An Integrated Active Perception Module for a Distributed Cognitive Architecture

Robert Eidenberger, Raoul Zoellner and Josef Scharinger

***Abstract*— This paper presents an active perception planning module embedded in a distributed functional cognitive architecture for complex environments. It discusses the functional module integration over the *wish list concept*, enabling distributed planning, reasoning and decision actions. Further the perception planning approach is depicted along with its components: the probabilistic framework for scene modeling, the correct association of observation data into the scene model, the probabilistic computation of object occlusions and the probabilistic action planning in order to select the most profitable prospective viewpoint.**

In an experimental setting a complete perception loop is presented and the achieved results are discussed. Perception and manipulation tasks in an everyday kitchen environment are chosen as the evaluation scenario. It can be stated that the so far developed active perception planning approach convinces in its applicability to high dimensional state spaces, its quick computation and its efficient planning strategy.

I. INTRODUCTION

This paper focuses on an active perception system for a household service robot acting in everyday environments. Further, the main scope of the perception system is to sustain the manipulation capabilities of the robot. The overall scenario we are looking for is on the one hand the perfect execution of manipulation tasks, and on the other hand the smooth integration of perception and action tasks. The scenario we consider is that our service robot is continuously active, steadily perceiving its surroundings and actively trying to enhance its internal representation of the environment. By doing this, the system will be able to fulfill manipulation tasks (e.g. fetch and carry etc.) in a very natural, reliable and fast way.

The requirements for the perception system are:

- Everyday environment: The system should be able to cope with realistic environments, which are typically cluttered due to the great amount of present objects. No prior assumptions are made about object positions and also unknown objects are considered for manipulation and path planning.
- Locally complete model: This requires the use of scene models which allow a quality assessment of the generated models enabling an active control of the perception. Hence the envisaged active perception must be implemented in a way that information over the local scene (in terms of space) will be maximized continuously.
- Robust and reliable: The used methods should consider uncertainties in all abstraction layers. This requires probabilistic models, which explicitly consider uncertainties and comprehensive noise models.
- Smooth integration in the overall system: The above requirements imply that the perception system has to continuously evaluate its knowledge and strive to maximize it. The required perception actions must be integrated into the overall system control in order to avoid local system optimization.

According to the above requirements the developed active perception system aims on the probabilistic modeling of multi-object scenes, where a scene consists of several items, henceforward called object instances. The number of instances is unknown. Each item instantiates an object class and all different, initially known classes establish the object database. The sensor measurement update comprises the correctly associated measurements with their corresponding object instances and the Bayesian update step.

Based on this probabilistic model we describe a measurement model which does not estimate object class and pose from sensing data only but also by incorporating scene knowledge, such as in particular object occlusions. The proposed concept for occlusion consideration is applied for both the real measurement and the observation prediction

Fig. 2. DESIRE robot performing a *sorting Task* on the CeBIT technology fair in Hannover, Germany.

This work was partly funded as part of the research project DESIRE by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IME01D.

R. Eidenberger and J. Scharinger are with the Department of Computational Perception, Johannes Kepler University Linz, 4040 Linz, Austria robert.eidenberger.ext@siemens.com, josef.scharinger@jku.at

R. Zoellner is with the Department of Intelligent Autonomous Systems, Information and Communication, Siemens AG, 81739 Munich, Germany raoul.zoellner@siemens.com

