2008 IEEE/RSJ International Conference on Intelligent Robots and Systems
Acropolis Convention Center
Nice, France, Sept, 22-26, 2008

# A Graph Matching Technique for an Appearance-based, visual SLAM-Approach using Rao-Blackwellized Particle Filters

Alexander Koenig, Jens Kessler and Horst-Michael Gross

*Abstract*— **In continuation of our previous work on visual, appearance-based localization in manually built maps [1], [2], in this paper we present a novel appearance-based, visual SLAM approach. The essential contribution of this work is, 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 concept with a Rao-Blackwellized Particle Filter (RBPF) as state estimator to a real-world suitable, online SLAM approach. In our system, each RBPF particle incrementally constructs its own graph-based environment model which is labeled with visual appearance features (extracted from panoramic 360$^o$ snapshots of the environment) and the estimated poses of the places where the snapshots were captured. The essential advantages of this appearance-based SLAM approach are its low memory and computing-time requirements. Therefore, the algorithm is able to perform in real-time. Finally, we present the results of SLAM experiments in two challenging environments that investigate the stability and localization accuracy of this SLAM technique.**

## I. INTRODUCTION

Robust self-localization plays a central role in our long-term research project PERSES (PERsonal SErvice System) which aims to develop an interactive mobile shopping assistant that can autonomously guide its user within a home improvement store [3] (see Figure 1). In everyday life and in mobile robotics, two main types of self-localization methods are typically used: landmark-based methods and appearance- or view-based approaches. The reason why people frequently and principally use natural landmarks (e.g. complete objects like buildings, doors, or salient parts of objects) in recognizing places is that people can segment the landmarks rapidly and accurately from scene images. On the other hand, people do not have strong ability in recognizing a large amount of valuable features of scene images in parallel, such as texture, color, edges, shapes and their global appearance features. Unlike people, computer-vision systems don't have problems to recognize location images and perform the function of place recognition using views or extracted appearance features only. That makes the appearance-based methods so interesting for robust localization and map building in mobile robotics. Instead by landmarks, appearance-based approaches only compare the appearance of the current view with those of the reference images to estimate the robot's pose ([15], [17]).

Fig. 1. Experimental platform - the interactive mobile shopping assistant TOOMAS based on a SCITOS A5 (by MetraLabs Ilmenau, Germany) during an interactive guided tour in a home store. The SonyRPU camera at the top of the robot head yields the panoramic images for the appearance-based approach presented here.

One objective of our ongoing research is to clarify whether appearance-based SLAM approaches are basically suited for large-scale and uniformly structured indoor environments, like the aforementioned home improvement store, and if so, how they can be made capable of working online and real-time. In our research on vision-based robot navigation, we are preferring the appearance-based approach for the following reasons: i) In a highly dynamic, populated and maze-like environment, a robust recognition of earlier selected natural landmarks cannot be guaranteed in any case. ii) Furthermore, the need for a robust and invariant detection of visual landmarks often results in highly computational costs and, therefore, map building is often performed off-line by these approaches. Because of the required online capability, we decided in favor of view-based techniques.

In our previous approach [1], [2] dealing with an appearance-based Monte Carlo Localization, a static, graph-based model of the environment was developed (see Fig. 2). The nodes (poses of the robot) of this environment model are labeled with appearance-based observations extracted from an omnidirectional image. Based on this environment model, we have developed an appearance-based visual SLAM approach that is using the Rao-Blackwellized Particle Filter (RBPF) concept [4].

The essential contribution of the approach presented here is, that we combined the appearance-based, visual localization concept with a Rao-Blackwellized Particle Filter as state estimator to a real-world suitable, online SLAM approach. In our system, each RBPF particle incrementally constructs its own graph-based environment model which is labeled with visual appearance features (extracted from panoramic 360$^o$ snapshots of the environment) and the estimated poses of the places where the snapshots were taken from. Another key idea of our approach is to utilize local graphs, representing a kind of short-term memory or time window of the current and

The research leading to these results has received funding from the State Thuringia (TAB-Grant #2006-FE-0154) and the AiF/BMWI (ALRob-Project Grant #KF0555101DF).

