Active Perception and Scene Modeling by Planning with Probabilistic 6D Object Poses

Robert Eidenberger and Josef Scharinger

Abstract— **This paper presents an approach to probabilistic active perception planning for scene modeling in cluttered and realistic environments. When dealing with complex, multi-object scenes with arbitrary object positions, the estimation of 6D poses including their expected uncertainties is essential. The scene model keeps track of the probabilistic object hypotheses over several sequencing sensing actions to represent the real object constellation.**

To improve detection results and to tackle occlusion problems a method for active planning is proposed which reasons about model and state transition uncertainties in continuous and high-dimensional domains. Information theoretic quality criteria are used for sequential decision making to evaluate probability distributions. The probabilistic planner is realized as a partially observable Markov decision process (POMDP).

The active perception system for autonomous service robots is evaluated in experiments in a kitchen environment. In 80 test runs the efficiency and satisfactory behavior of the proposed methodology is shown in comparison to a random and a step-aside action selection strategy. The objects are selected from a large database consisting of 100 different household items.

I. INTRODUCTION

Service robots - together with their perception systems - should be able to operate in everyday environments. In practice this expectation certainly is not met today. Inaccurate sensing devices, bad classifiers, object occlusions, poor lighting or ambiguities of object models limit the detection capabilities. Active perception approaches aim at deliberately utilizing appropriate sensing settings to gain more information about the scene.

In this paper we present an approach to probabilistic scene modeling and sensor action planning. Object hypotheses represented as probabilistic distributions are kept in the scene model. A hypothesis contains object class and pose information. The 6D pose representation describes the position and orientation of the object.

The proposed active planning methodology uses the POMDP concept to find an efficient strategy to satisfactorily recognize all objects in a scene. The predicted effects of future sensing actions on the scene model are compared on the basis of the expected information gain of state distributions. The next best sensing action is determined and executed to improve the scene knowledge.

The scene analysis described in this paper is a component of a service robot which autonomously performs manipulation tasks such as grasping objects, where precise knowledge about the object setting is required. Figure 1 shows the robot and some of the used household items.

Fig. 1. The service robot, operating in a complex environment.

The outline of the paper is as follows. The next section gives an overview of current approaches to active perception with focus on high-dimensional continuous pose modeling. Section III describes the concepts of scene modeling with probabilistic 6D poses. In Section IV and V the active perception architecture and its realization is detailed. The paper closes with experiments in Section VI which demonstrate a full perception loop and compare the outcomes of three planning strategies for various scenes.

II. RELATED WORK

Literature provides many approaches to active recognition and next best view planning. Surveys on perception planning by Chen[1] and Dutta-Roy[2] list the current state of the art. Here we discuss active perception approaches with emphasis on probabilistic continuous pose modeling and their applicability to complex scenarios according to their relevance for this paper.

The research on active perception varies with respect to the methods of evaluating sensor positions, the strategies for action planning and the field of application. It mainly aims at fast and efficient object recognition of similar and ambiguous objects, but does not cope with multi-object scenarios and cluttered environments. Forssen et al.[3] combine object recognition with an attention mechanism for obstacle avoidance to efficiently acquire scene information. This approach targets on rapidly identifying objects in cluttered environments, but neither models pose uncertainties nor takes

R. Eidenberger is with Information and Automation Technologies, Siemens AG, 81739 Munich, Germany robert.eidenberger.ext@siemens.com

J. Scharinger is with the Department of Computational Perception, Johannes Kepler University Linz, 4040 Linz, Austria josef.scharinger@jku.at

into account object occlusions. Lanz[4] describes a Bayesian framework for robust multi-object tracking in presence of occlusions. He takes into account the object visibility for computing the observation likelihood, but uses discrete state spaces.

Many authors claim the possible applicability of their active approaches to continuous, high-dimensional domains. Chli and Davison[5] show $2D$ active feature matching in continuous domains. Flandin and Chaumette[6] present an approach to active $3D$ object reconstruction by using spatial voxel representations. However, representations for orientations are not required.

The usage of $6D$ continuous state spaces for representing object poses requires adequate probabilistic representations to model pose uncertainties. Due to the theoretical and computational complexity of using multivariate distributions for describing orientations and positions, the use of sampled distributions in form of particle representations seems reasonable. However, their suitability to high-dimensional problems is questionable because of the great amount of particles needed. The practical applicability is generally only shown in simplified experiments with low-dimensional state spaces. Denzler and Brown[7] and Derichs et al.[8] formulate their solutions for continuous domains, but only demonstrate their active object class identification methodology and neglect pose determination. Ma and Burdick[9] present a methodology for 6-DOF pose estimation and tracking to actively recognize moving objects. However, the relevant aspects of probabilistic modeling are not detailed. In their experiments only a planar problem is considered, whereby $6D$ pose modeling is avoided.