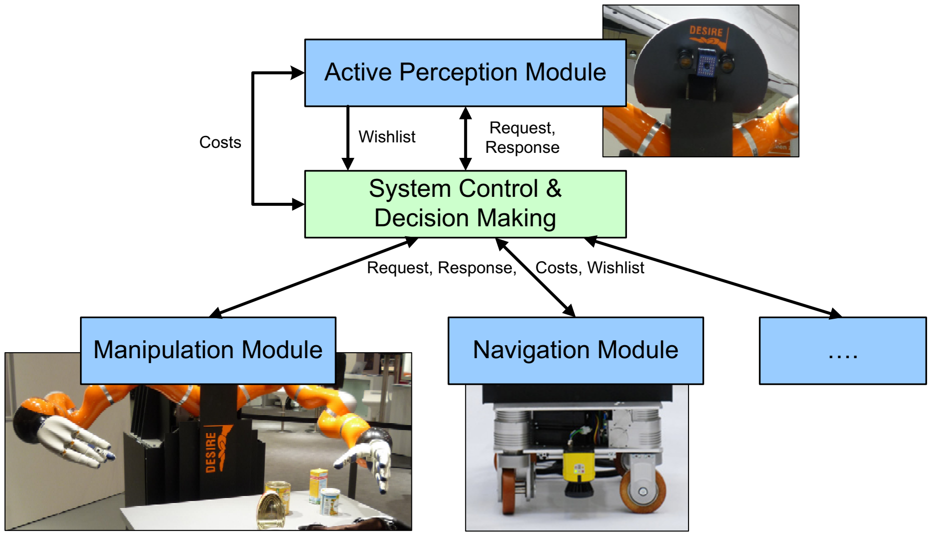

Fig. 1. Wish list concept. The system control and decision making component communicates with several modules via control requests and wish lists.

for perception planning. The probabilistic planner reasons by considering both, information theoretic quality criteria for probability distributions and control action costs.

The outline of the paper is as follows. The next section presents a brief overview of current approaches to multi-object scene modeling, active perception planning and some service robot systems with similar capabilities. Section III outlines the architecture integration and the perception framework, Section IV describes the perception system components in more detail. This paper closes with an experiment in Section V demonstrating a full scene perception loop.

II. RELATED WORK

Many different robotic systems capable of enhanced manipulation and perception skills have been developed so far. In the following we exemplarily list only a few in order to show the large variety. The Mobman [1] is a robot which uses a single arm with a two-finger gripper for grasping objects on the table. The ARMAR system [2] is more sophisticated in its hardware assembly and its perception and communication skills. In [3] manipulation tasks are embedded in the interaction loop with humans. Rollin' Justin [4] focuses more on the control theory for precise and fast actuation than on aspects of autonomous behavior. One of the most popular robots is the humanoid ASIMO [5], which convinces in its moveability.

All these robotic systems differ greatly due to different demands and realizations. However concerning the active perception and smooth integration of perception planning in the overall system control almost all presented systems are very limited.

In the area of active perception planning and especially for planning under uncertainty a partially observable Markov decision process is a very common general model. POMDPs statistically reason over costs for control actions for finding an optimal action policy. A very good overview on online planning algorithms for POMDPs is given in [6]. In [7] a POMDP is used for cost-sensitive feature acquisition and classification. The expected reward is calculated from the classification costs with respect to the current belief state. Spaan [8] suggests extending the planning strategy by combining costs with payoffs coming from information theoretic uncertainty measurements.

The research in active perception brought up very sophisticated approaches for viewpoint evaluation and next best view planning [9][10][11][12]. These works vary in their methods of evaluating sensor positions, in the strategies for action planning and in their field of application. They mainly aim on fast and efficient object recognition of similar and ambiguous objects, but do not cope with multi-object scenarios and cluttered environments. These topics are covered in [13].

III. SYSTEM DESIGN AND ARCHITECTURE

This section presents the architecture of the active perception module and its integration in the overall system.

A. *Perception system integration*

The global system architecture presented in Figure 1 follows the distributed paradigm of integrated cognitive functional modules. The system is based on several functional modules like *Active Perception*, *Manipulation*, *Navigation* etc. All of them perform planning and control actions autonomously according to their local models, knowledge and functional goals.

In order to enable the whole system to perform optimally on the one hand the exchange of local data between the functional modules has to be ensured and on the other hand the negotiation of partial or local goals for each module must be performed in an optimal way. This is achieved by following the *wish list concept*, presented in the following.

1) The wish list concept: The overall system control coordinates operations and assigns tasks to the specific functional modules. If a task cannot be completed the modules respond by sending a wish list, which means, that they suggest solutions in form of actions to properly accomplish the task. The global system control evaluates these wish lists in a reasoning process in order to select the most promising prospective task, which will be assigned as a new task to one or more functional modules.

2) Wish list processing: Since for this paper we mainly focus on the active perception module, the wish list processing is presented in this context. When calling the active perception module a specific request is executed. For instance a list of objects of the actual scene containing position and classification is returned. If the uncertainty in the belief of the actual scene model is unsatisfactorily high the request can not be completed and the perception planner decides to actively acquire more data in order to improve the scene model.

