2005 IEEE International Conference on Systems, Man and Cybernetics
Waikoloa, Hawaii October 10-12, 2005

Omniview-based Concurrent Map Building and Localization using Adaptive Appearance Maps

H.-M. Gross, A. Koenig, St. Mueller
Ilmenau Technical University, Department of Neuroinformatics and Cognitive Robotics
98684 Ilmenau, Germany
Horst-Michael.Gross@tu-ilmenau.de

Abstract - *This paper describes a novel omnivision-based Concurrent Map-building and Localization (CML) approach which is able to robustly localize a mobile robot in a uniformly structured, maze-like environment with changing appearances. The presented approach extends and improves known appearance-based CML techniques in a few essential aspects. For example, an advanced learning scheme in combination with an active forgetting is introduced to allow a complexity restricting adaptation of the environment model to appearance variations of the operation area. Moreover, a generalized scheme for fusion of localization hypotheses from several state estimators with different meaning and certainty and a distributed coding of the current observation by a weighted set of reference observations is proposed. Finally, several real-world localization experiments investigating the stability and localization accuracy of this novel omnivision-based CML technique for a highly dynamic and populated operation area, a home store, are presented.*

1 Introduction

Self-localization is the task of estimating the pose 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. Many solutions have been presented in the past to realize a robust self-localization in complex operation areas including methods based on feature or landmark extraction and tracking, and those based on appearance models of the environment. Robust self-localization also 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, a customer, to desired articles within a home store realizing a *guidance function*, or follow him as an attentive *service-companion* [1]. To accommodate the challenges that arise from the specifics of this interaction-oriented scenario and the characteristics of the operation area, a regularly structured, maze-like and populated environment, we placed special emphasis on vision-based methods for both human-robot interaction and robot navigation. In our previous localization approach [2], a graph-based representation of the operation area is employed for appearance-based Monte Carlo Localization. The static environment model is learned on-the-fly while manually joy-sticking the robot through the operation area. The nodes of the graph are labeled with both visual observations extracted from the omnidirectional image and information about the pose of the robot at the moment of node insertion. The main drawback of this and other appearance-based approaches is, that localization is only possible in previously mapped areas. The construction of an appearance map is a supervised process, and the learned map is only valid as far as no important modifications of the operation area occur. Because of the characteristics of the home store as a highly dynamic operation area with a changing appearance, we were forced to develop an alternative approach which realizes a Concurrent Map-building and Localization and can adapt the learned environment model to the changing environment.

Inspired by the work of Porta and Kroese [4, 5] and continuing our former work, an alternative technique was developed, which is able to perform an omnivision-based Concurrent Map-Building and Localization (CML) and to overcome this drawback. In this new approach, the static, predefined map of the operation area is replaced by an environment model which is obtained and continuously refined by the robot as it moves through the operation area. If the robot re-visits an already explored area, it can use the information previously stored to reduce the uncertainty of its position and to adapt its internal model. As extension to [5], our CML approach proposes improved fusion and learning methods and uses alternative observations which can be summarized as follows: i) While the model of Porta and Kroese can learn new reference views and passively forget the positions of irrelevant ones to deal with dynamic environments, our approach is in addition able to actively delete those views no more relevant for the environment model (e.g., due to permanent appearance changes at certain positions). This active forgetting is of central importance to keep the complexity and the number of reference views under control. ii) Moreover, we employ a generalized scheme for fusion of localization hypotheses from several position estimators according to their relevance and certainty. This allows us to superimpose the position hypotheses from very different information sources. iii) As observations and reference views, we

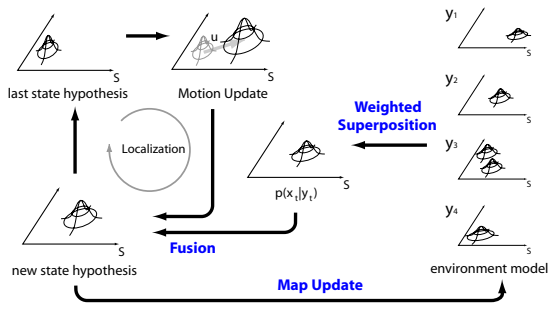

Figure 1: General idea of our Concurrent Map-building and Localization (CML) approach, from the modeling point of view closely related to [5]. In both approaches, Mixtures of Gaussians (MoG) are used to represent both the robot's state $\boldsymbol{x}_t = (x_t, y_t, \phi_t)^T$ in the environment (left) and to learn an environment model (right). The blue highlighted aspects "Fusion", "Map Update" and "Weighted superposition" mark the focus of this paper.