In this work scenarios are considered, which contain arrangements of several objects, belonging to different as well as to alike classes. Their pose uncertainties are represented by $6D$ multivariate probabilistic distributions.

III. PROBABILISTIC SCENE MODELING

Each real item, which is considered for the recognition process, is modeled by its characteristics such as geometry information or textures. All object models - or also denoted as object classes - build up the object database. In a scene various instances of these object classes might appear. Thus, the state space consists of n object hypotheses each represented by the tupel $q^i = (C^i, \phi^i)$. C describes its discrete class representation and ϕ its continuous pose. All n object-tuples build up the joint state $q = (q^1, q^2, ..., q^n)$. The entities are assumed to be mutually independent. Since q^i represents both discrete and continuous dimensions it will be further considered as a mixed state. The dimension of the state space varies as the perception process proceeds with recognizing new object instances.

The following section details the modeling of the pose and the probabilistic distributions over object hypotheses.

A. Probabilistic 6D Pose Modeling

In order to represent real world environments, a six dimensional pose representation is required to model the position and orientation of an object. In [10] we discussed several approaches of probabilistic $6D$ pose representations. In this work we decided to use a Rodrigues vector (or axis/angle) representation to describe the orientation because of its good applicability and simple comprehensibility.

The first three components of the pose vector $\phi^i = (x, y, z, \theta e^1, \theta e^2, \theta e^3)$ [11] describe the translational position. The orientation is represented by the axis and the angle. $(e^1, e^2, e^3)^T$ denotes the unit vector of the rotation axis. The length of the axis encodes the rotation angle $\theta \in]0; 2\pi]$, which the coordinate frame is rotated about. However, the Rodrigues vector representation has drawbacks such as duality and singularities, which cause problems for their mathematical processing. Duality means that one specific orientation can be described by two different Rodrigues vectors. This results from the fact that a rotation of θ radians about an axis equals rotating $2\pi - \theta$ radians about an axis pointing into the opposite direction. This effect is dealt with by setting a working point with respect to the orientation what possibly requires to use the pose representation with the Rodrigues vector in the opposite direction. A singularity arises at $\theta = 0$ radians.

In the simplest case the pose uncertainty is modeled by a multivariate Gaussian distribution with the pose vector ϕ^i as mean, where one 3D part consists of the components of the translation vector and the other 3D part of the components of the Rodrigues vector. The formulation of Gaussian distributions over the translational dimensions is simple due to the real, continuous and infinite domains. For a reasonable modeling in the rotational space the angular characteristics of periodicity, duality and singularities have to be considered. The probability density function of the rotational space would ideally be defined over the finite halfsphere with radius $0 < r_{HS} \leq 2\pi$. Nonetheless we use the infinite space for computational efficiency, but apply working point selection on the distributions. Generally, the representation as a single Gaussian is adequate for describing peaked distributions, where most of the probability mass lies within this halfsphere. In order to represent more complex, multi-peaked distributions, linear combinations of multivariate Gaussians are formulated as the multivariate Gaussian mixture distribution

$$p(\phi^i) = \sum_{k=1}^{K} w_k^i \mathcal{N}(\phi^i | \mu_k^i, \Sigma_k^i), \tag{1}$$

which consists of K Gaussian kernels. The mixing coefficient w_k^i denotes the weight of the mixture component with $0 \leq w_k^i \leq 1$ and $\sum_{k=1}^{K} w_k^i = 1$. μ_k^i is the mean and Σ_k^i the covariance of kernel k.

B. Probabilistic State Representation

The probabilistic joint state space $p(q)$ equals the product of all object instance distributions $p(q^i)$ as they are assumed to be independent. $p(q^i)$ is composed of the pose model $p(\phi^i|C^i)$ and the object class probability $P(C^i)$:

$$p(q^i) = P(C^i)p(\phi^i|C^i) \tag{2}$$

The probability distribution over the class is discrete and is described by a histogram over the object classes. The $6D$ pose space from Equation (1) of the continuous pose distribution $p(\phi^i|C^i)$ is conditioned on the object class, meaning that object class information, in particular the model origin, is required as reference for the pose distribution.

IV. ACTIVE PERCEPTION ARCHITECTURE

Active perception is a process where various actions $a_t \in A$ are compared to each other and executed in order to achieve intelligent sensing to build up to a good world model. From the wide range of different actions we only consider sensor positioning at different viewpoints as sensing actions in this paper. A viewpoint comprises both position and orientation of the sensor. The active perception architecture of the proposed approach for next best view selection, containing the *Observation Model*, the *Inference Model* and the *Planning Module*, is schematically illustrated in Figure 2.

