a geometric one, the spatial distance d_{ij}, and a visual one, the visual similarity S_{ij} (see Fig. 3). Both aspects are used to determine a matching weight w_i for the respective node e_i^L of the local map. Assuming an independence between the node weights of the local map, the total matching weight $w^{[k]}$ for a particle k is simply calculated as follows:

$$w^{[k]} = \prod_{i=1}^{n} w_i^{[k]} \tag{1}$$

with n describing the number of nodes in the local map. To evaluate the matching weight w_i we have to compute the probability that two observations i and j got a similarity S_{ij} by a given distance d_{ij}.

2.3 Adaptive sensor model

To compute the matching weights between corresponding nodes, an adaptive sensor model had been developed. In the context of appearance-based observations, the visual similarity between observations is not only depending on the difference in the positions but also on the environment itself. If the robot moves, for example over a wide plane, the appearance will not change at all when moving a short distance. While moving along narrow hallways the appearance of the scene will change more significant.

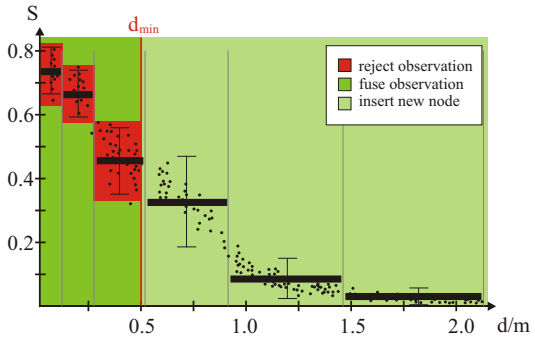

Fig. 4. The histogram model with mean value and σ per bin. The bins are not equally spaced. Points in the light green area ($d_{ij} > d_{min}$) are always included as new nodes. If points in the dark green area ($d_{ij} \leq d_{min}$) are within the model expectation (red area) they need not be included, otherwise they extend the local model by adding a new appearance variant to the node.

Hence, our sensor model estimates online the dependency between environment specific visual similarities $\hat{S}_{ij}$ of the observations $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ and their spatial distance $\hat{d}_{ij}$. Fig. 4 shows an example of such a function. Again the use of a local map is of help here. The samples to built such a model are taken from the nodes of the *local map* where each node is compared the others. We assume here that the resulting statistic represents $S(d)$ in the local map as well as in the global map. In [14] different approaches to approximate the model were investigated, e.g. the Gaussian Process Regression (GPR) of Rasmussen [9], while we used a parametric polynomial description of the sensor model and its variance. In this paper, we use an non-parametric histogram-based sensor model to create a more general model. For that purpose the distance d is divided in bins of different size, and for each bin the mean similarity and its variance are computed (see Fig. 4). The non-linear size of the bins is required to get a higher resolution for small distances, resulting in a higher accuracy of the matching process. The bin index idx is computed by a parametric exponential function $idx(d) = \alpha \cdot (d + \beta)^{\gamma} - 1$. The likelihood that two nodes i and j of particle k are matching the sensor model $S(d)$ is computed as follows, where $\hat{S}$ and $\hat{\sigma}$ are derived from the sensor model:

$$w_i^{[k]} = p\left(S_{ij}|d_{ij}\right) \approx \exp - \frac{\left(S_{ij} - \hat{S}\left(d_{ij}\right)\right)^2}{2 \cdot \hat{\sigma}(d_{ij})^2} \tag{2}$$

For our experiments we decided to use SIFT feature sets as appearance-based image description and similarity measure, because of their ability to re-detect the position of observation with good selectivity. For this specific image description we set the parameters of the bin indexing function to $\alpha = 8.0, \beta = 0.05, \gamma = 0.7$. These parameters are derived empirical.