use panoramic 360^o images, so-called omniviews, which describe a certain position in the environment under all possible heading directions of the robot. Therefore, this type of visual input is preferred to describe the appearance of a position. iv) To determine the reference observations taken by the robot in positions visited earlier, often a crisp 1-of-N mapping from the current observation to the best fitting internal representation is used. Our approach, however, employs a distributed coding, which uses a set of most similar reference observations to describe the current observation and to model the position hypothesis by a weighted superposition of the corresponding position estimations.

In the following section, we first introduce the necessary mathematical background of the CML technique. After that, we describe specific aspects of our omniview-based CML approach with direct relevance for active adaptation of the environment model. Finally, we present encouraging experimental results obtained with our CML system during localization experiments in the home store and conclude summarizing our work and pointing direction for further experiments.

2 Advanced CML approach

The basic CML approach of Porta and Kroese proposed recently [5] is a promising technique to simultaneously build an appearance-map of the environment and to use this map, still under construction, to improve the localization of the robot. Both in their and our system, the robot's state $\boldsymbol{x}_t = (x_t, y_t, \phi_t)^T$ in the environment is represented by a Mixture of Gaussians (MoG) (see Fig. 1, left). In general, a mixture $X_t = \{(\boldsymbol{\mu}_t^i, \boldsymbol{C}_t^i, w_t^i) | i \in [1, N]\}$ is a set of partial hypotheses in form of single Gaussians with center $\boldsymbol{\mu}_t^i$ and covariance matrix $\boldsymbol{C}_t^i$. The weights w_t^i $(0 < w_t^i \leq 1)$ provide information on the certainty of the partial hypotheses. With that, the current state of a robot is described as

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{u}_t, \boldsymbol{y}_t) = p(\boldsymbol{x}_t) = \sum_{i=1}^{N} w_t^i \phi(x | \boldsymbol{\mu}_t^i, \boldsymbol{C}_t^i) \quad (1)$$

with $\boldsymbol{x}_{t-1}$ as last state, $\boldsymbol{u}_t$ as motion information, and $\boldsymbol{y}_t$ as current observation. In the following, the conditional terms are left out, the global localization hypothesis is simply called $p(\boldsymbol{x}_t)$ to simplify matters. The generation of the current localization hypothesis $p(\boldsymbol{x}_t)$ is the result of a fusion of several state hypotheses coded as MoG. Given the last state hypothesis $p(\boldsymbol{x}_{t-1})$, the propagation of the motion data $\boldsymbol{u}_t$ by the motion model leads to a first hypothesis $p(\boldsymbol{x}_t | \boldsymbol{u}_t, \boldsymbol{x}_{t-1}) = p(\boldsymbol{x}_t | \boldsymbol{u}_t)$ for the current state. In the following, this hypothesis has to be fused with other state hypotheses, e.g. $p(\boldsymbol{x}_t | \boldsymbol{y}_t)$ which results from the environment model using the current observation $\boldsymbol{y}_t$, or state hypotheses from other sources of information $p(\boldsymbol{x}_t | ...)$ (see Fig. 1 and 2 middle). This requires an advanced fusion scheme in order to allow a weighted superposition of several information sources with different meaning, considering such aspect like reliability or stability of the hypotheses.

2.1 Fusion of state hypotheses

To fuse the two hypotheses generated by the motion model and the environment model (see Fig. 2, a-e), at first corresponding Gaussians in both hypotheses $a \in p(\boldsymbol{x}_t | \boldsymbol{u}_t)$ and $b \in p(\boldsymbol{x}_t | \boldsymbol{y}_t)$ representing similar positions are determined. As proposed by [5], this is done by a simple criterion based on the Mahalanobis distance in *State Space*:

$$D(a,b) = \begin{cases} 1 & : (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) (\boldsymbol{C}_a + \boldsymbol{C}_b)^{-1} (\boldsymbol{\mu}_a - \boldsymbol{\mu}_b) < \gamma \\ 0 & : else \end{cases} \quad (2)$$