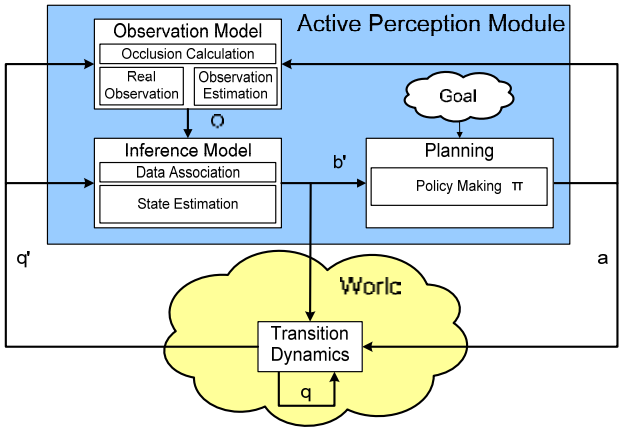


Fig. 2. Active perception framework

In order to find an optimal action policy π a sequence of prospective actions and observations is evaluated by the planning module. Decision making is based on the costs of the executed action and the expected reward from the belief $b_t(q')$, which denotes the conditional probability distribution over the state q', given a sequence of measurements. The inference module consists of data association and state estimation. It determines the belief distribution $b_t(q')$ by updating the initial distribution by incorporating an observation $O_t(a_t)$. 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. Object occlusions in multi-object scenarios are handled for more accurate observation prediction.

The following sections describe the probabilistic modeling of multi-object scenes and the Bayesian statistical framework. Also the observation model under consideration of occlusion estimation and probabilistic action planning are explained in detail.

A. Inference Model

In this approach we decided on a Bayesian state estimator. The system properties for the dynamic environment follow the equations

$$q' = g(a_t, q, \varepsilon_t) \quad (3)$$

$$O_t = h(a_t, q', \delta_t). \quad (4)$$

The index t denotes the time step, which covers the period of incorporating one action-observation pair into the state update. Often the system state q cannot be addressed right after the control action but only after the measurement. Thus the apostrophe $'$ marks the state transition due to action effects.

In the following the state estimation process for the joint state is derived following the dynamic environment model. Equation (3) describes that reaching a future state q' depends on the previous state q and on the applied action a_t. The system dynamics underlie the state transition uncertainty ε_t, meaning that each executed action influences the state distributions. We consider the probability distribution over the state

$$b_{t-1}(q) = p(q|O_{t-1}(a_{t-1}), ..., O_0(a_0)) \quad (5)$$

as the a priori belief for previous sensor measurements $O_{t-1}(a_{t-1}), ..., O_0(a_0)$ at time step $t-1$. Applying an action with its state transition probability $p_a(q'|q)$ containing ε_t leads to the probabilistic model for the prediction update

$$\begin{aligned} b_{t-1}(q') &= p(q'|O_{t-1}(a_{t-1}), ..., O_0(a_0)) \\ &= \int_q p_{a_t}(q'|q) b_{t-1}(q) dq. \end{aligned} \quad (6)$$

The measurement Equation (4) describes the influence of measurement uncertainties δ_t on the observation $O_t(a_t)$ which is performed at time step t after executing the control action a_t. In Bayesian context the measurement update is formulated as the probability distribution $P(O_t(a_t)|q')$.

A recursive Bayesian state estimator is used to calculate the posterior distribution $b_t^{O_t(a_t)}(q')$ by updating a priori information by the measurement $O_t(a_t)$. Combining the models of the transition update and the measurement update using Bayes' rule leads to

$$\begin{aligned} b_t^{O_t(a_t)}(q') &= p(q'|O_t(a_t), ..., O_0(a_0)) \\ &= \frac{P(O_t(a_t)|q') b_{t-1}(q')}{\int_{q'} P(O_t(a_t)|q') b_{t-1}(q') dq'}. \end{aligned} \quad (7)$$

The rules of probability, the Markov assumption and the theorem of total probability are applied to derive this expression. The observation $O_t(a_t)$ is assumed to be conditionally independent of previous measurements.

