in: Proc. ICPR 2004, 17th Int. Conf. on Pattern Recognition, Cambridge, pp. 266-269, IEEE Computer Society 2004

Robust Omniview-based Probabilistic Self-Localization for Mobile Robots in Large Maze-like Environments

Horst-Michael Gross, Alexander Koenig
Ilmenau Technical University, Department of Neuroinformatics
98684 Ilmenau, Germany
Horst-Michael.Gross@tu-ilmenau.de

Abstract

This paper extends our previous work on omniview-based Monte Carlo Localization. It presents a number of improvements addressing challenges arising from the characteristics of the given real-world application, the self-localization of a mobile robot in a regularly structured, maze-like and populated operation area, a home store. The contribution of this paper can be summarized as follows: we introduce a more specific extraction of color-based appearance features and propose a novel selective observation comparison method to determine the similarity between expected and actual observation allowing a better handling of severe occlusions or disturbances. Moreover, we present the results of a series of localization experiments studying the impact of the appearance-feature extraction and the observation comparison on the localization accuracy. Our improved approach can successfully demonstrate its omniview-based localization capabilities for a demanding, large operation area - a home store with a size up to $100 \times 60m^2$. *To the best of our knowledge, this is the most complex operation area that has been studied experimentally so far using appearance-based localization techniques.*

1. Introduction

Self-localization is the task of estimating the pose $\underline{x}$ of a mobile robot given a map of the environment and a history of sensor readings and executed actions. This includes both the ability of globally localizing the robot from scratch, as well as tracking the robot's position once its location is known. The current state-of-the-art localization methods often use distance sensors (laser range finders or sonar), but these sensor tend to be easily confused in environments with very regular, maze-like topology, like supermarkets or home stores, with their periodical hallways structure. Vision-based systems, however, do not show these limitations, but supply a much greater wealth of information about the 3D structure of the environment. Because of this, in recent years a number of vision-based localization approaches has been proposed. In most of them localization is typically performed in a straightforward way: the current input image or a more or less complex sub-space projection is compared to all references, and that location whose reference best matches the current image is considered to be the location currently taken by the robot. The similarity between current and all reference observations is typically determined by means of more or less complex distance measures (L_1 or L_2 norm, histogram intersection, Jeffrey divergence, etc.). Typical examples of these straightforward approaches are: the use of Eigenspace representations [1], the utilization of color histograms [6] or multi-dimensional histograms combining color with texture, etc. [8], or the use of Fourier spectra [5] as signatures for the omni-images.

In regularly structured, maze-like environments, however, localization on the basis of such a crisp mapping from observation $\underline{o}$ to pose $\underline{x}$ becomes very unreliable. Therefore, probabilistic methods, like Bayes filters, are required to quantify the ambiguity by means of beliefs for multiple location hypotheses. This allows to generate and track a set of alternative location hypotheses in parallel which can be disambiguated in the further course of the localization process. Instead of using highly sophisticated image transformation and feature extraction techniques followed by a classifying mapping from the current observation to the best fitting internal reference, inspired by Fox et al. [2] we have favored the opposite way, namely the extraction of relatively simple appearance features in combination with a probabilistic multi-hypothesis estimation by means of Particle Filters. In [3] we introduced the basic idea of our omniview-based Monte Carlo Localization (MCL) and presented very first experimental results which could demonstrate the general feasibility of this approach. Up to now, only a few other approaches are known that successfully combine omnivision with Bayesian localization techniques, e.g. that of Menegatti [5] which is based on Fourier spectra of the omni-images. However, most of these approaches

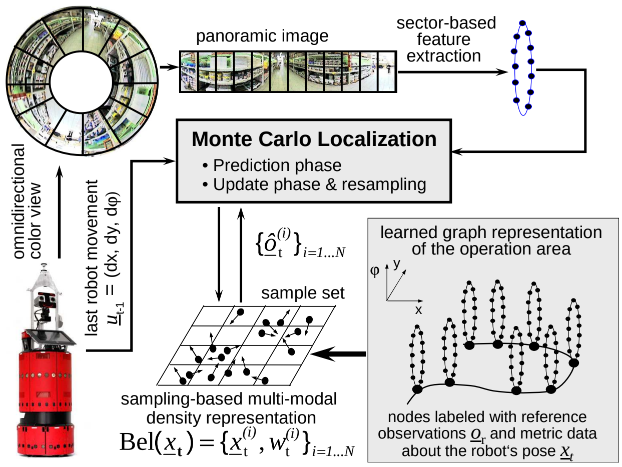

Figure 1. General idea of our omniview-based probabilistic self-localization.

have been studied experimentally only under lab conditions (small operation area with simple structure, constant illumination conditions, etc.). Till now, they have not been able to demonstrate their robustness and reliability in demanding, large-scale operation areas handling occlusions by people and illumination disturbances.