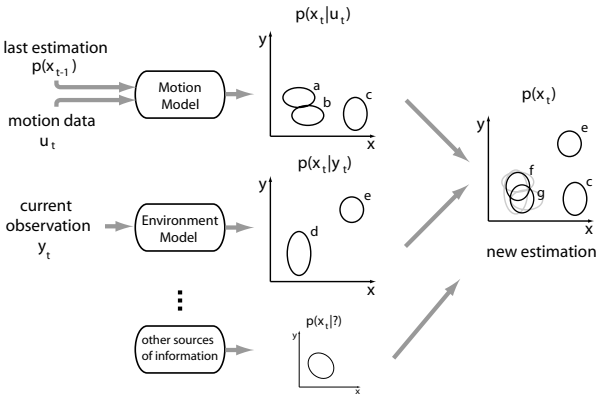

Figure 2: Advanced fusion of several state hypotheses: state hypothesis generated by the last estimation and moved according to the motion data and the motion model (top, a-c), new observation-based state hypothesis from environment model (middle, d-e), and additional state hypotheses suggested by other position estimators (bottom) can be superimposed and merged to a new distribution coding the current state hypothesis (right).

If $D(a,b) = 1$, the two corresponding Gaussians are merged by Covariance Intersection similar to [4]. The objective of this fusion is to minimize the variance of the resulting distribution. Therefore, the resulting Gaussian gets the following parameters:

$$\boldsymbol{C} = ((1-\alpha)\,\boldsymbol{C}_a + \alpha\,\boldsymbol{C}_b)^{-1} \quad (3)$$

$$\boldsymbol{\mu} = \boldsymbol{C}\left[(1-\alpha)\,\boldsymbol{C}_a^{-1}\boldsymbol{\mu}_a + \alpha\,\boldsymbol{C}_b^{-1}\boldsymbol{\mu}_b\right] \quad (4)$$

The control parameter α determines which Gaussian dominates the merging result. Alternative to [5], we take the weights w_a and w_b of both Gaussians into account to realize a weighted fusion:

$$\alpha = \frac{w_b\; det(\boldsymbol{C}_a)}{w_b\; det(\boldsymbol{C}_a) + w_a\; det(\boldsymbol{C}_b)} \quad (5)$$

To determine the new weights w^i of the resulting Gaussians, a little bit effort is needed because one component can have more than one corresponding counterpart (e.g. component d in Fig. 2). For this purpose, for each Gaussian first the weights of all corresponding Gaussians within the allowed neighborhood (i.e., all components with $D(a,b) = 1$) is determined:

$$\hat{w}_a = \sum_{j|D(a,j)=1} w^j_{p(\boldsymbol{x}|\boldsymbol{y})} \quad (6)$$

$$\hat{w}_b = \sum_{j|D(j,b)=1} w^j_{p(\boldsymbol{x}|\boldsymbol{u})} \quad (7)$$

With that, the weight w^i of the Gaussian is given by

$$w^i = w_a \frac{w_b}{\hat{w}_a} + w_b \frac{w_a}{\hat{w}_b} \quad (8)$$

Gaussians in $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ and $p(\boldsymbol{x}_t|\boldsymbol{u}_t)$ without a counterpart in the other distribution (see Fig. 2, components c and e) are simply copied into the resulting MoG. This way, also completely new hypotheses can be integrated. After this first fusion, the resulting MoG could be fused again with another state hypothesis to integrate further information sources (see Fig. 2 bottom). This has to be done, however, still before the final weight normalization to preserve a similar influence of all sources of information. During the final normalization of the resulting MoG, the weights of all unsupported components will be reduced passively. This way, their certainty is continuously reduced, and after a number of update cycles they can be deleted. To simplify the resulting MoG, overlapping adjacent Gaussians can be merged to a single Gaussian similar to the method described in section 2.3.