B. Observation Model

The observation model encapsulates all processes from data acquisition to providing the joint measurement likelihood distribution $P(O_t(a_t)|q')$ for the current measurement

$O_t(a_t)$. Under the assumption of using interest point detectors this measurement can be expressed as the detection of the set of F features

$$O_t(a_t) = \{f_1(a_t), ..., f_F(a_t)\} \tag{8}$$

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

While for a real measurement the set of features is acquired directly from the detector, we generate this set explicitly when predicting an observation, where the measurement is simulated. Feature characteristics and occlusion events are considered to determine the visibility of features. In [12] the methodology of occlusion calculation is detailed. Given the set of (expected) visible features and in consideration of the measurement uncertainty we formulate the likelihood of seeing the feature $f(a_t)$ as $P(f(a_t)|q')$. The measurement likelihood distribution $P(O_t(a_t)|q')$ is computed as the product of conditional feature likelihoods by applying the naive Bayes rule and assuming all features to be conditionally independent:

$$P(O_t(a_t)|q') = \prod_j^F P(f_j(a_t)|q'). \tag{9}$$

C. Perception Planning

Sequential decision-making consists of the processes of evaluating future actions and finding the best action sequence with respect to a specific goal. The probabilistic planning concept is realized in form of a partially observable Markov decision process, as proposed in [13]. The probabilistic planner reasons by considering information theoretic quality criteria of the expected belief distribution $b_t^{O_t(a_t)}(q')$, which is abbreviated by b' in the following equations, and control action costs $r_{a_t}(b')$. The objective lies in maximizing the long term reward of all executed actions and the active reduction of uncertainty in the belief distributions. The value function

$$V_t(b') = \max_{a_t} \left(R_{a_t}(b') + \gamma \int V_{t-1}(b') P(O_t(a_t)|q') dO_t \right) \tag{10}$$

with $V_1(b') = max_{a_t} R_{a_t}(b')$ is a recursive formulation to determine the expected future reward for sequencing actions. γ denotes the discount rate for penalizing later actions and $R_{a_t}(b')$ is the reward. The continuous domains and the high-dimensional state spaces make the problem intractable. As the value function is not piecewise linear, it is evaluated at specific positions, which demands the online calculation of the reward for these specific actions and observations.

The control policy

$$\pi(b') = \underset{a_t}{\operatorname{argmax}} \left(R_{a_t}(b') + \gamma \int V_{t-1}(b') P(O_t(a_t)|q') dO_t \right) \tag{11}$$

maps the probability distribution over the states to actions. Assuming a discrete observation space the integral can be replaced by a sum.

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

$$R_{a_t}(b') = -\alpha E_O[h_{b_t}(q'|O_t(a_t))] + \int r_{a_t}(b') b(q) dq, \tag{12}$$

which relates benefits and costs of future actions a_t with the relation factor α. The first term states the expected benefit of applying the control action, the second term expresses the respective costs with $r_a(b')$ denoting the action efforts. In perception problems the quality of information is usually closely related to the probability density distribution over the state space. The information theoretic measure of the differential entropy is suitable for determining the uncertainty of the belief distribution. Since the computation of the differential entropy both, numerically or by sampling from parametric probability density distributions is costly in terms of processing time, the sum of the upper bound estimates over the object instances

$$\begin{aligned} h_{b_t}^U(q'|O_t(a_t)) &\geq h_{b_t}(q'|O_t(a_t)) \\ &= \sum_i \sum_{k=1}^{K^i} w_k^i [-log w_k^i + \frac{1}{2} log((2\pi \exp)^D |\Sigma_k^i|)] \end{aligned} \tag{13}$$

is used to approximate and determine the expected benefit [14]. D denotes the dimension of the state, $|\Sigma_k|$ denominates the determinant of the kth component's covariance matrix.

V. REALIZATION OF THE ACTIVE FRAMEWORK

This section details the basic concepts by applying them on a robotic scenario. As a sequence of observations from various viewpoints is fused, the measurements have to be correctly associated with the corresponding object instances of the prior knowledge. The essential process of data association in the inference model is described in this section as well as the modeling of system and measurement uncertainties and the realized planning concept.

A. Data association

State estimation in the high-dimensional joint space is challenging. Thus, we factor the joint space into subspaces, each describing one object instance distribution. The Bayes' update Equation (7) can almost be directly applied, except for the measurement likelihood $P(O_t(a_t)|q')$. It has to be decomposed and associated with the priors of the corresponding object instances. The process of multi-target data association aims at finding the assigned measurement likelihoods $P(O_t^i(a_t)|q')$. In this work we combine global nearest neighbor (GNN) data association and geometry-based data association to find corresponding measurement components. Both build up association tables. GNN data association uses the Mahalanobis distance measure to probabilistically compare Gaussian kernels of pose distributions. This is only applicable for components of the same object classes. Geometry-based data association accomplishes the association task over classes by checking object constellations of physical plausibility, meaning objects instances must not intersect. Therefore, samples are draw from the distributions

and each sample arrangement is checked for intersections according to their object geometries.

If the entries of the association tables are within the validation gate, the corresponding measurements are associated. Otherwise, unassigned measurements establish a new object instance distribution which is fused in the Bayes' update with a uniform prior, resulting in an increase of the dimension of the joint state.