The result of the planning step is a wish list containing a set of possible control actions including their benefits and costs, which is sent to the system control. Based on the evaluation of the overall system state a new request in form of a single action of the wish list is passed to the perception module. Afore other tasks like navigation to a different location in order to change the view point on the focused scene might be accomplished.

The wish list concept is applied to any other module too. For instance the manipulation component suggests new grasping positions if an object cannot be grasped, or the navigation module proposes solution strategies such as removing obstacles if a target cannot be reached.

B. Active Perception Module

The active perception module aims on choosing control actions $a \in A$ to reach a specific goal g. In this paper we only consider sensor positioning at different viewpoints as sensing actions. The framework for selecting the best prospective action policy π is schematically illustrated in Figure 3.

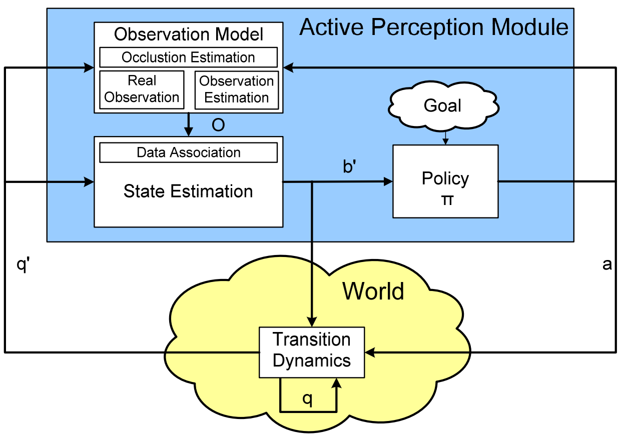

Fig. 3. Active perception framework

In order to find an optimal action policy a sequence of prospective actions and observations has to be evaluated. Decision making bases on the costs of the executed action and the reward from the expected belief $b(q')$, which denotes the conditional probability distribution over the state q', given a sequence of measurements. State estimation determines this belief distribution by updating the initial distribution by incorporating an observation O. The observation model provides the measurement data for state estimation. The expected observation is predicted from the chosen sensing action and the state distribution after the transition update. For more accurate observation prediction, object occlusions in the multi-object scenarios are estimated.

The following sections depict the components of the perception module in more detail.

IV. ACTIVE PERCEPTION COMPONENTS

At first this section depicts the used probabilistic methods of object modeling for describing multi-object environments. Subsequently the active perception components such as state estimation, the observation model and policy making are detailed.

State estimation describes the fusion of measurement data with current scene information and associates observations to corresponding hypotheses. The visibility influences between objects are considered in form of occlusions in the estimation process. Perception planning reasons over estimated belief distributions for finding the best action policy in order to reduce the state uncertainty. For executing actions we consider state transition uncertainties due to unprecise navigation and modeling inaccuracies. All these components are explained in detail in the following sections.

A. Multi-object scene modeling

We consider the single object instance ι with $0 \le \iota \le I$, where I denotes the temporary total number of object instances. Let $b^{\iota}(q)$ be the belief distribution of item ι over the object state $q = (C_i, \phi^T)^T$, which is a tupel containing the discrete class representation C_i and its continuous m-dimensional pose $\phi = (\phi_1, ..., \phi_m)^T$ with $\phi \in \mathbb{R}^m$. C_i is element of the set of object models $(C_0, C_1, ..., C_c)$, which establish the object database C, containing all c different and initially known object classes and the rejection class C_0.

Hence we model the uncertainty of each object instance by individual probability distributions. As each belief distribution integrates to one all items would be equal in its probability mass. To allow assigning different weights to object instance distributions to express the certainty of which this distribution in fact corresponds to an item or not we introduce the rejection class C_0. Assigning weights to this class reduces the probability mass of the other object classes and decreases the occurrence likelihood of this object instance.

In order to consider the total probabilistic scene we introduce $b(q)$ as the set of all instance distributions $b^{\iota}(q)$. Note that $b(q)$ is no probability distribution after all, but a conjunction of distributions.