In this paper, we present a number of improvements of our omni-view based probabilistic approach addressing challenges arising from the characteristics of our real-world application, the vision-based self-localization of a mobile robot in a regularly structured, maze-like and populated large-scale operation area, a typical home store. Against this background, the contribution of this paper can be characterized as follows: First, we introduce a more specific feature extraction based on color-histograms instead of mean RGB-values determined per segment of the omni-image so far [4]. Second, we present a novel selective observation comparison method to determine the similarity between expected and actual observation. This new comparison method allows to better deal with severe occlusions or disturbances of the omni-image caused by, e.g. people standing nearby the robot or local illumination artifacts. And third, we present the results of a great number of comparative localization experiments studying the impact of the new feature extraction and observation comparison on the accuracy and robustness of the localization.

2. Omniview-based MCL

The general idea of our omniview-based MCL is illustrated in Fig. 1. As map of the environment, our approach employs a graph representation of the environment (Fig. 1, bottom right) which is learned on-the-fly while manually joy-sticking the robot through the operation area. Each node of the graph is labeled with both visual reference observations $\underline{o}^r(x, y, \varphi)$ extracted from the respective panoramic view at position x, y in heading direction φ and the respective metric data about the robot's pose during node insertion. A new reference point is inserted, either if the position distance to other reference points in a local vicinity is larger or if the similarity between the current observation $\underline{o}_t$ and the reference observations $\underline{o}^r(x, y, \varphi)$ at adjacent reference points is smaller than given threshold values. It should be noted, that the labeling of the reference points with correct pose data necessitates a powerful odometry correction because of the increasing error over time. For that purpose, we developed an efficient vision-based method [7].

Extraction of visual appearance features: The following criteria determined the selection of appropriate appearance features to describe the omniview in a specific pose of the robot: a) To allow for an on-line localization, the calculation of the features should be as easy and efficient as possible. b) The feature set should include the orientation of the robot as prerequisite to estimate its heading direction - therefore the feature set needs to preserve at least global spatial relations within the image. c) It should allow an easy generation of expected observations for hypothetical poses. Considering these criteria, we realized the following approach: both during map-building and self-localization, the omniview image is transformed into a panoramic image (Fig. 1, top) which is partitioned into a fixed number of segments. Then for each segment simple color features are extracted. In the past, per segment only mean RGB-values were determined, while our new feature extraction computes a color histogram per segment. To compare the computational costs, the preservation of similarities under changing illumination conditions and image capturing positions, and the localization accuracy as integral performance measure, we experimentally studied several color spaces for histogram building. Best results we achieved with the HSV-color space and low-dimensional histograms consisting of 16 bins for the H-component (coding color type) and 4 bins for the V-component (coding brightness of color).

Selective observation comparison: During particle filter update, for each sample $s^{(i)}$ an importance weight $w_t^{(i)}$ is computed which describes the probability that the robot might be located at the pose $\underline{x}_t^{(i)} = (x, y, \varphi)_t^{(i)}$ of the respective sample i. The importance weights $w_t^{(i)}$ are determined on the basis of the similarity $S_t^{(i)}$ between the expected observation $\underline{\hat{o}}_t^{(i)}$ of the sample i in its current pose $\underline{x}_t^{(i)}$ and the actual observation $\underline{o}_t$ applying a camera-specific observation model $p(\underline{o}_t|x_t)$. Our *non-selective comparison method* used so far simply determines this value by means of the L_2-norm of the difference vector $\underline{o}_t - \underline{\hat{o}}_t^{(i)}$. The general idea of our new *selective observation comparison* can be described as follows: If no local occlusions or disturbances occur in a specific robot pose, the expected and the actual observation should match largely. In case of local disturbances, however, the affected segments will show significant differences and, therefore, should be left out from computing the total similarity. Because of this, our new algorithm takes only those segments into account that show the highest similarity between expected and actual observation as well as maximize a similarity function (see below). This way, the negative influence of local occlusions or disturbances on the total similarity can be largely reduced. First, for each segment j of the n segments a segment-specific similarity s_j is determined by means of the Histogram Intersection (HI) between the expected histogram $\underline{\hat{o}}_j$ and the actually observed histogram $\underline{o}_j$ of the corresponding segments. After that, the segment similarities s_j are ranked in descending order yielding those segments most important for the current observation comparison. Thereafter, the largest total similarity is searched for by iterating the number of segments k taken into account. To avoid a trivial summation of all segment similarities, a penalty function is introduced which is counteracting this effect by preferring a small number of segments k. In a series of experiments, several penalty functions were investigated. The nonlinear term $\sqrt{1/(k \cdot n)}$ yielded the best selection results. This way, the total similarity for sample i can be determined as follows:

$$S^{(i)} = \max_{k=1}^{n} \left(\left(\sum_{j=1}^{k} s_j \right) \cdot \sqrt{\frac{1}{k \cdot n}} \right)$$

Finally, $S^{(i)}$ is used to determine the weight $w_t^{(i)}$ of i.