2.4 Dynamic Changes

To meet the challenges of dynamic environments, we face two main problems. At first, the map grows unlimited while observing the environment and including every estimated position and observation. Second, the visual impression at the same place can change due to different lighting conditions, occlusions and moving objects. So we have to select which of the new positions and observations need to be included into the map. If the distance d to the nearest neighbor exceeds a distance d_{min} a new node with the corresponding observation is added to the map. So we do not add nodes within a circle of d_{min} around existing nodes. We just have to consider which observations have to be added to existing nodes. Hence we extend the nodes of the global map to collect observations of different appearance states by using the similarity model $\hat{S}(d_{ij}), \hat{\sigma}(d_{ij})$ (see Fig. 4) to decide whether new observations have to be included into an existing node or not. For each node of the local map e_i^L we assume to be in a certain position $\boldsymbol{x_i}$ with a global map $\boldsymbol{m}$. We choose the highest similarity S_{ij} between all observations stored in the nearest node e_j^G in the global map $\boldsymbol{m}$ and the observation stored into node e_i^L to evaluate the existing sensor model ($S_{max} = max_k(S_{ij}(k))$). According to equation 2, we can calculate the probability $p\,(S_{max}|d)$. If this probability is above a certain threshold ξ, the observation matches the expected similarity (as we have seen this particular scene in a similar configuration before) and can be ignored. If the observation does not match the model, it is associated to the current node e_j^G in the global map.

3 Experiments and results

To evaluate the extensions of our approach, we used two alternative test environments with specific characteristics, the home improvement store shown in Fig. 1 with large straight hallways, and a small home-like section of our lab with narrow rooms with little space to navigate. All data for the analysis were recorded under realistic conditions, i.e. people walking through the operation area, shelves were rearranged, and other dynamic changes (e.g. illumination, occlusion) happened. In addition, laser data were captured to generate a reference map to evaluate the results of our approach.

The resulting graph of the store (Fig. 5, left) covers an area of 120 x 50 meters and was generated by a mean value of 250 particles (max. 2000 particles) in the RBPF. Figure 5 only shows the most likely final trajectory and a superimposed occupancy map for visualization. The home like area (see Fig. 5, right) was much smaller (10 m x 10 m). To evaluate the visual estimated path shown as trajectories in Fig. 5 a ground truth path and map built by means of a Laser-SLAM algorithm were calculated (GMapping of G. Grisetti [8] taken from www.openslam.org). The Laser-SLAM estimated reference path was used to determine the mean and the variance of the position error of our approach, shown in table 1. These experimental results demonstrate, that our approach is able to create a consistent trajectory and based on this, a consistent graph representation, too. Furthermore, in contrast to grid map approaches (up to 4GB

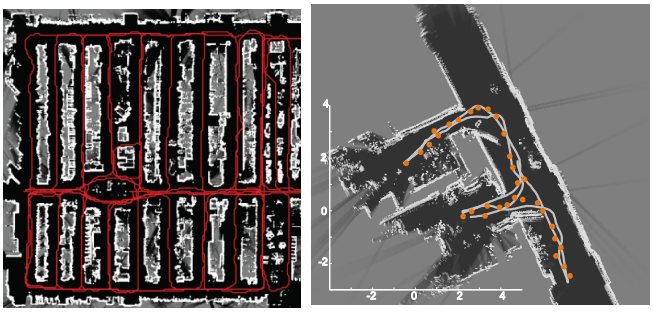

Fig. 5. In the home store: The red path shows the robot's movement trajectory only estimated by our appearance-based SLAM approach. In the home lab: additionally the nodes of the resulting graph are shown. For visualization of the localization accuracy, a laser-based occupancy map was built in parallel and superimposed to the estimated movement trajectory (visual SLAM).

Table 1. Overview of the achieved results in all environments.

Experiment	**Home store**	**Home lab**
Size of area	50x120m	10x10m
Total path length	2400m	20m
# of particles	250 (mean) *2000* (max)	50 (mean) *150* (max)
Error Mean/Var/Max	**0.53**/0.21/1.76 m	**0.21**/0.14/0.42 m
Time per cycle	0.250 s	0.250 s

for our home store environment), topological maps require less memory (up to 1.5GB for 4800 observations) because of the efficient observation storage. The observation's features are stored in a central data set and are only linked to the nodes in all maps and only the positions have to be stored in each node. So the memory requirements are nearly independent from the used number of particles. The results show that in small environments, like the home-like lab, where the robot is able to move along smaller loops, the visual SLAM approach achieved a higher accuracy than in large loops. We are able to demonstrate that our approach not only works in large scale environments but also in narrow home environments.