B. State estimation and data association

This work uses the Bayesian state estimator introduced in [14] and considers uncertainties in the state transition and in the measurement for state estimation. We state the probability distribution over the state

$$b_{t-1}^{\iota}(q) = p^{\iota}(q|O_{t-1}(a_{t-1}), ..., O_0(a_0)) \quad (1)$$

as the a priori belief of the object instance ι for previous sensor measurements $O_{t-1}(a_{t-1}), ..., O_0(a_t)$. Applying an action with its state transition probability $p_a(q'|q)$, which contains the inaccuracies resulting from robot actions, leads to the probabilistic model for the prediction update

$$p_a^{\iota}(q'|O_{t-1}(a_{t-1}), ..., O_0(a_0)) = \int_q b_{t-1}^{\iota}(q) p_a(q'|q) dq. \quad (2)$$

The posterior distribution $b_t^{\iota}(q')$ is calculated according to Bayes' rule by updating the prediction update with the new observation $O_t(a_t)$ for each object instance.

$$\begin{aligned} b_t^{\iota}(q') &= p^{\iota}(q'|O_t(a_t), ..., O_0(a_0)) \quad (3) \\ &= \frac{P^{\iota}(O_t(a_t)|q') p_a^{\iota}(q'|O_{t-1}(a_{t-1}), ..., O_0(a_0))}{P^{\iota}(O_t(a_t), ..., O_0(a_0))} \end{aligned}$$

The evidence term $P^{\iota}(O_t(a_t), ..., O_0(a_0))$ is determined by integrating over the state distribution applying the theorem of total probability

$$\begin{aligned} &P^{\iota}(O_t(a_t), ..., O_0(a_0)) = \quad (4) \\ &\int_{q'} P^{\iota}(O_t(a_t)|q') p_a^{\iota}(q'|O_{t-1}(a_{t-1}), ..., O_0(a_0)) dq'. \end{aligned}$$

The actual measurement model provides the total measurement likelihood $P(O_t(a_t)|q')$.

In this work all probability distributions, including this likelihood, are represented as multivariate Gaussian mixtures. The abilities to describe multifaceted, multi-peaked distributions and their suitability to high dimensional state space due to its parametric computation are favorable.

The measurement likelihood $P(O_t(a_t)|q')$ contains i Gaussian mixture components describing the observation, but does not possess the desired information about the object instance categorization

Thus, this input data is associated with the corresponding object instances, which the probability distribution needs to be fused with. We split up the complex measurement likelihood distribution to simple, single peaked components

$$P(O_t(a_t)|q') = \sum_i P^i(O_t(a_t)|q') \quad (5)$$

with $P^i(O_t(a_t)|q') = \mathrm{N}(w_i, \mu_i, \Sigma_i)$. Weight w_i, mean μ_i and covariance Σ_i are the parameters of the Gaussian kernel. We compare each component with each object instance prior distribution $b_{t-1}^{\iota}(q)$ by applying the Mahalanobis-distance measure

$$d = \sqrt{(\mu_1 - \mu_2)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)} \quad (6)$$

on both distributions. The similarity is defined by a specific threshold dependent on the object class' geometry. More precisely, a function of two parameters, the Mahalanobis distance of the distributions of the object centers and their geometric dimensions, is used for determining object instance allocations. When two distributions are considered to be similar $P^{\iota}(O_t(a_t)|q')$ is set to $P^i(O_t(a_t)|q')$ (or added to it if already one component was assigned before). If a component i cannot be associated with any object instance distribution a new object instance $I+1$ has been found and its distribution is assigned to a new object instance measurement $P^{I+1}(O_t(a_t)|q')$. The corresponding prior distribution for the Bayes update is assumed to be uniformly distributed. The associated object instance likelihood are used for the Bayes' update in Equation (3).

C. Occlusion estimation in the observation model

Generally we want the observation model to estimate the observation likelihood $P(O_t(a_t)|q')$ for the current measurement $O_t(a_t)$. Under the assumption of using interest point detectors this observation can be expressed as the detection of a set of N features

$$O_t(a_t) = \{f_1(a_t), ..., f_N(a_t)\}, \quad (7)$$

as a subset of all database features. These features are considered to be the currently visible interest points.