A. Koenig, J. Kessler, H.-M. Gross are with Neuroinformatics and Cognitive Robotics Lab, Ilmenau University of Technology, 98684 Ilmenau, Germany. alexander.koenig@tu-ilmenau.de

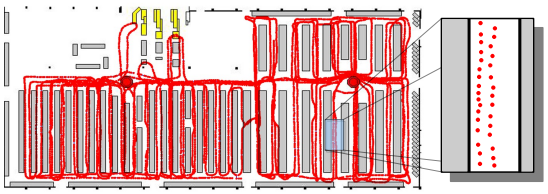

Fig. 2. A manually built map of the operation area, a regularly structured, maze-like home improvement store with a size of $100x50m^2$ (taken from [2]). The red dots show the positions of stored reference observations, i.e. the places where they were taken from. The correct pose was provided by an external reference system. To localize the robot in the home store, a Particle Filter was used which determines the weights of the particles by calculating the similarities between the current observation and the respective reference observations stored in the graph.

the very latest observations and pose estimations, instead of single observations typically used in the field of appearance-based localization/mapping to perform the evaluation step. Based on this, we introduce a graph-matching technique to compare the local graph of each particle with its particular global graph to determine the best matching map. Another novel idea consists in online estimating an environment-depending sensor model. This sensor model is used by the graph-matching algorithm to evaluate the likelihood of each map. The appearance-based SLAM approach presented subsequently is able to robustly build consistent maps of environments with a low demand on computational and memory costs. Therefore it works in real-time.

The rest of the article is structured as follows: The next section first gives an overview of related works. After that, the developed appearance-based SLAM approach based on RBPF and the novel key ideas of our approach, the graph-based environment model, the employed graph matching technique, and the adaptive sensor model, are described in detail. Experimental results achieved with our SLAM approach in different environments are presented and discussed in the section following that. A conclusion and suggestions for future work are given at the end of the paper.

## II. Related work

Many solutions have been presented in the past to realize a robot self-localization in more or less complex environments including methods based on feature or landmark extraction and tracking and those based on appearance models of the environment. A short overview of the most relevant research is given in the following:

**Feature/Landmark-based approaches:** In many SLAM approaches, the map representation is assumed to be a vector of point-like feature positions (landmarks) [7]. The attractiveness 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 robustly recognize the landmarks 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 [8], [9] has often been used so far. Further important feature/landmark-based approaches are those by Davison using Stereo vision [10] or monocular vision [11]. To estimate the landmark positions, popular methods like the Extended Kalmanfilter (EKF) [11], Rao-Blackwellized Particle Filters (RBPF) [12] or FastSLAM [13] are applied.

**Appearance-based SLAM/CML approaches:** The Concurrent Mapping-building and Localization (CML) approach of Porta and Kroese proposed in [19] 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. [20]. Here, a topological map stores nodes with appearance-based features and edges which contain relations between nodes and their poses. Essential drawbacks of this approach are, however, the required offline relaxation phase and the computational costs for calculation of the SIFT features. To avoid these requirements, our approach uses simpler and, therefore, faster methods to extract appearance features and also uses an RBPF [4] to avoid off-line relaxation methods. Moreover, the method to estimate the pose difference between images applying the image similarity introduced by Andreasson [20] has been picked up and extended in our SLAM approach. Further approaches that use a topological map representation are described in [22], where a Bayesian inference scheme is used for map building, and in [23], where a fast image collection database is combined with topological maps that allows an online mapping, too.

## III. Appearance-based SLAM approach with RBPF

### A. RBPF with local and global graph models

Our appearance-based SLAM approach also utilizes the standard Rao-Blackwellized Particle Filter approach to solve the SLAM problem, where each particle contains a pose estimate $\mathbf{x}_i$ (position $x, y$ and heading direction $\varphi$) as well as a map estimate (see Fig. 3). The environment model (map) used in our appearance-based approach is a graph representation, where each node $i$ representing a place in the environment is labeled with appearance-based features (for details see subsection III-D) $\mathbf{z}_i$ extracted from the panoramic view captured at that place and the estimated pose $\mathbf{x}_i$. 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* $\mathbf{m}^G = \mathbf{x}_{1:(l-1)}, \mathbf{z}_{1:(l-1)}$, which represents the already known environment model learned so far and a *local map* $\mathbf{m}^L = \mathbf{x}_{l:t}, \mathbf{z}_{l:t}$ representing the current and the $n$ latest observations and the local path between them (e.g. the last two meters of the robot's trajectory). The global map stores all observations before the time-step $l = t - n$. Thus, in extension of known RBPF techniques, in our approach each particle (see Fig. 3) estimates and stores a local map, a global map, and the currently estimated pose. Figure 4 schematically shows the local and global map in more detail. Serving as a short-term time-window of observations the local map is used to compute the likelihood

