2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics
Suntec Convention and Exhibition Center
Singapore, July 14-17, 2009

Probabilistic Occlusion Estimation in Cluttered Environments for Active Perception Planning

Robert Eidenberger, Raoul Zoellner and Josef Scharinger

***Abstract*— This paper presents a probabilistic framework for scene modeling and active perception planning in complex environments. It tackles the problems of representing detection and transition uncertainties in multi-object scenes without knowledge of the total number of objects in the scenario. The correct association of observation data with scene information is essential for reasonable incorporation of sequencing measurements into the scene model. This work also deals with the probabilistic computation of object occlusions for probabilistic action planning in order to select the most profitable prospective viewpoint. Concepts from computer graphics and statistics are combined for the efficient and precise estimation of future observations.**

In an experimental setting this active perception system, integrated into an autonomous service robot, is evaluated in a kitchen scenario.

I. INTRODUCTION

In previous works [1][2] we developed an active perception planning approach for a service robot which convinces in its applicability to high dimensional state spaces, its quick computation and its efficient planning strategy. As in most other approaches in literature we also considered single object scenes in the past. As a matter of fact in natural, realistic scenarios the constraint of considering solitary items only does not hold. Quite the contrary in household environments mostly arrangements of different or alike objects, such as food, furniture, dishes et cetera are found. Consequently we seek for a service robot which is able to detect and interpret multi object constellations both for perception and manipulation.

In this paper we extend previous concepts by applying them on more complex scenarios. 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. Thus a realistic scenario contains various object instances of alike or different object classes. In this work we aim on the probabilistic modeling of multi-object scenes and correctly associating measurements with their corresponding object instances.

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

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. From these object relations expected visible (and also invisible) features can be predicted and used for improving measurement results. The proposed concept for occlusion consideration is applied for both the real measurement and the observation prediction for perception planning.

In Section II current approaches to multi-object scene modeling and occlusion estimation are listed.

Section III outlines the probabilistic active perception framework, Section IV describes the concepts in more detail.

This paper closes with an experiment in Section V which demonstrates the application of the proposed theoretical concepts on an active vision system.

II. RELATED WORK

The research in active perception brought up very sophisticated approaches for viewpoint evaluation and next best view planning [3][4][5][6]. 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. The occurring problem of object occlusion is treated by Ercan et all [7] by considering static and dynamic occluder objects. However the object poses are not modeled probabilistically, so the problem is solved by applying common algorithms from computer graphics. Koostra et all [8] approach the cluttered scene problem by keypoint clustering of interest points from very robust feature detectors, which allows feasible detection of strongly occluded objects. Hence, they avoid to model occlusions explicitly and rely on the robustness of the detector. In [9] object recognition is combined with an attention mechanism for occlusion avoidance to efficiently acquire scene information. This approach targets on rapidly identifying objects in cluttered environments, but does not model pose uncertainties.

For multi-object tracking the probabilistic comprehension of occlusion seems more relevant than in current active systems. Lanz [10] describes a Bayesian framework for robust multi-object tracking under the presence of occlusions. He takes into account the object visibility for computing the observation likelihood by analyzing each image pixel. Tamminen [11] presents a similar approach for face recognition. The probabilistic representation by inverse-Wishart

978-1-4244-2853-3/09/$25.00 ©2009 IEEE

functions induces great computational costs and outweighs the advantages of the occlusion model. Gupta et all [12] describe a very fast and resource effective vision system for people part tracking. They use a heuristic for probabilistically determining body occlusions which is calculated from the number of visible voxels. This approach does not convince due to the discretization of space.

In this paper we present an active perception system which estimates object occlusions by considering probabilistic object poses, object geometries and feature alignments.

III. ACTIVE PERCEPTION ARCHITECTURE FOR MULTI-OBJECT SCENE RECOGNITION

In active perception we aim on choosing control actions $a \in A$ to reach a specific goal g. There is a wide range of different actions, largely depending on the application. In this paper we only consider sensor positioning at different viewpoints as sensing actions and neglect sensor parameter adjustments. The framework for selecting the best prospective action policy π is schematically illustrated in Figure 1.

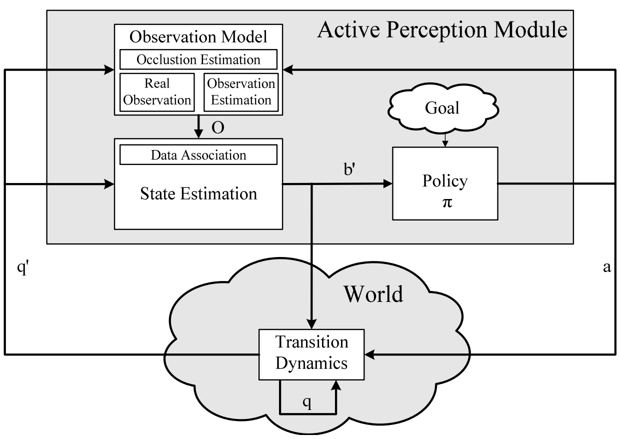


Fig. 1. Active perception framework

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 probabilistic modeling of multi-object scenes and the Bayesian statistical framework including observation association for state estimation. Also the observation model under consideration of occlusion estimation and probabilistic action planning are explained in detail.