We generate this set of features explicitly when predicting an observation, where we simulate the measurement. Feature characteristics and occlusion events are considered. While for a real measurement the set of features is acquired directly from the detector, during the simulation of the observation we estimate the visibility of a feature j from its occurrence likelihood $P(f_j(a_t))$. For determining a feature's visibility the pose distribution of the object instance, which the feature belongs to, is of importance. In order to compute the probabilistic visibility we draw S samples from the prior beliefs $b_{t-1}(q)$ of the object instance distributions. Each sample describes a set of states q_s containing one state q_s^{ι} of each instance. Thus, the sample state q_s can be interpreted as one specific object constellation. For this state $P_s(f_j(a_t)|q)$ is calculated taking into account the view frustums of the sensor and the features, features on back faces, and possible object occlusions. Adding up the likelihoods over all samples leads to

$$P(f_j(a_t)) \sim \frac{\sum_{s=1}^{S} P_s(f_j(a_t)|q_k)}{S}, \quad (8)$$

which states an approximation for the probability that the feature $f_j(a_t)$ is visible from viewpoint a_t.

Now, given the set of expected visible features, $P(O_t(a_t)|q')$ is computed by applying the naive Bayes rule and assuming the features to be conditionally independent:

$$P(O_t(a_t)|q') = \prod_j^N P(f_j(a_t)|q'). \quad (9)$$

D. Perception planning

The probabilistic planning concept in form of a partially observable Markov decision process, as proposed in [14], is used for finding optimal action policies. Due to realtime constraints and the fact that performing an observation usually greatly influences the beliefs and makes proposed policies obsolete, this concept is slightly simplified to the 1-horizon planning strategy:

$$\pi(b) = \underset{a}{\operatorname{argmax}}\, R_a(b) \tag{10}$$

The prospective action policy π is determined by maximizing the expected reward

$$R_a(b) = \begin{cases} \int r_a(b)b(q)dq & if\, t < T \\ \alpha h_b(q'|O_t(a_t)) & if\, t = T \end{cases} \tag{11}$$

by applying a Greedy-technique to propose the control action to execute. b is an abbreviation of $b(q)$. The factor α relates the value of information and the sensing costs. The reward $R_a(b)$ for executing action a in state q' is calculated by comparing the sensing costs $r_a(b)$ for consecutively moving the sensing device with the quality of the belief distribution after incorporating an observation, which is performed at time T. In order to determine this quality the information theoretic measure of the differential entropy $h_b(q'|O_t(a_t))$ of the estimated belief distribution is used. Equation (3) describes the calculation of the belief distribution for each object instance. For evaluating the entire scenario the entropy over all objects instances is acquired by summing up the individual entropies

$$h_b(q'|O_t(a_t)) = \sum_{\iota} h_{b^\iota}(q'|O_t(a_t)) \tag{12}$$

As the number of object instances remains constant for all sensing action during planning, this summation is justified.

E. Transition uncertainty in state estimation

The transition uncertainty is defined as the linear Gaussian

$$p_a^\iota(q'|q) = \sum_{k=1}^{K} w_k \mathcal{N}(q|\mu_k + \Delta(a), \Sigma_k(a)), \tag{13}$$

with Gaussian kernels equal in the number of components and mean values to the belief distribution. $\Delta(a)$ indicates the change in state dependent on the action with a covariance $\Sigma_k(a)$.

V. EXPERIMENTAL RESULTS

In this experiments section the detection algorithms, which are used in the observation model, and the modeling of detection uncertainties are described. The experimental setup, including object settings and viewpoint arrangement, is specified and a perception loop is performed. This means that a sequence of measurements is performed which are selected by the proposed action planning algorithm.

A. Uncertainties in the observation model

In this work we use a stereo camera system and the SIFT-detection algorithm [15] for object recognition. In an offline process the object database is build by acquiring 396 images of each object from different viewing angles. All interest points are calculated from this data. In contrast to our previous works [16][14] we do not consider each interest point separately, but use a more abstract representation. We cluster all features from one view to a single, more abstract feature. Thus we get a far less number of total features at the drawback of a loss of geometric information and restrictions in the ability of differentiating between ambiguous objects. This simplification helps for comprehensibility but does not restrict the occlusion estimation process.

The measurement model provides $P(O_t(a_t)|q')$ as a mixture distribution for the state update. The mean values of the measurement distribution are determined from the stereo matching algorithm on the basis of feature correspondences [17]. The uncertainties result from relations between seen and expected interest points, matching errors, sensor and feature characteristics.

B. Experimental setup

The proposed approach is demonstrated with the environmental setup shown in Figure 4. The scenario consists of four different types of grocery objects, a tee box, a rusk box, a sauerkraut tin and a ham tin. In an initial acquisition process the object geometry and the SIFT interest points per view are stored in the object database. The SIFT algorithm finds interest points in each camera image. By comparing these features with the object database we get object hypotheses. From stereo matching of these features based on both camera images we estimate the $6D$ locations of the object hypotheses. In order to express the uncertainty of a measurement the number of consistent stereo matches and the scale of the interest points are taken into account. The uncertainties of the SIFT detector in the distance estimation are usually higher than in their orientation, what results in higher variances in viewing direction.