3. Experimental results

All experiments were carried out in a home store with our experimental platform PERSES, a standard B21 robot additionally equipped with an omnidirectional camera for vision-based navigation and human-robot interaction (Fig. 1, left). As opposed to earlier experiments, we introduced a new regime for the localization experiments which consequently distinguishes between *training tours* traveled to build the graph and *test tours*, executed to acquire unknown data for the localization experiments. For the recent experiments, we learned a new graph representation of a section of the home store as shown in Fig. 2. Each dot represents a node labeled with both the pose of the robot and the reference observation. The total distance to be travelled during training was about 1500 meters. Using the recorded test data acquired during several test tours, we conducted a great number of localization experiments to determine accuracy, robustness, and localization speed comparing both the feature extraction and observation comparison methods. In all experiments, we studied the worst-case scenario: our robot had no prior information about its initial pose. Each localization experiment was repeated 50 times to determine the mean localization errors and their variances.

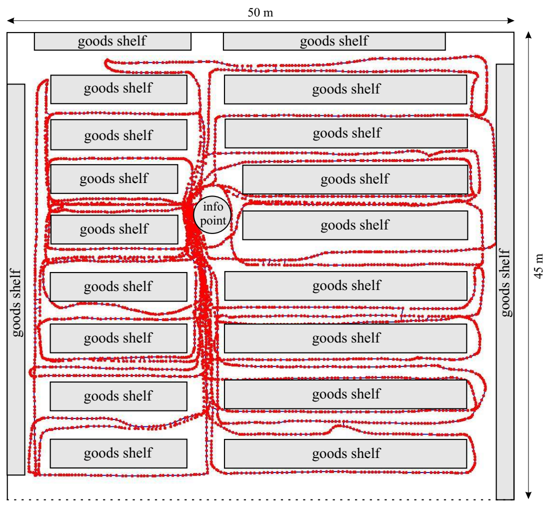

Figure 2. Learned graph of the operation area in the home store (3500 reference points).

Mean RGB-values vs. color histograms: The objective of these experiments was to compare the localization accuracy and loss rate achieved for the two feature extraction methods (RGB v. Hist). The localization is regarded as lost, if the position error is larger than 2.5 m per step. Table 1 compares the mean position errors of several test tours traveled a couple of days after the training tour for both feature extraction techniques - at first using the *non-selective comparison method*. The experiments clearly express the superiority of the histogram-based feature extraction.

	TT1	TT2	TT3	TT4	⊘
Position error (cm)					
mean RGB-value	75	192	39	51	68
Histogram	72	51	41	47	50
Loss rate (%)					
mean RGB-value	0	6.5	0.1	0	0.7
Histogram	0	0	0	0	0

Tabelle 1. Comparison of the mean position errors and the loss rate during several test tours (TT)

Selective vs. non-selective comparison: In the following experiments we compared the two observation comparison methods, in this case using the simple RGB-based feature extraction. It turned out that our novel selective method **(Sel)** in all experiments produces lower position errors than the non-selective method **(NSel)**. In TT 2, a very demanding tour with a great number of natural occlusions and disturbances, the non-selective method fails almost completely. Due to the disturbances, it generates numerous alternative localization hypotheses and is not able to disambiguate the correct position. The experiments clearly express the merits of the our selective method in handling critical situations.

tour	Sel. method loc. error (in m)	Sel. method loc. σ (in m)	NSel. method loc. error (in m)	NSel. method loc. σ (in m)
TT1	0.71	0.25	0.75	0.33
TT2	0.93	1.47	1.92	2.07
TT4	0.45	0.20	0.51	0.28

Tabelle 2. Comparison of the mean localization errors and their variances.