B. Measurement uncertainties

In this work the object recognition process uses local interest points from the SIFT algorithm as features for object classification and pose determination. From stereo images spatial information is gained and $6D$ poses are calculated, which is detailed in [15].

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. In order to determine the measurement accuracy of the object pose we evaluated the detection precision of several objects in a total of 260 measurement against ground truth data. The deviation of the translational components is plotted in Figure 3 with the sensor pointing into z-direction. The covariance ellipsoid of this point cloud is used to approximate the measurement likelihood uncertainty. The uncertainty of the object class recognition is encoded in the mixture weight. It results from relations between seen and expected interest points, matching errors, sensor and feature characteristics.

The measurement model for the state prediction slightly differs from the model for the state update as the observation needs to be simulated. The mean object pose and the average spreading is determined from a heuristic based on training data. The interest points of 360 stereo images per objects are calculated, matched and stored to build up the object database. The $3D$ locations of the features are mapped onto the object geometry to generate the database model. During measurement simulation we determine the expected visible features and estimate the consistent stereo matches, which are used together with the results of the occlusion estimation process to adapt the covariances of the simulated measurement components. While the shape of the ellipsoid of the translational components is modeled according to the detection results from Figure 3, the rotational covariance axes are set to equal length.

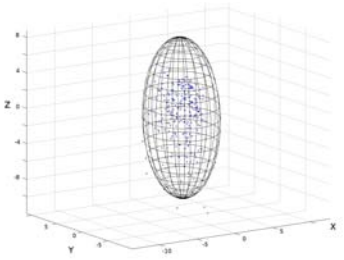
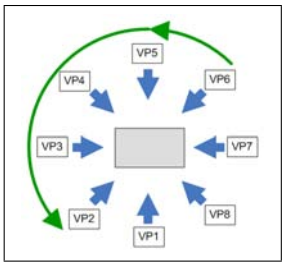


Fig. 3. Measurement uncertainty in xyz-directions with the covariance ellipsoid (97%quantile). The sensor points into z-direction.

(a) Multi-object scenario (image taken from VP1)



(b) Viewpoint arrangement and executed action sequence

Fig. 4. Experimental setup

C. State transition uncertainties

The transition uncertainty is defined as the linear Gaussian mixture

$$p^{\iota}_{a_t}(q'|q) = \sum_{k=1}^{K} w_k \mathcal{N}(q|\mu_k + \Delta(a_t), \Sigma_k(a_t)), \qquad (14)$$

with Gaussian kernels equal in the number of components and mean values to the belief distribution. $\Delta(a_t)$ denotes the change in the mean dependent on the action with covariance $\Sigma_k(a_t)$. Ground truth experiments have shown standard deviations due to state transitions of $16.8mm$ in xy- and $19mm$ in e^3-direction for experiments on the service robot. The other dimensions are negligible.

D. Greedy planning

The incorporation of observations usually changes the belief distribution enormously and possibly makes the current action policy ineffectual. Hence a 1-horizon planning strategy is more efficient for the considered scenarios which can be achieved by applying a Greedy-technique to plan an action sequence until a measurement is performed. The iterative planning algorithm terminates when the desired quality criteria in form of distribution entropies are reached. In this work the differential upper bound entropy is estimated and evaluated for each object instance separately, aiming at detecting each object instance with a specific precision.

VI. EXPERIMENTS

In this section the proposed approach is evaluated. A perception cycle is demonstrated with focus on the localization process. The performance of the developed active planning approach is compared to two other strategies for viewpoint selection.

A. Localization process in the perception cycle

1) Experimental setup: The proposed approach is demonstrated on the example of the object constellation shown in Figure 4a. The complex scenario consists of 11 different objects belonging to 9 object classes. The object database contains a total of 100 different object classes, all household items.

The viewpoint arrangement consists of 8 different, circularly aligned viewpoints and is schematically illustrated in Figure 4b. A sensing action is defined as a change in

TABLE I

DETECTION RESULTS: NUMBER OF RECOGNIZED OBJECTS AND NUMBER OF COMPONENTS OF MIXTURE DISTRIBUTIONS

Observation from VP	Recognized objects		Mixture components		
	class	accurate pose	$b_{t-1}(q)$	$O_t(a_t)$	$b_t(q')$
VP 6	7	4	0	18	9
VP 5	10	7	9	22	12
VP 2	11	11	12	17	13

TABLE II

CORRELATING COSTS AND VALUE FOR CALCULATING THE REWARD R_{a_t} FROM EQUATION (12) FOR SELECTED VIEWPOINTS FOR $\alpha = 2.0$