The measurement model for the state prediction slightly differs from the model from the state update as the observation needs to be simulated. Hence the measurements are simulated by estimating the mean object pose and their average spreading. Therefore we formulate the function $g_v(f_{total}) = k_v f_{total} + \delta_v$ for the coherence between the total number of interest points in a view f_{total} and the number of consistent stereo matches $g_v(f_{total})$ in this view for a detection. The factor k_v denotes the percentage of interest points which are detected on average. δ_v describes additive Gaussian noise. These descriptions and the results of the occlusion estimation process are used for determining the object class likelihood and pose covariance in the simulated measurement model.

A sensing action is defined as a change in the robot's viewpoint. The viewpoint arrangement of 9 different control actions is illustrated in Figure 4, where each blue cone represents a viewpoint pointing in the direction of the tip

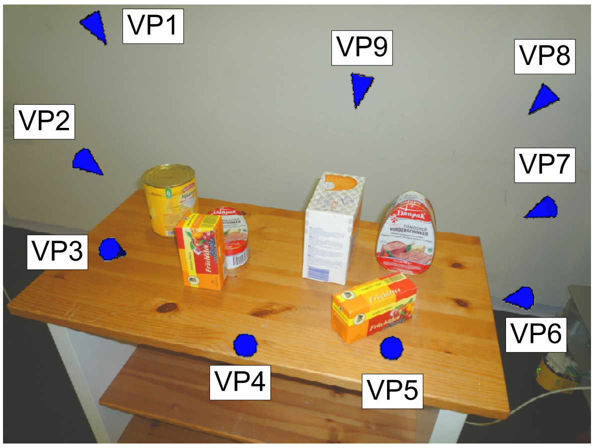

Fig. 4. Experimental scenario with viewpoint arrangement. The blue cones, whose tips point into sensing directions, represent the viewpoints.

of the cone. These actions are evaluated in this experiment. We derive the costs of the robot's movement to another viewpoint from the movement angle and distance. The costs for accomplishing a new observation without moving are set to a value greater than 0 for the current viewpoint to avoid remaining at the same location at all times. The measurement reward is calculated according to Equation (12) from the entropy of the estimated distribution.

C. Performing a perception loop

In the following an action sequence resulting from the proposed planning approach is presented. Figure 5 shows the detection result in each step.

Initially we do not have any scene knowledge, so each action promises identical benefits. Thus the first measurement could be acquired from any viewpoint. In this experiment we start from the current robot position, namely viewpoint 6. A total of 5 new object instances are detected. The first column in Figure 5 illustrates the result for this measurement. The top image shows the acquired image by the left camera. The middle plot illustrates the projections onto the ground plane of the translational covariance ellipsoids, which cover 97 percent of the probability mass. The bottom plot depicts a 3-dimensional side view, almost from the same viewpoint as the image in Figure 4 was taken. Again the translational covariance ellipsoids are shown together with the table board which the objects are positioned on. While four objects are detected very well, the uncertainty of the ham tin is significantly higher. This results from the bad viewing angle and the shadow on a significant object surface. The tea box in the back was not detected at all. In the sequencing planning phase the uncertainty needs to be reduced. Viewpoint 4 is suggested as the best prospective action to accomplish. This is because it is quite close to the current viewpoint and allows seeing all objects, as none would be occluded. Viewpoint 5 is declined due to possible occlusion by either the tea or the rusk box. Moving counterclockwise around the table seems promising only for reducing the uncertainty of the ham tin, but not for the other objects such as the sauerkraut and ham tin on the left.

The desired effect of a more precise location of the right ham tin is achieved after performing an observation from viewpoint 4. As all other object instances are detected too, their distributions are sharpened. However a new tea box object instance is found. The tea box distribution consists of a two component mixture distribution, both of about the same weight, but one with a far larger covariance. The next planning step aims at increasing the position knowledge of this tea box and on optimizing all other belief distributions and makes the robot move to viewpoint 5. The resulting belief contains six well located object instances. As the specified accuracy is reached, the algorithm terminates.