Occlusion experiments: To investigate the efficiency and robustness of the two feature extraction and observation comparison methods dealing with occlusions and disturbances, in the following experiments we utilized the mean postion error as performance measure again. In the occlusion experiments, the omni-images of unknown test observations were randomly occluded by artificial image segments showing Gaussian-distributed color noise. The impact of occlusion effects was gradually controlled by the percentage of image content covered by these artificial images. The influence of the feature extraction and observation comparison methods on the accuracy of the pose estimation is illustrated in Fig. 3 for various degrees of occlusion and two different test tours. In both cases, the selective comparison in combination with segment-specific color histogram features shows the highest robustness against occlusions - it can handle rates up to 60%.

4. Summary and Outlook

We introduced several methodical improvements of our original approach [4] addressing challenges arising from the characteristics of our real-world scenario, a home store, and presented a series of new experimental studies. To better deal with severe image occlusions or disturbances, we proposed a novel, very robust selective observation comparison method as prerequisite for an effective particle filter update. Moreover, we introduced a more specific color-histogram based feature extraction instead of the mean RGB-values used so far. We conducted a series of localization experiments studying the impact of the feature extraction and the observation comparison on localization accuracy. The results of these experiments confirm the greater robustness and superiority of the improved approach in handling critical real-world situations, e.g. situations with large occlusions or local illumination artifacts. Our advanced omnivision-based MCL enables real-time localization, because the complete sampling/importance re-sampling of the particles (500 for tracking, 5000 for global localization) only takes between 30 and 300 ms on a 1.5 GHz PC. In most recent experiments, our approach could demonstrate its capabilities for a very demanding, large-scale operation area, a populated home store with a size of $100 \times 60m^2$ consisting of more than 40 hallways. To the best of our knowledge, this is the most complex operation area that has been studied experimentally so far using appearance-based visual localization techniques. Currently, we are intensely dealing with the on-line adaptation of the reference observations in order to better handle extreme appearance variations at the reference points due to, e.g., changes in the operation area.

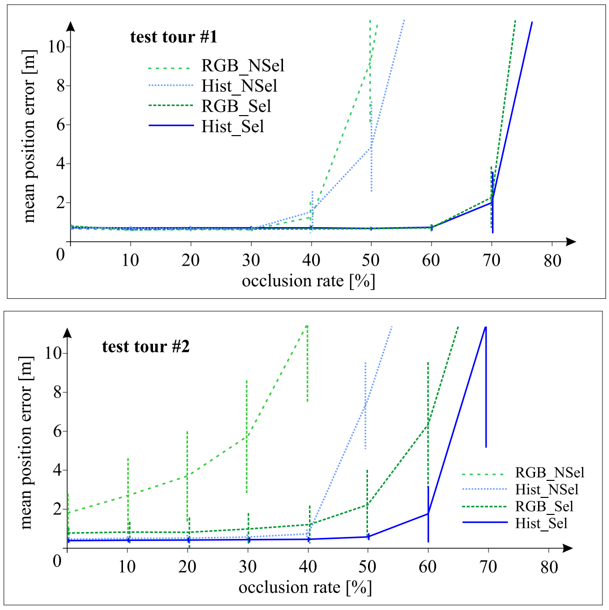

Figure 3. Results of occlusion experiments showing the mean position errors and their variances.

References

[1] M. Artac, M. Jogan and A. Leonardis. Mobile Robot Localization Using an Incremental Eigenspace Model. In: *Proc. ICRA 2002*, 1025-1030

[2] D. Fox, W. Burgard, F. Dellaert, S. Thrun. Monte Carlo Localization: Efficient Position Estimation for Mobile Robots. In: *Proc. AAAI-99*, 1999

[3] H.-M. Gross, H.-J. Boehme. PERSES - a Vision-based Interactive Mobile Shopping Assistant. in: *Proc. IEEE Intern. Conf. on Systems, Man and Cybernetics, 2000*, pp. 80-85

[4] H.-M. Gross et al. Vision-Based Monte Carlo Self-localization for a Mobile Service Robot Acting as Shopping Assistant in a Home Store. *Proc. IROS 2002*, pp. 256-262

[5] E. Menegatti et al. Hierarchical Image-based Localisation for Mobile Robots with Monte-Carlo Localization. *Proc. Europ. Conf. on Mobile Robots (ECMR) 2003*, pp. 13-20

[6] I. Ulrich and I. Nourbakhsh. Appearance-based Place Recognition for Topological Localization. *Proc. ICRA 2000*, pp. 1023-1029

[7] C. Schroeter et al. Extraction of Orientation from Floor Structure for Odometry Correction in Mobile Robotics. in: *Proc. DAGM 2003*, pp. 410-17, LNCS 2781, Springer 2003

[8] C. Zhou, Y. Wei and T. Tan. Mobile Robot Self-Localization Based on Global Visual Appearance Features. *Proc. ICRA 2003*, pp.1271-1276