4 Conclusion and future work

In this paper we presented extensions of our appearance-based SLAM approach to limit the memory consumption and to deal with dynamic changes that occur in common real-world environments. These improvements allow an appearance-based on-line SLAM in dynamic real-world environments at long term time windows. For the near future we plan to implement a visual scan-matching technique to limit the number of required particles to close loops correctly and to further

increase the accuracy of estimated maps. Furthermore, we want to study the influence of dynamic changes in the home environment in more detail.

References

1. H. Andreasson, T. Duckett, and A. Lilienthal. Mini-slam: Minimalistic visual slam in large-scale environments based on a new interpretation of image similarity. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 4096–4101, 2007.
2. A.Ranganathan, E. Menegatti, and F. Dellaert. Bayesian inference in the space of topological maps. In *IEEE Trans. on Robotics, vol.22, no.1*, pages 92–107, 2006.
3. T D. Barfoot. Online visual motion estimation using fastslam with sift features. In *Proc. of 2005 IEEE/RSJ IROS*, pages 579–585, 2005.
4. A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. Int. Conf. on Computer Vision(ICCV'03)*, pages 1403–1410, 2003.
5. A. Davison and D. Murray. Simultaneous localization and map-building using active vision. *IEEE Trans. on PAMI*, 24(7):865–880, 2002.
6. P. Elinas, R. Sim, and J. J. Little. σSLAM: Stereo vision SLAM using the Rao-Blackwellised particle filter and a novel mixture proposal distribution. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 1564–1570, 2006.
7. F. Fraundorfer et al. Topological mapping, localization and navigation using image collections. In *Proc. IEEE/RSJ IROS*, pages 3872–3877, 2007.
8. G. Grisetti et al. Improved techniques for grid mapping with rao-blackwellized particle filters. In *IEEE Transactions on Robotics*, pages 34–46, 2006.
9. Rasmussen et al. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
10. S. Se et al. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int.Journal of Robotics Research*, 21(8):735–758, 2002.
11. D. Fox. Kld-sampling: Adaptive particle filters. In *Advances in Neural Information Processing Systems 14*. MIT Press, 2001.
12. H.-M. Gross, A. Koenig, H.-J. Boehme, and Chr. Schroeter. Vision-based monte carlo self-localization for a mobile service robot acting as shopping assistant in a home store. In *Proc. IEEE/RSJ IROS*, pages 256–262, 2002.
13. H.-M. Gross, A. Koenig, Chr. Schroeter, and H.-J. Boehme. Omnivision-based probabilistic self-localization for a mobile shopping assistant continued. In *Proc. IEEE/RSJ IROS*, pages 1505–1511, 2003.
14. A. Koenig, J. Kessler, and H.-M. Gross. A graph matching technique for an appearance-based, visual slam-approach using rao-blackwellized particle filters. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1576–1581, 2008.
15. D.G. Lowe. Object recognition from local scale-invariant features. In *Proc. Int. Conf. on Computer Vision ICCV'99*, pages 1150–1157, 1999.
16. E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro. Hierarchical Image-based Localisation for Mobile Robots with Monte-Carlo Localisation. In *Proc. 1st European Conf. on Mobile Robots (ECMR'03)*, pages 13–20, 2003.
17. J. M. Porta and B. J.A. Kroese. Appearance-based Concurrent Map Building and Localization using a Multi-Hypotheses Tracker. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'04)*, pages 3424–3429, 2004.
18. R. Smith, M. Self, , and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *Robotics Research, 4th Int. Symposium*, pages 467–474, 1988.
19. I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. IEEE ICRA*, pages 1023–1029, 2000.