A. Multi-object scene modeling

We consider the single object instance ι with $0 \leq \iota \leq I$, 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 probability 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 [2] and considers uncertainties in the state transition and in the measurement for state estimation. We state the probability distribution over the state

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

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

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

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.

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

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

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

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 ability 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

$$P(O_t(a_t)|q') = \sum_i P^i(O_t(a_t)|q') \tag{5}$$

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 $b^\iota_{t-1}(q)$ by applying the Mahalanobis-distance measure

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

on both distributions. The similarity is defined by a specific threshold dependent on the object class' geometry. 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 likelihoods 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

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

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 the 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^ι_s 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

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

which states an approximation for the probability that 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:

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

D. Perception planning

The probabilistic planning concept in form of a partially observable Markov decision process, as proposed in [2], 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:

$$\pi(b) = \operatorname*{argmax}_a R_a(b) \tag{10}$$

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 actions during planning, this summation is justified.

IV. REALIZATION OF THE PROBABILISTIC FRAMEWORK

This section details the basic concepts of state and occlusion estimation and describes how uncertainties are modeled and used in the robotic scenario.

A. Transition uncertainty in state estimation

The transition uncertainty is defined as the linear Gaussian

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

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

B. Uncertainties in the observation model

In this work we use a stereo camera system and the SIFT-detection algorithm [13] 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 [1][2] 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 [14]. The uncertainties result from relations between seen and expected interest points, matching errors and sensor and feature characteristics.

C. Probabilistic occlusion estimation

The proposed analysis of feature visibility follows the 5-step algorithm illustrated in Figure 2. It is demonstrated on the example of the two-object constellation pictured in Figure 3a.

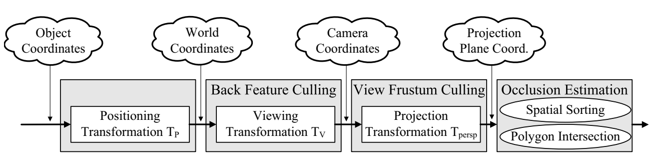
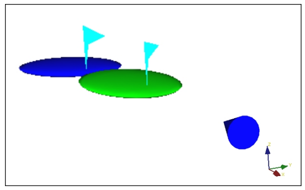


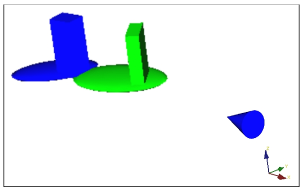
Fig. 2. Occlusion estimation algorithm

1) Initially when drawing a sample from all object instance distributions $b_{t-1}(q)$ we get a state q_s^{ι} for each object instance. From the object instance class C^{ι} the object geometry, such as the bounding box, and the set of surface features are acquired from the corresponding data base object. All geometric data is given with respect to the object instance's origin.
2) In order to represent all instances within one coordinate reference frame the position transform T_P is applied to transfer all items into world coordinates. The pose ϕ^{ι} specifies the sample location. Figure 3b shows the drawn sample from the probability distributions, Figure 3c pictures the bounding boxes of the two-object scene.
3) The viewing transformation T_V transforms features and bounding boxes into camera coordinates. In the first stage of visibility analysis back feature culling is accomplished by determining the feature visibility due to the object instance's pose. It aims on identifying all invisible features by executing following methods:
   - Feature at the back side of the object, which point away from the sensor, are identified by checking the direction vector of the feature. These are considered to be undetectable.
   - Features with a z-location of their center behind the camera are eliminated due to their invisibility.
   - To find out if a feature, although pointing towards the sensor, is outside its visibility volume, the conic view frustums of the sensor and the feature are checked. Both have to lie within the others visibility cone.
4) In the sequencing projection transformation step the object bounding boxes are transformed into a 2-dimensional projection plane, which is clipped according to the sensor properties. This process is called view frustum culling. The transformation matrix of a perspective projection $T_{persp} = (n^T V)^T - (nV)I_4$ can be easily derived from the normal of the projection plane n and the current homogeneous viewpoint V with I_4 as the 4×4 identity matrix [15]. In [16] this matrix is expanded for clipping the scene by a pyramidal view volume in order to consider only bounding box vertices which are within the sensor range. The convex hull of the projected bounding box's vertices specifies the resulting, visible $2D$ bounding polygon. Figure 3d shows the perspective projection of the sample object instances, Figure 3e illustrates the bounding polygons in the projection plane.
5) In previous steps the feature visibility due to its object's geometry was depicted. Now we deal with object occlusion in multi-object scenes which are caused by other object instances. First the object instances are spatially sorted with reference to the distance of their origin from the sensor following methods such as depth-sort-algorithm and binary space partitioning. From this sorted object list all possible occluder objects are selected for each object instance. Each object's bounding polygon is compared with each occluder's polygon for an intersection by using boolean set operations. An intersection means that the objects occlude each other. In this case the features of the hindmost