View-points	VP6	VP6			VP5		
	R_{a_t}	costs	value	R_{a_t}	costs	value	R_{a_t}
VP1	-0.75	-0.75	-1.0	-2.75	-1.00	-0.53	-2.06
VP2	-1.00	-1.00	-0.00	-1.0	-0.75	-0.00	**-0.75**
VP3	-0.75	-0.75	-0.95	-2.65	-0.50	-0.65	-1.80
VP4	-0.50	-0.50	-0.88	-2.26	-0.25	-0.28	-0.81
VP5	-0.25	-0.25	-0.24	**-0.73**	-0.50	-0.24	-0.94
VP6	**0.00**	0.50	-0.13	-0.76	-0.25	-0.49	-1.23
VP7	-0.25	-0.25	-0.28	-0.81	-0.75	-0.64	-2.03
VP8	-0.50	-0.50	-0.14	-0.78	-0.50	-1.00	-2.50

the robot's viewpoint, equivalent to moving the sensor to a different location. These actions are evaluated in this experiment on the basis of the predicted belief distributions. We derive the costs of the robot's movement to another viewpoint from the movement angle and distance.

2) Perception Cycle: An action sequence resulting from the proposed planning approach and its effects on the probabilistic scene model are presented in this section.

Initially we do not have any scene knowledge, so each action promises identical benefits. We start from the current robot position, namely viewpoint 6. After performing the first measurement and accomplishing the data association and state estimation process we retrieve the probabilistic scene distribution consisting of a total of 7 detected object instances, 3 of them of unsatisfactory pose accuracy, though. Table I lists the recognition results and the number of mixture components of the probability distributions during the state update. Figure 5a displays characteristics of the measurement process. The top left image shows one of the stereo images acquired by the camera. The bottom left graphics illustrate the current scene with object models positioned at the means of the corresponding probability distributions. The middle plots show the horizontal and upright projection of the translational covariance ellipsoids of each mixture component of all object instance distributions. The right plot pictures the translational and rotational covariance ellipsoids in a single graphic. The axis of the orientation of each mixture component is drawn in black as a vector of unit length, originating from the translational mean. The original length representing the rotation angle is colored. At the tip of the unit vector the covariance ellipsoid of the Rodrigues components is attached. The weight of the mixture component is visualized by the transparency of the ellipsoids. The table board is shown in grey. All graphics are drawn from the perspective of viewpoint 1.

3) Recognition results from viewpoint 6: The first observation is performed at viewpoint 6. The salt box and the tomato soup can are recognized very well. The recognition results of the stapled soup boxes, especially for the lower blue one, are worth looking at. Two mixture components, one for the green and one for the blue box, are assigned to the object instance of the bottom soup box. This effect results from the similarity of the objects as they are almost identical in their textures, implying they have many similar interest points. For the Amicelli box in the front - due to reflections - and for some back objects no hypotheses are found as they are too far away or beyond the image. The planning algorithm aims at differentiating between the soup boxes and sharpening the knowledge of the Ceylon tee and the jam tins. It determines moving to viewpoint 5 as best future action. The planning results are shown in Table II. Due to the heavy occlusion of the bottom soup box, many viewpoints are considered as disadvantageous.

4) Recognition results from viewpoint 5: Three new objects are recognized in the second measurement step and most goals, except for improving the knowledge of the left jam tin, are reached. Note, that the uncertainty of the jam tin even grew due to state transition effects. For the right jam tin and the Fenchel tea two mixture components depict the respective object instance distributions. As these objects have similar textures on several sides of their surface, a clear identification is not possible from this observation. While the two mixtures of the Fenchel tea are almost equally weighted, the larger mixture component of the jam tin has very little weight assigned. Due the fairly low entropy of this distribution, it is still regarded as satisfactorily recognized.

5) Recognition results from viewpoint 2: Based on the current belief distributions the planning algorithm proposes a prospective measurement from viewpoint 2 from where all uncertain objects are supposed to be clearly visible. Except for the right jam tin, all objects are observed. The ham tin is newly discovered. This results in 11 strongly peaked hypotheses distributions, which makes the algorithm terminate. The green arrows in Figure 4b show the completed action sequence.

The detection of new object instances has great effects on the planning algorithms. Thus, especially in complex scenes with a large number of objects, the detection of all objects cannot be guaranteed. The algorithm terminates when all recognized objects are of high class and pose accuracy, but does not have explorative behavior. The consideration of occlusion is essential for a reasonable observation sequence.

B. Evaluation against other strategies

In a series of experiments we compare the proposed approach to two other strategies, namely random viewpoint selection and an incremental strategy. The random strategy arbitrarily chooses the next sensing action. The incremental strategy decides on either moving clockwise or counterclockwise around the table and performs measurements step by step, always moving right to the next viewpoint.