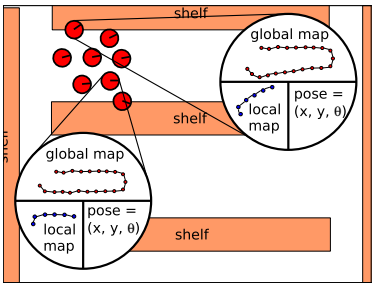


Fig. 3. The map representation of the particles in our approach: Each particle models its current pose estimation based on its path history, a full global map, and a local map. Due to the path history, all particles are slightly different and their maps differ as well. Some maps are more likely to be correct and consistent than others.

of each global map to be correct. This has two advantages. First, the likelihood of a given global map can be evaluated in a simple way by comparing the local and the global map directly. Based on this comparison, the RBPF can determine for each particle the likelihood of the trajectory and the global map estimated by that particle. Second, the still more relevant advantage is that the local map provides both geometric and visual information about the lastly observed places. Only this way, correct comparisons can be made taking both spatial relations and visual observations into consideration. The approach assumes that a short part of the robot's trajectory is measured correctly. Hence, the length of the local map has to be as short as possible. On the other hand, for reliable comparison results, the graph of a local map has to contain enough nodes for the matching process. We achieved the best results with a local trajectory length of about 2 meters and average distances between the nodes of 0.2 meters. Note, that this values are depending on the accuracy of the robot's odometry sensors.

### B. Graph matching

Our approach does not compute the probability distribution $p(\mathbf{z}|\mathbf{x},\mathbf{m})$ directly, but the likelihood of a given map to be correct is estimated by comparing the local and the global map. In the context of RBPF, these distribution determine directly the importance weight $w = p(\mathbf{z}|\mathbf{x},\mathbf{m})$ of a particle. For this purpose, corresponding pairs of nodes in both maps are selected by a simple nearest neighbor search in the position space. The relation between each selected pair of corresponding nodes $e_i^L$ (of the local map $\mathbf{m}^L$) and $e_j^G$ (from

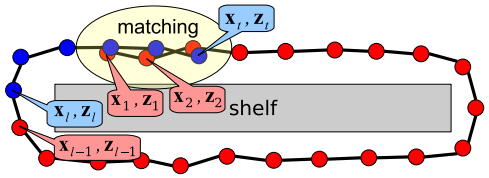


Fig. 4. Graph-based environment representation of our appearance-based approach: The red nodes show the global map of a single particle with respect to the path estimated by this particle. The blue nodes code the local map, whose data represent a short-term time-window of observations (the current and the $n$ latest observations) used for map matching to determine the likelihood of the respective global map. The idea of our appearance-based RBPF is, that only particles with correctly estimated trajectories are able to build correct maps, so that the matching between local and global map provides a higher matching value than wrongly estimated trajectories.

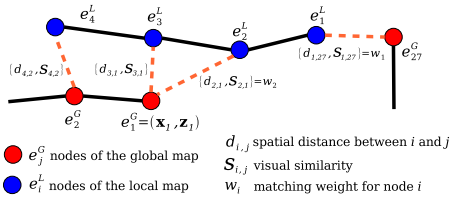


Fig. 5. Basic idea of our map matching algorithm: the likelihood of a given global map (particle-specific) is determined by comparison of the spatial distances $d_{ij}$ and visual similarities $S_{ij}$ between each pair of nodes $i$ (in the local graph) and $j$ (in the global graph). Corresponding nodes $e_i^L$ and $e_j^G$ are defined by the minimum spatial distances $d_{ij}$. The matching weights $w_i$ per node in the local map are calculated in dependence of the spatial distance $d_{ij}$ and the visual similarity $S_{ij}$ (for more details see Section III-C).

the global map $\mathbf{m}^G$) provides two pieces of information, a geometric one (spatial distance $d_{ij}$) and a visual one (visual similarity $S_{ij}$), depicted in Fig. 5. Both aspects of each relation $ij$ are used to determine a matching weight $w_i$ for the respective node $i$ of the local map. Assuming an independence between the node weights of the local map, the total matching weight $w^{[k]}$ between the local and the global graph of 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.