This example shows only one possible action sequence for starting from viewpoint 6. Due to the noise effects, coming either from the real world influences or being modeled in prediction phase, measurements and predictions vary and lead to different strategies. It is not guaranteed that all object instances are found at all times, as the algorithm tries to reduce uncertainties of knows beliefs, but does not explore space up to now.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents an active perception module and its embedding in an overall system control of a service robot, which operates in household environments. The wish list concept is introduced as a main part of the proposed distributed architecture of cognitive functional modules. Further the implementation of the active perception module based on estimating and selecting best prospective viewpoints in multi-object environments by considering perception benefits and action costs is discussed. Some aspects of the fully probabilistic framework, which works in continuous high-dimensional domains, are described at the example of components of the perception module. Since the actual implementation lacks the capability of the detection of white spots in the actual scene model further research will tackle this issue.

REFERENCES

[1] G. v. Wichert, C. Klimowicz, W. Neubauer, T. Woesch, G. Lawitzky, R. Caspari, H.-J. Heger, P. Witschel, U. Handmann, and M. Rinne, “The robotic bar - an integrated demonstration of man-robot interaction in a service scenario,” in *Proceedings of the 2002 IEEE Int. Workshop ROMAN, Berlin*, 2002.

[2] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schroeder, and R. Dillmann, “Toward humanoid manipulation in human-centred environments,” *Robotics and Autonomous Systems*, vol. 56, pp. 54–65, 2008.

[3] E. Sisbot, L. Marin-Urias, R. Alami, and T. Simeon, “A human aware mobile robot motion planner,” *IEEE Transactions on Robotics*, vol. 23 (5), pp. 874–883, 2007.

[4] C. Ott, O. Eiberger, W. Friedl, B. Baeuml, U. Hillenbrand, C. Borst, A. Albu-Schaeffer, B. Brunner, H. Hirschmueller, S. Kielhoefer, R. Konietschke, M. Suppa, T. Wimboeck, F. Zacharias, , and G. Hirzinger, “A humanoid two-arm system for dexterous manipulation,” in *Proceedings of IEEE-RAS International Conference on Humanoid Robots*, Genova, Italy, 2006, pp. 276–283.