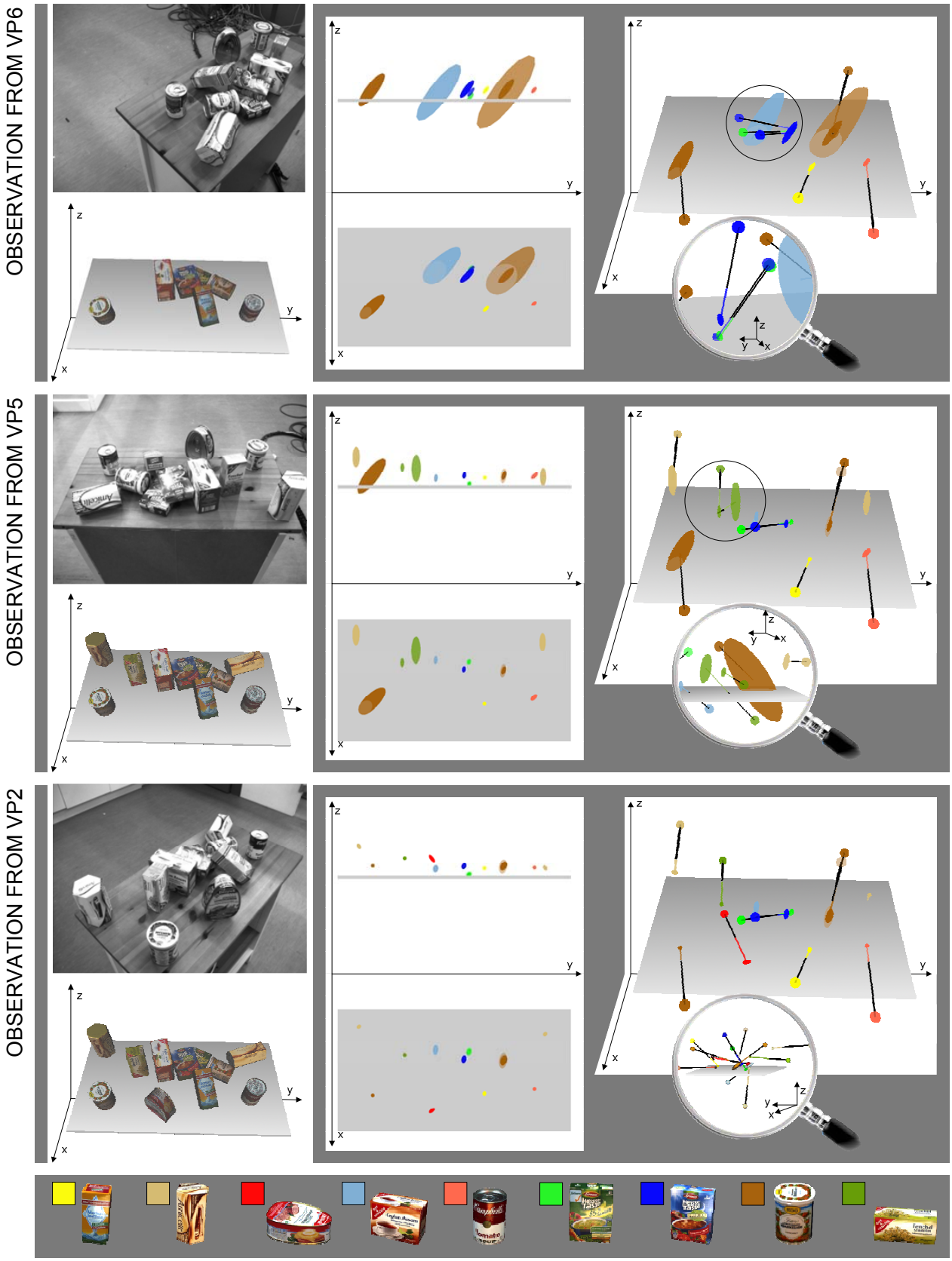
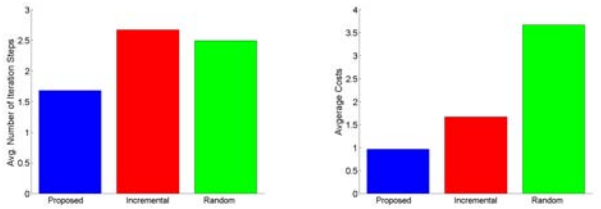
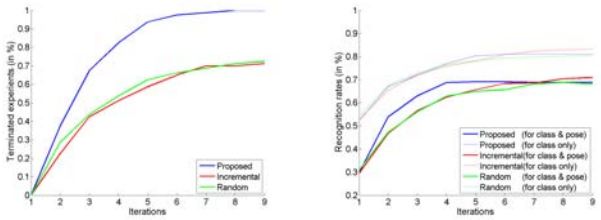


Fig. 5. Detection results for three sequential measurements. The top left plot shows one stereo camera image, the bottom left plot the current scene model. The other plots illustrate the covariance ellipsoids (97%quantile) of each mixture component of all object instance distributions. The middle column shows the horizontal and upright projection of the translational components, the right column a $3D$ view on the translational and rotational ellipsoids. On the bottom right important components are enlarged.