Because this graph matching needs to consider both spatial and visual similarities, we have to distinguish between several cases taking the different spatial situations (narrowness, wideness) in the local vicinity of the robot into consideration: in case of a correctly built global map, corresponding nodes with low spatial distance $d_{ij}$ code nearby places that show a similar appearance. This results in a high visual similarity $S_{ij}$ which should be considered in a high matching weight $w_i$. If the nearest corresponding nodes $e_i^L$ and $e_j^G$ are spatially more distant, however, then no evidence is given whether the map could be correct or not. In this case an average matching weight $w_i$ should be assigned. And finally, if the corresponding nodes in the local and global map seem to have only a low spatial distance but also show only a low visual similarity, then the respective global map must be wrong, which should result in low matching weights. A graph matching algorithm needs to take these different cases into account. Before the following section describes how this is realized by means of the adaptive sensor model, a short summary of our RBPF with topological maps will be given. Estimating correct maps with RBPF approaches requires loops in the trajectory. When the robot closes a loop, the local and global maps overlap and the graph-matching algorithm can estimate the likelihood of a map. In the case of non-overlapping maps no weights are determined by graph-matching, but set to the apriori weight computed by the average weight of all particles with overlapping maps. This prevents particle depletion before a loop is closed. If there are no particles with overlapping maps (e.g. at the start) the uniform distribution is used for particle weighting. Note, that

in the context of RBPF no explicit loop closing detection is performed and, therefore, no explicit trajectory (or map) correction can be done.

### C. Adaptive sensor model

To compute the matching weights between corresponding nodes, an adaptive sensor model had been developed which respects the aforementioned cases and requirements. In the context of appearance-based observations, the visual similarity between observations is not only depending on the difference in position but also on the environment itself. If the robot, for example, moves in a spacious environment with much free-space, the similarity between observations from slightly different positions will be very high. In a narrow environment with many obstacles, however, observations at positions with low spatial distance are already drastically influenced, which leads to low visual similarities. To that purpose, our sensor model estimates the dependency between surrounding-specific visual similarities $\hat{S}_{ij}$ of the observations $\mathbf{z}_i$ and $\mathbf{z}_j$ and their spatial distance $\hat{d}_{ij}$. An example for this dependency is shown in Fig. 6. The samples (blue dots) to built such a model are taken from the nodes of the local map where each node is compared to each other. This results in $n^2/2$ pairs of $\hat{S}_{ij}$ and $\hat{d}_{ij}$ representing samples describing the appearance variability in the local environment. Different approaches to approximate the model were investigated, e.g. the Gaussian Process Regression (GPR) of Rasmussen [24]. Despite of the advantageous of non-parametrical description, we decided against the use of GPR because of the computational costs associated with the algorithm, since the model has to be computed after each motion step. Furthermore, the model has to be applied to each node of the local map of each particle while the graph-matching process. Therefore, our approach uses a parametrical polynomial description of the sensor model and its variance. The parameter are estimated by a simple least square optimization. Then, the model provides two values, an expected similarity $\hat{S}(d_{ij})$ and a reliability $\hat{\sigma}(d_{ij})$ for a given distance $d_{ij}$. The likelihood that two nodes $i$ and $j$ of particle k are matching is computed as follows:

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

With that adaptive sensor model and the aforementioned graph-matching algorithm, the importance weight of each particle (the likelihood of a map) can be determined.

### D. Visual features

Before the experimental results will be presented in the following section, possible image features for our approach are introduced. Concerning the reasons discussed in Section I, we prefer an appearance-based approach utilizing holistic appearance features extracted from panoramic snapshots obtained from the omnidirectional camera located at the top of our experimental robot platforms (see Fig. 1). The environment observation for a given position of the robot is described by one 360° view. Because the heading direction of the robot only results in a rotated omniimage, the appearance-based features extracted from such a snapshot image are independent from the robot's heading and can simply be brought in reference orientation by simple rotation.