Afterwards, the resulting MoG coding the new state hypothesis $p(\boldsymbol{x}_t)$ (see Fig. 1, bottom left or Fig. 2, right) is used for updating the environment model. Only unimodal hypotheses are employed for this update step, otherwise the reconstruction of states between two consecutive, unimodal hypotheses takes place according to [5]. Experiments showed that this update must be delayed to avoid a positive feedback between state estimation and environment model and vice versa. In this case, the changed $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ would directly influence $p(\boldsymbol{x}_t)$ after a short time. Thereto, observation-state estimation pairs $(\boldsymbol{y}_t, p(\boldsymbol{x}_t))$ are stored in a short queue which is also used to buffer the time-consuming update operations of the environment model.

2.2 Improved environment model

The purpose of our improved environment model U is to allow an estimation of the current state $\boldsymbol{x}_t$ under the condition of the current observation $\boldsymbol{y}_t$ in a dynamic operation area with appearance variations. In our case, $\boldsymbol{y}_t$ represents the omnidirectional view in form of a describing feature vector. Inspired by the idea of [3], for feature extraction we use a Fourier transformation over averaged local view-segments of the omnidirectional image. This way, an efficient determination of rotation-invariant distances between the current view and the reference views is possible. The whole *Observation Space* is represented by a variable set of reference views $\boldsymbol{y}_i$. To each $\boldsymbol{y}_i$ a learned MoG X_i in the *State Space* is assigned.

$$U = \{(\boldsymbol{y}_i, X_i)\,,\; i = 1, \ldots, R\} \quad (9)$$

Each X_i is coding all those positions in the operation area the respective observation was captured before.

$$X_i = \{\left(w_{ij}, \eta_{ij}, \boldsymbol{\mu}_{ij}, \boldsymbol{C}_{ij}\right) | j = 1, \ldots, M_i\} \quad (10)$$

Every component j in X_i has a weight $w_{ij} > 0$ describing its certainty and relevance for the fusion operation. In contrast to the modeling of $p(\boldsymbol{x})$ by a Mixture of weighted Gaussians, these weights do not need to sum up to one. Thus, the term $\sum_{j=1}^{M_i} w_{ij}$ describes the total importance of the respective reference view $\boldsymbol{y}_i$ for the environment model.

In extension to [5], we explicitly want to consider and model appearance variations and fluctuations occurring at identical positions within the environment. Therefore, we had to introduce another control parameter η_{ij} that counts the number of observations at a position covered by a single Gaussian j. This parameter η_{ij} is necessary for a stable estimation of the distribution X_i. If, for example, X_i is the result of many fitting previous observations, the respective Gaussian j could accumulate a high η_{ij} value coding a stabile and reliable estimation. In this case, a single new position hypothesis will get less influence on the resulting distribution.

Given the current observation $\boldsymbol{y}_t$, the localization hypothesis $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ can be simply generated by finding the most similar reference-view $\boldsymbol{y}_i$ in the model U, as proposed by [5]. However, to achieve a smoother approximation of the hypothesis $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$, in our approach we realize a weighted superposition of several position hypotheses. To be precise, we superimpose the position hypotheses X_i of all that reference views $\boldsymbol{y}_i$ most similar to the current observation $\boldsymbol{y}_t$ in the *Observation Space*. As illustrated in Fig. 3-top left, only those reference-views are considered in this superposition whose distance to the current observation is lower than the limit distance E^{max}. Therefore, a similarity measure $S(\boldsymbol{y}_t, \boldsymbol{y}_i)$ was introduced which is 1.0 for a perfect matching and continuously decreases to zero up to the E^{Max}. The

final state hypothesis $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ is computed as weighted sum over all X_i whereas the individual weights w_{ij} of the Gaussians are multiplied by the similarity values $S(\boldsymbol{y}_t, \boldsymbol{y}_i)$ of the respective reference views.

Because the heading direction of the robot only results in a rotated omniimage (under the prerequisite the camera is mounted coaxial to the center of rotation), the comparison of views is invariant against different directions. Therefore, the orientation components of the selected MoG X_i only have to be rotated to refer the current view. The best matching rotation angle can be simply determined by the maximum cross-correlation between $\boldsymbol{y}_t$ and $\boldsymbol{y}_i$ over the angle.