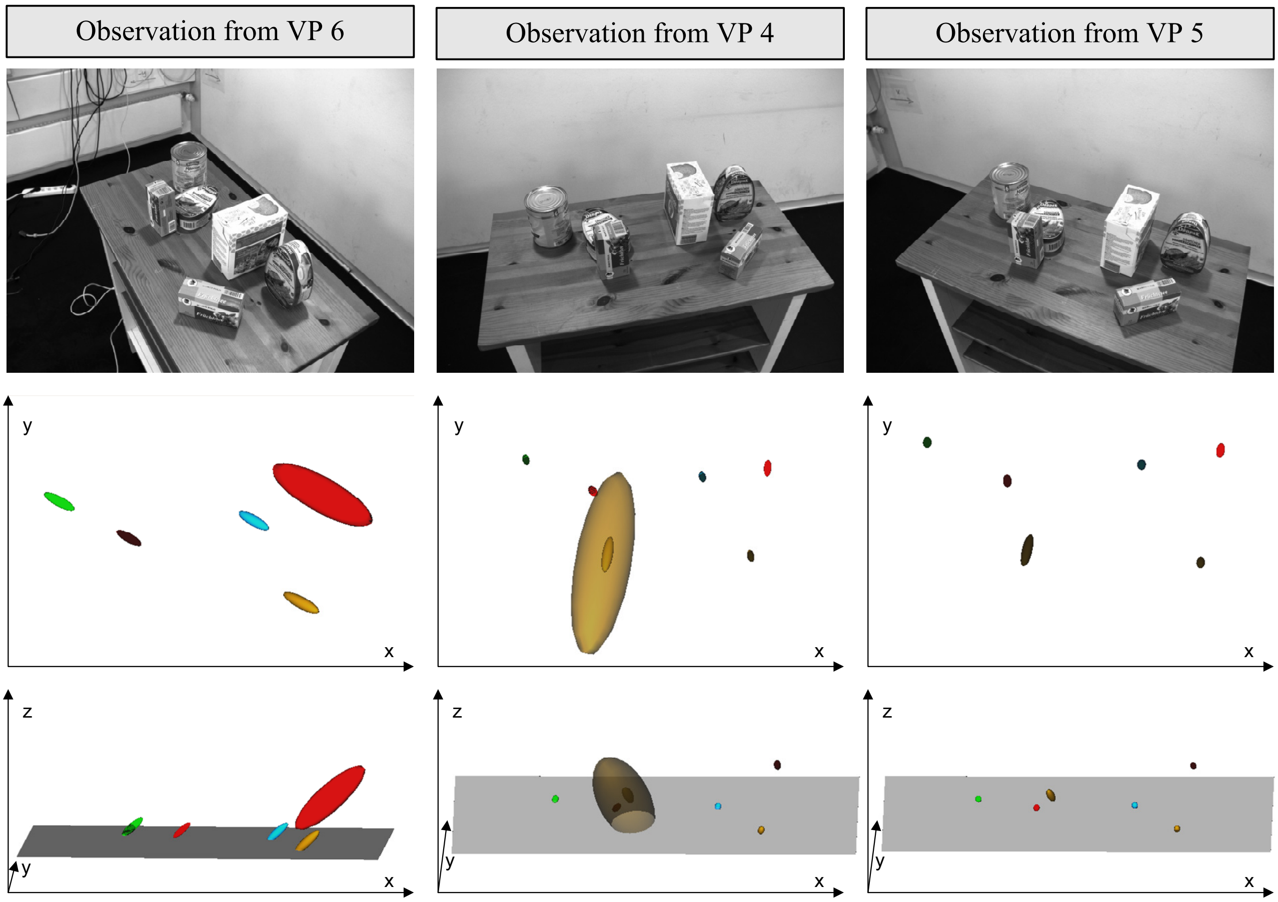

Fig. 5. Probabilistic representation of detection results for the selected sequence of control actions. The top plots picture the acquired camera images from the selected viewpoints, the middle plots show the translational covariance ellipsoids representing the object instance's pose uncertainties. The bottom plots illustrate a 3-dimensional view of the covariance ellipsoids and the table board for better clearness. The tee box is colored brown, the rusk box light blue, the sauerkraut tin green and the ham tin red.

[5] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent asimo: system overview and integration," in *IEEE/RSJ International Conference on Intelligent Robots and System*, vol. 3, 2002, pp. 2478–2483.

[6] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa, "Online planning algorithms for POMDPs," *Journal of Artificial Intelligence Research*, vol. 32, pp. 663–704, 2008.

[7] S. Ji and L. Carin, "Cost-sensitive feature acquisition and classification," *Pattern Recognition*, vol. 40, pp. 1474–1485, 2007.

[8] M. T. J. Spaan, "Cooperative active perception using POMDPs," in *AAAI 2008 Workshop on Advancements in POMDP Solvers*, 2008.

[9] J. Denzler and C. M. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 145–157, 2002.

[10] J. Vogel and N. de Freitas, "Target-directed attention: sequential decision-making for gaze planning," in *International Conference on Robotics and Automation*, 2008.

[11] M. Chli and A. J. Davison, "Active matching," in *European Conference on Computer Vision*, 2008.

[12] K. Deguchi and H. Ohtsu, "An information theoretic approach for active and effective object recognitions," *International Conference on Pattern Recognition*, vol. 2, p. 622, 2006.

[13] R. Eidenberger, R. Zoellner, and J. Scharinger, "Probabilistic occlusion estimation in cluttered environments for active perception planning," in *Proceedings of the IEEE International Conference on Advanced Intelligent Mechatronics*, 2009.

[14] R. Eidenberger, T. Grundmann, and R. Zoellner, "Probabilistic action planning for active scene modeling in continuous high-dimensional domains," in *Proceedings of the IEEE International Conference On Robotics and Automation*, 2009.

[15] D. G. Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision*, 1999, pp. 1150–1157.

[16] R. Eidenberger, T. Grundmann, W. Feiten, and R. Zoellner, "Fast parametric viewpoint estimation for active object detection," in *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 309–314.

[17] T. Grundmann, R. Eidenberger, R. Zoellner, Z. Xue, S. Ruehl, M. Zoellner, R. Dillmann, J. Kuehnle, and A. Verl, "Integration of 6d object localization and obstacle detection for collision free robotic manipulation," in *Proceedings of the IEEE International Symposium on System Integration*, 2008.