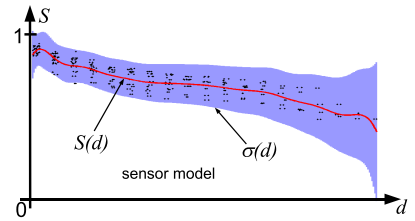


Fig. 6. The adaptive sensor model (shown by the red curve *similarity* and the blue area *reliability* ) is modeling the dependency of the visual similarity between two appearance observations and their spatial distance in the local environment described by the local map. The model uses a parametrical polynomial description which is continuously re-estimated and updated by means of the visual and metric labels of the local graph nodes (black dots).

A number of possible appearance-based features has been studied in our lab so far with respect to their capability to visually distinguish neighbored positions in the environment, the computational costs, and the preservation of similarities under changing illumination conditions and occlusions. The respective features experimentally investigated include local RGB mean values [1], HSV-histograms [2], FFT-coefficients (as proposed by Menegatti [17]), and SIFT features [8], [20] as appearance-based image description. All these features exhibit several pros and cons. RGB, HSV and FFT are very fast to compute and to compare, furthermore, their memory requirements are small. However, occlusions and illumination changes are a problem, whereby the HSV features are influenced least. In contrast to that, SIFT is slow in computation and comparing, but the advantage is the ability to determine the position of the salient observation with high accuracy. In our SLAM approach, however, where an adaptive sensor model is continuously estimated (see Section III-C), the computation time grows quadratically with the number of nodes in the local map. Considering the findings and constraints of these preliminary investigations, in our final implementation we decided in favor of the HSV histogram features (lower memory usage and computational costs than SIFT), especially because of the real-time requirements of our SLAM approach and the high robustness of these features. The extraction of the HSV features and details of an automatic luminance stabilization and color adaptation to cope with highly varying illumination conditions are explained in detail in [2].

## IV. Experiments and Results

Our appearance-based SLAM-approach and its general map building and localization capabilities were investigated in two different environments each of them showing specific constraints and challenges. For the first experimental investigation, we used our default test environment [2], the

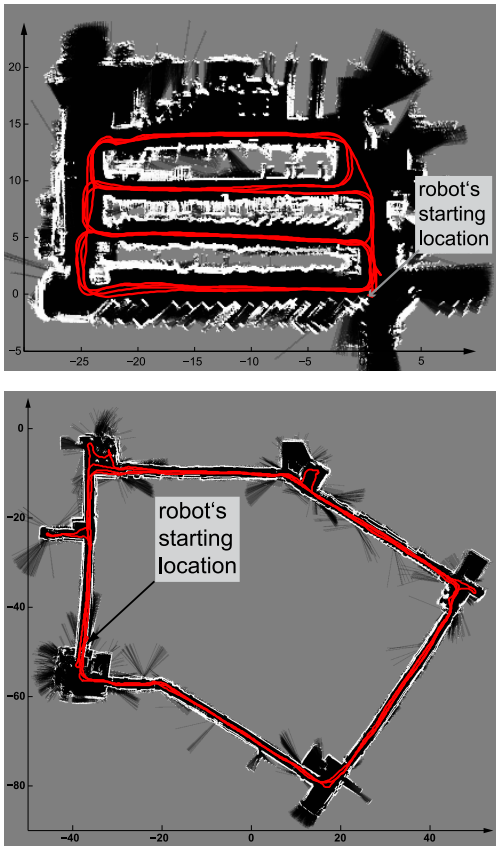


Fig. 7. **Experiment 1** (top) - in the home improvement store: The red path shows the robot's movement trajectory only estimated by our appearance-based SLAM approach. The map shows a high accuracy with only small alignment errors. For visualization of the localization accuracy, a laser-based occupancy map was built in parallel and superimposed to the estimated movement trajectory (visual SLAM). **Experiment 2** (bottom)- in a university building (Kirchhoff building): In this corridor-like environment it is very challenging to build up a map without distance measuring sensors because of the uniformity of the hallways and walls and the huge loop (250 m) that needs to be closed for place recognition.

home improvement store already introduced in Section I and shown in Fig. 2, where many dynamic effects occur. The second series of experiments was performed in a large pentagonal university building, the Kirchhoff-building, where the corridors are arranged in a closed loop. All data for the analysis were recorded under realistic conditions, i.e. people walked through the operation area, shelves were rearranged and with that their appearance, and other dynamic changes (e.g. illumination, occlusion) happened. Additionally, laser data were captured to generate a ground truth to evaluate the results of our appearance-based visual SLAM approach. Note, that the laser scanner in all cases only was used for comparison and visualization, but not for mapping or localization.