(a) Average number of iterations (b) Average costs

Fig. 6. Evaluation of number of iterations and costs for each strategy.



(a) Rates of successful termination (b) Recognition rates

Fig. 7. Experimental results, analyzed at each iteration step.

In the following we compare these strategies in 80 experiments with different scenes consisting of up to 10 objects. The proposed strategy outperforms the others in the number of average iteration steps with an average of 1.68 steps as shown in Figure 6a. In the movement costs a clear increase can be seen from the proposed to the incremental strategy. The random strategy is weakest as expected. Figure 6b shows the results.

In Figure 7a the rates of successful termination of the algorithms are plotted against the iterations. Significantly more experiments terminated after few iterations for the proposed strategy than for others. Also, the proposed strategy almost always terminates, while the others sometimes cannot achieve the desired termination precision. Note, that it does not necessarily mean that all objects are successfully detected, when the algorithms stop, as no explorative behavior is implemented. Thus, sometimes the precise recognition of a subset of all objects meets the specified termination criterion already.

The recognition rates with respect to the different strategies are presented in Figure 7b. The rates are again plotted over the iteration steps to show how the results change with the number of incorporated measurements. We differentiate between the detection rates for the object class, illustrated by the dashed line and the recognition rate, which additionally takes into account pose accuracy (solid line). As presumed the class detection rate is similar for all strategies due to the lack of exploration abilities. It grows with the number of iterations as object instances accumulate with the number of observations up to the total number of objects in the scene. The recognition rate including the pose accuracy is higher for the proposed strategy between the second and the forth iteration step, what proves the deliberate reduction of uncertainty in the scene model.

VII. CONCLUSIONS AND FUTURE WORKS

In this work we realized an active perception framework which probabilistically reasons over belief distributions to efficiently plan future actions. It is integrated in a real robot and operates in realistic environments. In experiments the functionality of the proposed approach is demonstrated. Its good performance is shown in comparison to two other viewpoint selection strategies.

In future works object relations could be used to improve localization errors and measurement data. Also, it could be of interest to model occluded and invisible space to give the robot more profound knowledge for action selection and enable it to explore the environment.

VIII. ACKNOWLEDGMENTS

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.

REFERENCES

[1] S. Chen, "On Perception Planning for Active Robot Vision," *IEEE SMC Society eNewsletter*, vol. 13, December 2005.

[2] S. Dutta Roy, S. Chaudhury, and S. Banerjee, "Active recognition through next view planning: a survey," *Pattern Recognition*, vol. 37, no. 3, pp. 429–446, 2004.

[3] P.-E. Forssen, D. Meger, K. Lai, S. Helmer, J. J. Little, and D. G. Lowe, "Informed visual search: Combining attention and object recognition," in *IEEE International Conference on Robotics and Automation*, 2008, pp. 935–942.

[4] O. Lanz, "Occlusion robust tracking of multiple objects," *International Conference on Computer Vision and Graphics*, 2004.

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

[6] G. Flandin and F. Chaumette, "Visual data fusion for objects localization by active vision," in *Proceedings of the 7th European Conference on Computer Vision*, 2002.

[7] 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.

[8] C. Derichs, F. Deinzer, and H. Niemann, "Integrated viewpoint fusion and viewpoint selection for optimal object recognition," in *British Machine Vision Conference*, 2006, pp. 287–296.

[9] J. Ma and J. W. Burdick, "Dynamic sensor planning with stereo for model identification on a mobile platform," in *Proceedings of IEEE International Conference Robotics, Automation and Control*, 2010.

[10] W. Feiten, P. Atwal, R. Eidenberger, and T. Grundmann, "6d pose uncertainty in robotic perception," in *Proceedings of the German Workshop on Robotics*, 2009.

[11] M. W. Spong, *Robot Dynamics and Control*. New York, NY, USA: John Wiley & Sons, Inc., 1989.

[12] 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.

[13] 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.

[14] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for gaussian mixture random vectors," in *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008.

[15] T. Grundmann, R. Eidenberger, M. Schneider, and M. Fiegert, "Robust 6d pose determination in complex environments for one hundred classes," in *Proceedings of the 7th International Conference On Informatics in Control, Automation and Robotics*, 2010.