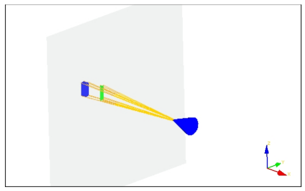
(a) Sample constellation containing the soup box (green) and the salt box (blue).



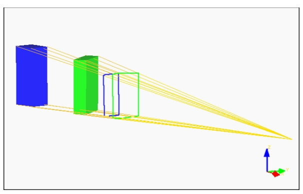
(b) Each flag represents an object instance pose q_s^t. The covariance ellipsoids depicts the uncertainty in the belief distribution.



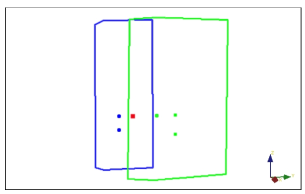
(c) Bounding boxes of the sample constellation in world coordinates.



(d) Perspective projection of the bounding boxes into the projection plane.



(e) $2D$ bounding polygons in projection plane for occlusion check.



(f) Occlusion event with two intersecting bounding polygons. One feature (red) is occluded.

Fig. 3. Occlusion estimation algorithm for a sample object constellation

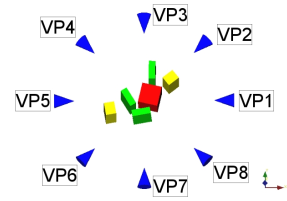
object are projected onto the projection plane. If a feature lies within the polygon intersection area it is considered to be occluded. The probability of the feature visibility is acquired from the percentage of the occluded feature area. Performing all occlusion checks leads to the probability distribution $p(f)$ which describes the estimated likelihood of seeing a feature. Figure 3e pictures an occlusion event.

After performing the matrix multiplication chain $T = T_{persp} T_V T_P$ occlusion estimation considers relations between objects in order to probabilistically determine the feature visibility. The number of samples required for determining occlusions is closely related to the shape of the respective probability distribution, the sample is drawn from.

V. EXPERIMENTAL RESULTS

In this experiment the proposed approach is demonstrated with the environmental setup shown in Figure 4a. The scenario consists of three different types of objects, a salt box, a soup box and a sauerkraut tin. The object database contains the object geometries and the SIFT interest points per view. We assume the measurement model to be identical, both for the state update and the state prediction. The calculation of the measurement probability bases on simulated data in order to run through a great variety of observation sequences with varying outcomes for one scenario. The characteristics of the measurements are derived from real observation data.

(a) Multi-object scenario



(b) Viewpoint arrangement

Fig. 4. Experimental setup

In real measurements in the illustrated scenario we acquired the object poses and their average spreading. Additionally we formulated 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. This descriptions and the results of the occlusion estimation process are used for determining the object class probability and pose covariance in the simulated measurement model.

A sensing action is defined as a change in the robot's viewpoint. Figure 4b illustrates 8 different, circularly aligned control actions, which are evaluated in this experiment. We derive the costs of the robot's movement to another viewpoint from the movement angle. The costs for accomplishing a new observation without moving are set to 0.5 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. The problem is assumed to be planar, so we only consider locations in the xy-plane and rotations around the z-axis for the sake of clarity.

In the following an action sequence resulting from the proposed planning approach is presented. Figure 5 shows the detection result in each step, Table I lists costs, values and rewards in each detection step of all actions.

As initially no scene knowledge is available, each action promises identical benefits. We acquire the first measurement from the current robot position, namely viewpoint 6. The measurement results with varying characteristics are associated and merged to three newly detected object instances. In Figure 5 the translational components of the uncertainty ellipsoids covering 97 percent of the probability mass are plotted. Based on the belief distributions viewpoint 5 is suggested for the prospective measurement. Some single measurements of this sensing actions are associated with the already detected salt and soup box, leading to a significantly more certain pose. The sauerkraut can is not visible as being occluded by another, first-time found soup box. Hence the sauerkraut's uncertainty even grows due to the state transition inaccuracies. Two further new object instances are found in the background. The next planning step aims on optimizing all 6 belief distributions and makes the robot move to viewpoint 3. While most object instance poses are determined quite accurate, the uncertainty of the closest object instance's pose is hardly reduced. This is because only