In the first experiment, the robot was moved several times through the home store along a path of about 500 meters length. Repeated loops around the goods shelves were driven, because the RBPF SLAM algorithm explicitly requires loop closures. The resulting graph (Fig. 7, top) covers an area of 20 x 30 meters and was generated by means of 250 particles in the RBPF. Thus, 250 global graph maps had to be built, whereas Fig. 7 only shows the most likely final trajectory and a superimposed occupancy map for visualization. This occupancy map was created by means of raw laser data arranged along the pose trajectory estimated by means of the appearance-based SLAM approach. The superimposed map gives an impression of the quality and accuracy of our visual SLAM algorithm. It only shows marginal alignment errors, all hallways and goods shelves are arranged very precise along straight lines. To evaluate the visually estimated path shown as red trajectory in Fig. 7, in addition a ground truth path and map built by means of a Laser-SLAM algorithm were calculated. A first result was, that the trajectories estimated by both SLAM approaches are very similar. This is also expressed by a mean localization error of 0.27 meters (with a variance $\sigma \approx 0.13m$) compared to the laser-based SLAM results. The maximum position error in this experiment was about 0.78 meters. These experimental results demonstrate, that our approach is able to create a consistent trajectory and for this reason, a consistent graph representation, too. Furthermore, in contrast to grid map approaches, topological maps require less memory because of the efficient observation storage, where each observation's features of each snapshot are linked to all global maps.

The second indoor experiment was carried out in the Kirchoff building where the robot drove a huge corridor loop with a path length of about 250 meters (Fig. 7, bottom). The challenge here is to close the loop at the right position. This is really hard because the hallways in that building all look very similar. In contrast to laser-based SLAM algorithms which often use scan-matching techniques to keep the odometry error small until a loop is closed, our RBPF approach cannot improve the odometry before a loop closing occurs. Thus, a position improvement cannot be done until the robot reenters areas it has already visited. The robot SCITOS used for these experiments produces a position error of 2-3 meters at the end of the 250 meters loop. Therefore, our algorithm needs to cope with this odometry uncertainty. Using a higher number of particles (1,500 particles) to approximate this state space, our algorithm also achieves good localization results with a mean localization error of 0.57 meters ($\sigma \approx 0.26m$) and a maximum error of 1.39 meters.

A final overview of the achieved results and the computational costs of our visual SLAM approach in all experiments and environments is given in Table I. It becomes evident, that in those environments, like the home improvement store, where the robot could be driven along smaller loops, the visual SLAM approach achieved the highest accuracy. The computational and the memory costs of the algorithm depend linearly on the number of particles and quadratically on the number of nodes in the local maps required for graph matching. At a single-core CPU with 1.8GHz, the computation of the adaptive sensor model is done in approximately 60 ms (independent from map and particle size), whereby 20 nodes of the local map were used for the estimation of the similarity model. The rest of the computational costs (see

TABLE I

OVERVIEW OF THE ACHIEVED RESULTS IN ALL EXPERIMENTS AND ENVIRONMENTS.

| Experiment | Home store | University building |
|---|---|---|
| **Size of area** | 30x20m | 90x120m |
| **Total path length** | 500m | 1400m |
| **# of particles** | 250 | 1500 |
| **Error Mean/Var/Max** | **0.27**/0.13/0.78 m | **0.57**/0.26/1.39 m |
| **Time per cycle** | 0.08 s | 0.16 s |

Table I) is spent for map updates, weights determination, and the resampling process. For a small number of particles (as used in experiment 1), our approach is running in real-time. With the higher number of particles used in experiment 2, the SLAM algorithm nearly works in real-time. In this case, the robot only has to be joy-sticked a bit slower through the environment.

## V. CONCLUSION AND FUTURE WORK

We presented a novel appearance-based SLAM approach which is able to localize a mobile robot in demanding real-world environments and to create a global map of the environment simultaneously. To the best of our knowledge, this is the first approach allowing an appearance-based on-line SLAM in large-scale and dynamic real-world environments. The essential contribution of this paper is the introduction of the visual appearance-based approach into the concept of Rao-Blackwellized Particle Filters (RBPF) to solve the SLAM problem. The key ideas of our approach are the use of global and local graph models per particle, the introduction of an adaptive sensor model, and the sophisticated graph-matching technique to compare the local graph of the particle with the respective global graph to determine the best matching map, and with that the best particles for the resampling step. The essential advantages of our appearance-based SLAM approach are its low memory and computing-time requirements. Therefore, our algorithm is able to perform in real-time which is a prerequisite for all on-line working map builders or mapping assistants.