2.3 Updating the environment model

While using the model for generation of state hypotheses, new observations are made that must be integrated. In the simplest case, the model is only updated by pairs of observation $\boldsymbol{y}_t$ and unimodal state distribution $p(\boldsymbol{x}_t)$, i.e. a single Gaussian $\phi(\boldsymbol{x}|\boldsymbol{\mu}_t, \boldsymbol{C}_t)$. A dynamic environment and the arising stability requirements, however, need a more complex update regime that has to take into consideration the following premises: i) one observation can only be generated by one position in the area and ii) each position does only show one appearance at a time. Therefore, the update of the environment model is typically carried out in three phases.

1) Insertion of a new reference view: If the feature distance between the current observation and all learned reference views is larger than the limit distance E^{max}, the current observation $\boldsymbol{y}_t$ has to be stored as a new reference view $\boldsymbol{y}_n$. Initially, the position estimation X_n of this new reference-view is empty, in the further course of learning state hypotheses are inserted (see next point).

2) Update of the Mixtures of Gaussians: A new Gaussian representing the current state hypothesis $p(\boldsymbol{x}_t)$ is added to the already existing Gaussian Mixtures X_i using the similarity values $S(\boldsymbol{y}_t, \boldsymbol{y}_i)$ mentioned above as gain control. The basic idea behind this update step is as follows: since each X_i is the result of all former observations, a balanced insertion of new Gaussians has to be realized. Therefore, if a new component k has to be inserted into X_i with the describing parameters $(w_{ik}, \eta_{ik}, \boldsymbol{\mu}_t, \boldsymbol{C}_t)$, the initial weight w_{ik} is set to zero, i.e. this Gaussian still is without relevance for the MoG at the beginning. The initial observation counter η_{ik} is set to a value describing the ratio between the current similarity value and the sum over the similarity values of all reference-views:

$$\eta_{ik} = \frac{S(\boldsymbol{y}_i, \boldsymbol{y}_t)}{\sum_{l=1}^{R} S(\boldsymbol{y}_l, \boldsymbol{y}_t)} \tag{11}$$

The resulting MoG can be simplified by a merging of similar components to a single one, where the parameters η and w of the involved Gaussians have to be summed up. If two Gaussians n and m fulfill the following distance criterion:

$$(\boldsymbol{\mu}_{in} - \boldsymbol{\mu}_{im})^T (\boldsymbol{C}_{in} + \boldsymbol{C}_{im})^{-1} (\boldsymbol{\mu}_{in} - \boldsymbol{\mu}_{im}) < \delta \tag{12}$$

they are reduced to a single Gaussian k replacing n and m with the new mean

$$\boldsymbol{\mu}_{ik} = \eta_{in}\boldsymbol{\mu}_{in} + \eta_{im}\boldsymbol{\mu}_{im} \tag{13}$$

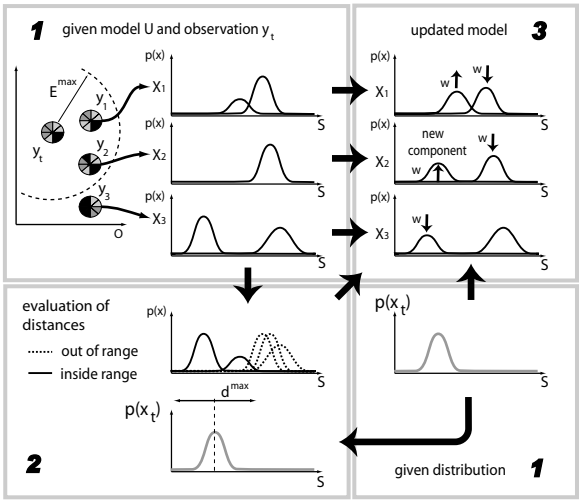

Figure 3: Schematic illustration of the weight updating in the Environment Model in three phases: (**Phase 1**) given are the old model and the new observation $\boldsymbol{y}_t$ to be inserted (top left) and its current position estimation $p(\boldsymbol{x}_t)$ (bottom right); (**Phase 2**) at first, similarities between the Gaussians in X_i and $p(\boldsymbol{x}_t)$ are determined; (**Phase 3**) the similarity of observation $\boldsymbol{y}_t$ and known reference views $\boldsymbol{y}_i$ decides how the certainty weights of the Gaussians have to be adapted (see equations (16)-(19))

and the new covariance matrix

$$\begin{aligned} \boldsymbol{C}_{ik} &= \eta_{in}(\boldsymbol{C}_{in} + (\boldsymbol{\mu}_{in} - \boldsymbol{\mu}_{ik})(\boldsymbol{\mu}_{in} - \boldsymbol{\mu}_{ik})^T) \\ &+ \eta_{im}(\boldsymbol{C}_{im} + (\boldsymbol{\mu}_{im} - \boldsymbol{\mu}_{ik})(\boldsymbol{\mu}_{im} - \boldsymbol{\mu}_{ik})^T \end{aligned} \tag{14}$$