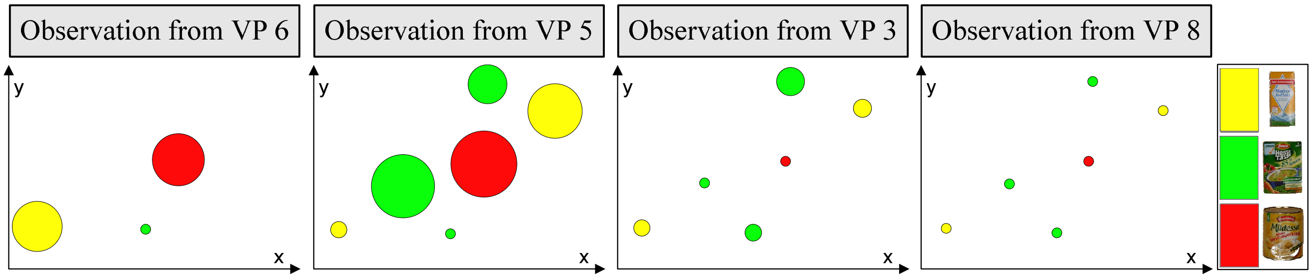


Fig. 5. Covariance distributions of xy components for the selected sequence of control actions

TABLE I

CORRELATING COSTS AND VALUE FOR CALCULATING THE REWARD R_a FROM EQUATION (11) FOR SELECTED VIEWPOINTS GIVEN $\alpha = 0.75$

View-points	VP6 R_a	costs	VP6 value	R_a	costs	VP5 value	R_a	costs	VP3 value	R_a	costs	VP8 value	R_a
VP1	-0.92	-0.92	-0.82	-1.51	-1.00	-0.26	-1.01	-0.70	-0.64	- 1.17	-0.38	-0.43	-0.71
VP2	-1.00	-1.00	-0.49	-1.23	-0.92	-0.40	-1.10	-0.38	-0.64	-0.93	-0.70	-1.00	-1.53
VP3	-0.92	-0.92	-0.00	-0.69	-0.70	-0.00	**-0.53**	-0.50	-0.61	-0.98	-0.92	-0.04	-0.74
VP4	-0.70	-0.70	-0.68	-1.21	-0.38	-0.27	-0.56	-0.38	-0.61	-0.90	-1.00	-0.43	-1.18
VP5	-0.38	-0.38	-0.22	**-0.50**	-0.50	-0.33	-0.70	-0.70	-0.38	-0.91	-0.92	-0.44	-1.13
VP6	**-0.00**	-0.50	-0.61	-0.99	-0.38	-1.00	-1.28	-0.92	-1.00	-1.69	-0.70	-0.19	-0.72
VP7	-0.38	-0.38	-1.00	-1.28	-0.70	-0.09	-0.62	-1.00	-0.86	-1.61	-0.38	-0.00	**-0.29**
VP8	-0.70	-0.70	-0.37	-0.90	-0.92	-0.11	-0.80	-0.92	-0.00	**-0.69**	-0.50	-0.38	-0.76

the small side of soup box is visible, which contains few interest points Viewpoints close to the current do have similar expected rewards, viewpoint 8 outperforms these due its great value despite the costly movement. The resulting belief contains 6 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 modeled noise, 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.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents an active perception module for planning next best viewpoints in multi-object environments based on contrasting perception benefits and action costs. This approach is embedded in a fully probabilistic framework and works in continuous high-dimensional domains. It considers object relations by a probabilistic representation of object occlusions for efficient decision making. Currently the reasoning process bases on present scene knowledge. In future works it could be of interest to also model occluded and invisible space to give the robot more profound knowledge for action selection and enable it to explore the environment.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

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

[6] 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.

[7] A. O. Ercan, A. El Gamal, and L. J. Guibas, "Camera network node selection for target localization in the presence of occlusions," in *SENSYS Workshop on Distributed Smart Cameras*, October 2006.

[8] *Active exploration and keypoint clustering for object recognition*. IEEE, 2008.

[9] P.-E. Forssen, D. Meger, K. Lai, S. Helmer, J. J. Little, and D. G. Lowe, "Informed visual search: Combining attention and object recognition," 2008, pp. 935–942.

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

[11] T. Tamminen and J. Lampinen, "A bayesian occlusion model for sequential object matching," in *Proceedings of the British Machine Vision Conference*, 2004, pp. 547–556.

[12] A. Gupta, A. Mittal, and L. S. Davis, "Cost: An approach for camera selection and multi-object inference ordering in dynamic scenes," in *Proceedings of the Internation Conference On Computer Vision*, 2007, pp. 1–8.

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

[14] 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.

[15] D. Marsh, *Applied Geometry for Computer Graphics and CAD*, 2nd ed. Springer, 2005.

[16] J. D. Foley, A. V. Dam, S. K. Feiner, and J. F. Hughes, *Computer graphics: principles and practice*, 2nd ed. Addison-Wesley, 1996.