We conducted a number of indoor SLAM experiments investigating the impact of the path length until loop closing, the variability of the visual appearances captured from the surroundings along the driven route, and the odometry quality on the stability and localization accuracy of this SLAM technique. Based on these encouraging results, several long-term experiments are planned for the near future to determine the accuracy of this approach in the complete store ($100x80m^2$), and to investigate the influence of dynamic changes in the environment (changing filling of the goods shelves, re-arrangements in the hallways, occlusions by people, changing illumination), to the robustness and long-term stability of our approach.

## REFERENCES


[1] 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 Int. Conf. Intell. Robots and Systems (IROS'02)*, 2002, pp. 256–262

[2] 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 Int. Conf. on Intelligent Robots and Systems (IROS'03)*, 2003, pp. 1505–1511.

[3] H.-M. Gross and H.-J. Boehme. PERSES - a Vision-based Interactive Mobile Shopping Assistant. In *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics (SMC'00)*, 2000, pp. 80–85.

[4] K. P. Murphy. Bayesian map learning in dynamic environments. In *Advances in Neural Information Processing Systems 12 (NIPS'99)*, 1999, pp. 1015–1021.

[5] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proc. of the AAAI Nat. Conf. on Artificial Intelligence*, 2002, pp. 593–598.

[6] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. Fastslam 2.0: An improved particle filtering for simultaneous localization and mapping that provably converges. In *Proc. of Int. Joint Conf. on Artificial Intelligence (IJCAI'03)*, 2003, pp. 1151–1156.

[7] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *Robotics Research, 4th Int. Symposium*, MIT Press, 1988, pp. 467–474.

[8] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. Int. Conf. on Computer Vision ICCV'99*, 1999, pp. 1150–1157.

[9] 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. 21, no. 8, pp. 735758, 2002.

[10] A. Davison and D. Murray. Simultaneous localization and map-building using active vision. *IEEE Trans. on PAMI*, vol. 24, no. 7, pp. 865 880, 2002.

[11] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. Int. Conf. on Computer Vision (ICCV'03)*, 2003, pp. 1403–1410.

[12] P. Elinas, R. Sim, and J. J. Little. $\sigma$SLAM: 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 (ICRA'06)*, 2006, pp. 1564–1570.

[13] T D. Barfoot. Online visual motion estimation using fastslam with sift features. In *Proc. of 2005 IEEE/RSJ Int. Conf. on Intelligent Robots and System (IROS'05)*, 2005, pp. 579–585.

[14] F. Pourraz and J.L. Crowley. Continuity properties of the appearance manifold for mobile robot position estimation. In *Proc. Symp. Intelligent Robotics Systems (SIRS)*, 1998, pp. 251260.

[15] I. Ulrich and I. Nourbakhsh. Appearance-based Place Recognition for Topological Localization. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA'00)*, 2000, pp. 1023-1029.

[16] P. Blaer and P. Allen. Topological mobile robot localization using fast vision techniques. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA02)*, 2002, pp. 10311036.

[17] 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)*, 2003, pp. 13-20.

[18] F. Dellaert, W. Burgard, D. Fox, and S. Thrun. Using the Condensation Algorithm for Robust Vision-based Mobile Robot Localization. In: *Proc. IEEE Int. Conf. on Comp. Vision and Pattern Recogn*, 1999.

[19] 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)*, 2004, pp. 3424-3429.

[20] 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 (ICRA'07)*, 2007, pp. 4096–4101.

[21] D. Haehnel, W. Burgard, D. Fox, and S. Thrun. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'03)*, 2003, pp. 206–211.

[22] A.Ranganathan, E. Menegatti, F. Dellaert. Bayesian inference in the space of topological maps. In *Robotics, IEEE Transactions on, vol.22, no.1*, pp. 92-107, Feb. 2006

[23] F. Fraundorfer, C. Engels, D. Nister. Topological mapping, localization and navigation using image collections. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'07)*, 2007, pp. 3872-3877

[24] C. E. Rasmussen, C. Williams. Gaussian Processes for Machine Learning, the MIT Press, 2006