The resulting new Gaussian k approximates the weighted sum of the two density functions. Because η_{ik} is increased for each fitting observation, the influence of a single update step decreases with time, and the MoG stabilizes gradually.

3) Adaptation of the Gaussian's weights: Now, the certainty weights w_{ij} of the Gaussians are updated according to the premises introduced above. For this, first the feature distances $E(\boldsymbol{y}_t, \boldsymbol{y}_i)$ and smilarities $S(\boldsymbol{y}_t, \boldsymbol{y}_i)$ in the *Observation Space* are determined as described earlier (see Fig. 3, top left). Furthermore, the similarity of each Gaussian j in X_i to the single Gaussian describing the unimodal current state $p(\boldsymbol{x}_t)$ is determined by means of the Mahalanobis distance (see Fig. 3, bottom left)

$$d(\boldsymbol{\mu}_a, \boldsymbol{\mu}_b) = (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{ij})^T (\boldsymbol{C}_t + \boldsymbol{C}_{ij})^{-1} (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{ij}) \tag{15}$$

Now, the weights w_{ij} of every Gaussian j in all X_i are adapted, if one of the following conditions is fulfilled:

1. *Weight increasing:* If $\boldsymbol{y}_t$ belongs to this reference observation ($E(\boldsymbol{y}_t, \boldsymbol{y}_i) < E^{max}$) and the current position estimation $p(\boldsymbol{x}_t)$ resembles the Gaussian currently considered ($d(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{ij}) \leq d^{max}$) then the weight of this Gaussian is increased with β as learning rate (see Fig. 3, top right - left Gaussian in X_1):

$$w_{ij} := \beta\, S(\boldsymbol{y}_t, \boldsymbol{y}_i) + (1 - \beta\, S(\boldsymbol{y}_t, \boldsymbol{y}_i)) w_{ij} \tag{16}$$

2. *Passive forgetting:* If $\boldsymbol{y}_t$ is matching the reference observation ($E(\boldsymbol{y}_t, \boldsymbol{y}_i) < E^{max}$) but the considered Gaussian j of X_i does not match to $p(\boldsymbol{x}_t)$ ($d(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{ij}) > d^{max}$), then the weight of the respective Gaussian must be decreased as follows (see Fig. 3, top right - right Gaussians in X_1 and X_2):

$$w_{i,j} := (1 - \beta\, S(\boldsymbol{y}_t, \boldsymbol{y}_i))w_{ij} \tag{17}$$

3. *Active forgetting:* If reference observation $\boldsymbol{y}_i$ doesn't match $\boldsymbol{y}_t$ ($E(\boldsymbol{y}_t, \boldsymbol{y}_i) >= E^{max}$) but has a component j at the same (x, y)-position as $p(\boldsymbol{x}_t)$ with ($d^{Euclid_{xy}}(\boldsymbol{\mu}_t, \boldsymbol{\mu}_{ij}) < d^{max}$) then the respective weight has to be decreased too, to gradually forget former observations at this position, which will not appear again (see Fig. 3, top right - left Gaussian in X_3):

$$w_{ij} := (1 - b)w_{ij} \tag{18}$$

where b is a function of distance, and the variances of $p(\boldsymbol{x}_t)$ determine the speed of forgetting:

$$b = \beta_{for} \cdot max\left\{1 - \frac{d^{Euklid_{xy}}(\boldsymbol{\mu}_t, \boldsymbol{\mu}_i)}{d^{for}}; 0\right\} \frac{d^{gain}}{\sigma_x + \sigma_y} \tag{19}$$

To consider uncertainty in the position of the new observation, the forgetting rate is modulated by the variances σ_x and σ_y. So the forgetting of former hypotheses happens slower if the new observation influences a wider spatial area in the *State Space*. The necessary parameters σ_x^2 and σ_y^2 are extracted from the main diagonal of $\boldsymbol{C}_t$. Parameter d^{for} determines the maximum spatial area within forgetting takes place, while d^{gain} controls the dependency of forgetting rate on the variances.

After updating, Gaussians with too low weights are removed in X_i and, as consequence of this, all reference views with empty X_i are deleted. This way, the last update rule realizes a limited number of Gaussians in a restricted local area. Moreover, it guarantees, that old, irrelevant observations at those positions with variable appearances can be replaced by new ones. By means of this forgetting, the complexity of the algorithm, which is determined by the number of reference views, is linear in the area the robot is operating in. If a former observation, however, should appear again at the same position, it can be inserted as a new observation into the environment model again.

3 Experiments and results

We first investigated the general learning and localization capabilities of our algorithm in a static operation area, a part of the home store with a size of 25m by 10m. Figure 6 shows the localization path determined by our CML approach (blue curve, 7 laps around goods shelves). Observable is a localization displacement to the reference positions (gray/yellow curve) growing with distance to the initial position. The reason for this behavior is the erroneous odometry used during the first lap for building the initial environment model.

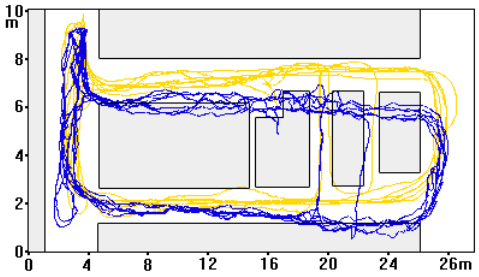

Figure 4: Localization results of our vision-based CML approach in the operation area, a part of the home store. The initial position lies in the upper part of the left hallway (top left). As reference for visualization, the gray/yellow path shows the true position (ground truth), while the dark/blue one depicts the localization estimates.

This is a general problem of this class of CML-approaches, but in the field we want to apply this technique this problem is secondary. In our desired application, the autonomous navigation in a home store, the main task is not to build a model of a completely unknown area but to continuously adapt the model learned before to a changing environment. While building the map for the first time, supplementary localization information could be given to achieve a higher initial precision (e.g., by a manual position calibration at the edges of the operation area).

In the following experiment, the behavior of the actively forgetting environment model in an operation area showing highly dynamic effects is examined. To construct such a demanding situation, we executed a long-term experiment in our institute building. Here, we could actively influence the appearances of local surroundings by switching the lights and closing the curtains. The results of this experiment clearly demonstrate the merits of our model (see table 1). Using a model similar to the one presented in [5], the number of reference views was growing continuously as long as the environment was changed ($\rightarrow$ CML with passive forgetting). Our approach ($\rightarrow$ CML with active forgetting) is able to better handle the situation by replacing irrelevant reference views by new ones, leading to a limited number of references for this restricted operation area. Based on these encouraging lab results, we then investigated the active adaptation of the environment model and the resulting localization accuracy under realistic home store conditions, i.e. with customers and employees walking through the operation area, rearranged or cleared out goods shelves, and other dynamic changes (illumination, etc.). To increase the dynamic effects, two data sets

Table 1: Comparison of the number of learned reference views required for a continuously changing environment (10m x 15m): CML with passive forgetting (similar to [5]) versus CML with active forgetting and replacing of irrelevant reference views.

	localization steps			
	2000	4000	6000	8000
CML w. passive forgett.	1534	2835	4326	5560
CML w. active forgett.	1253	1807	2096	2070

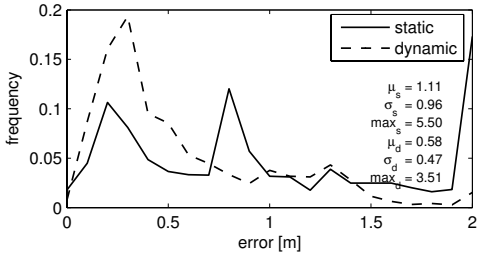

Figure 5: Histogram of localization errors in a dynamically changing environment. The diagram shows that the errors arising from use of a static map are significantly higher (→ static), whereby our dynamic environment model is able to improve the localization accuracy (→ dynamic). Higher values on the left side are a result of more small position errors and correspond to a low average error μ.

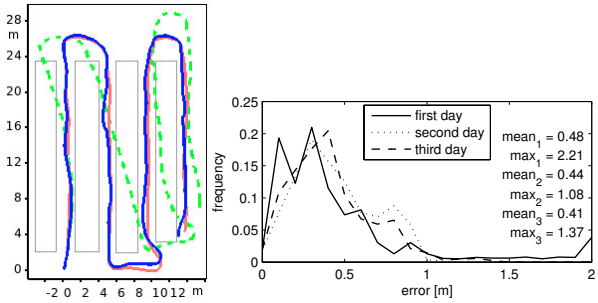

Figure 6: Results of our long-term localization and learning experiment executed in the home store. (Left) As reference for visualization the goods shelves (grey rectangles), true path (red/light grey curve), estimated path (blue/dark grey curve) and odometric data (green/dotted curve) are shown. (Right) The error histogram clarifies the improvements made by the CML algorithm with every new data (from day to day). The average error is improved from 48 cm to 41 cm, the maximum error could be reduced from 221 to 137cm.

were recorded with a time difference of four weeks. For that purpose, the robot was moved about 2,000 meters through the operation area (50 x 80 meters) by joy-sticking. The driven test routes typically have a length of about 300 meters. In the experiment, at first an "old" static map was built up offline with manually corrected odometric data. Afterwards this model was used to localize the robot on data recorded four weeks later. During this time, many locations in the home store have been rearranged which caused a high average position error of 111cm in the beginning (see Fig. 5). After that, the capability of our CML algorithm to adapt the learned map was activated. As a result, the mean localization error could be decreased to 58cm by using the adapted map. Due to the changes in the environment, in the experiment using the non-adapted map numerous localization failures produce high maximum localization errors (see Fig. 5) that influence the average error negatively.

In continuation of the basic experiment described first, the objective of the last experiment was to concurrently localize the robot and to build up the environment model over a longer period of operation. The map was built online without corrected odometric data, and the algorithm started with an empty map. Therefore, it is important that the CML algorithm can recognize already visited areas to create a stable and consistent map. Figure 6 (left) shows the navigation path determined by the CML approach for a relatively large local area. The map was built up and updated while the robot was driven through the environment on three different days. A remarkable result is that the average localization error could be reduced from day to day because of the adapting map (Fig. 6, right). At the end, an average localization error of 41cm was achieved. This experiment shows that it is possible to localize and concurrently build up an environment model with the omnivision-based CML approach. But it is relatively time-consuming to get a consistent map because of the necessity of a repeated observation of the same operation area.

4 Conclusion and future work

To better deal with a dynamic environment, we developed a novel omnivision-based Concurrent Map Building and Localization (CML) algorithm that allows to better handle appearance variations in the environment. In this paper, we introduced the essential extensions and improvements of our approach: a more flexible hypotheses fusion, a distributed coding of the current observation by a weighted set of reference observations, and an advanced learning scheme in combination with an active forgetting to better deal with appearance changes. We conducted a number of encouraging localization experiments investigating the impact of these extensions on the stability and localization accuracy of this CML technique. Based on these results, several long-term experiments are planned for the near future to determine the accuracy of the global localization in the complete store, and to observe the development of the number of reference views used in the environment model with respect to the model complexity and long-term stability.

References

[1] H.-M. Gross, H.-J. Boehme, A contribution to vision-based localization, tracking and navigation methods for an interactive mobile service-robot. in: *Proc. IEEE-SMC 2001*, pp. 672-677

[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-IROS 2003*, pp. 1505-1511

[3] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, "Hierarchical Image-based Localisation for Mobile Robots with Monte-Carlo Localisation." in: *Proc. of 1st European Conference on Mobile Robots (ECMR'03)*, pp. 13-20

[4] J. M. Porta and B. J.A. Kroese, "Appearance-based Concurrent Map Building and Localization," in: *Proc. Int. Conf. on Intelligent Autonomous Systems (IAS'04)*, pp. 1022–1029

[5] J. M. Porta and B. J.A. Kroese, "Appearance-based Concurrent Map Building and Localization using a Multi-Hypotheses Tracker", in: *Proc. of IEEE/RSJ-IROS 2004*, pp. 3424-3429