Fast Parametric Viewpoint Estimation for Active Object Detection

Robert Eidenberger
Department of Computational Perception
Johannes Kepler University Linz
4040, Linz, Austria

Thilo Grundmann, Wendelin Feiten, Raoul Zoellner
Siemens Corporate Technology
Information and Communications
81739, Munich, Germany

{robert.eidenberger.ext, thilo.grundmann.ext, wendelin.feiten, raoul.zoellner}@siemens.com

Abstract— **Most current solutions to active perception planning struggle with complex state representations or fast and efficient sensor parameter selection strategies. The goal is to find new viewpoints or optimize sensor parameters for further measurements in order to classify an object and precisely locate its position.**

This paper presents an exclusively parametric approach for the state estimation and decision making process to achieve very low computational complexity and short calculation times. The proposed approach assumes a realistic, high dimensional and continuous state space for the representation of objects expressing their rotation, translation and class. Its probability distribution is described by multivariate mixtures of Gaussians which allow the representation of arbitrary object hypotheses.

In a statistical framework Bayesian state estimation updates the current state probability distribution based on a scene observation which depends on the sensor parameters. These are selected in a decision process which aims on reducing the uncertainty in the state distribution. Approximations of information theoretic measurements are used as evaluation criteria.

I. INTRODUCTION

Many computer vision systems which interpret single observations face the problem of having to suggest the object class and object state from insufficient measurement data. These uncertainties might result from either inaccurate sensing devices, weak classifiers, occlusions, poor lighting or ambiguity of object models. Active perception approaches aim on incorporating further measurements under different sensing settings to gain more precise information on the scene. Their direct application to real world problems such as object recognition, visual surveillance, inspection, target tracking, visual search, monitoring systems, 3D object reconstruction or scene exploration and interpretation underlines their importance.

The developed work is part of a service robot which autonomously performs manipulation tasks such as grasping objects. The demand of suitability for everyday environments requires the localization of objects under challenging environmental conditions such as bad lighting, reflections or occlusions. As a matter of fact the unique classification and localization of an object is often not possible given a single sensor measurement. Thus fusing several observations would lead to better detection results at the costs of sensor movements and additional computational power. In order to minimize these expenses a minimal sequence of views would be desired which provides excellent classification and localization results.

Section II outlines current state of the art approaches and deals with the problems of viewpoint selection and state estimation.

In section III an overview of the perception architecture is given. The more detailed concept regarding state estimation for parametric state distributions is explained in section IV.

Section V explicates a fast sensor parameter selection strategy, which chooses the next best viewpoint from estimated posterior probability distributions.

This paper closes with the evaluation Section VI where experiments on proposed theories are described.

II. RELATED WORK

Lots of research work on active vision brought up very sophisticated approaches for dealing with this problem. Since in this paper we decided to use a basic widely-used statistical framework [1], we only highlight relevant theories including Bayesian state estimation and online strategies for sensor action evaluation. Several state of the art approaches are presented which examine active perception problems. Their architectural properties and their capability for continuous high dimensional state representations are descussed.

In [2] an appearance-based approach to active object recognition is proposed. The method combinedly handles the effects of shape, pose, reflections and illumination by projecting an image into a parametric eigenspace. In an automatic learning procedure for a large amount of images the eigenspace is constructed. In order to find a new viewpoint each possible movement from the current eigenspace entry which is closest to the one from the input image is evaluated. Therefore the entropy loss is used as a criterion. Despite the applicability of this approach to high dimensions, the fact that only single object scenes can be considered and the difficulties of continuous pose representation constitute drawbacks.

Starting from Bayesian state estimation [3] uses entropy maps as selection criterion for next best view planning, which have to be build offline at high costs. The framework is extended in [4] by using a measure which builds up on Fisher's linear discriminant analysis for the evaluation of a viewpoint. The advantage of computational efficiency and the fast algorithm unfortunately only have been proven for a one-dimensional discretized state space for the object pose.

A probabilistic Bayesian object recognition technique is presented in [5] and [6] which bases on the statistical representation of 3D objects by several multi-dimensional receptive field histograms. Information theoretic evaluation approaches such as mutual information or Shannon's entropy allow both, the prediction of the performance of the object recognition process and a measure for the discrimination of a viewpoint. As well the adaptability of all equations to a continuous state space is claimed, still experiments mainly show their purpose for sampled distributions.

Another approach [7], quite similar to the previous one, systematically reduces the uncertainty in the state estimation process. Mutual information is used for gaze control and the selection of an optimal sequence of sensing actions. Experiments for discrete probabilities endorse the quality of the Bayesian statistical framework and the information theoretic evaluation. The authors also consider continuous probability density representations. Therefore the classifier is swapped with the more sophisticated eigenspace classifier [8] [2], and the sensing parameters are chosen after Monte Carlo evaluation of the differential mutual information. Experiments show the effectiveness in matters of computation time, detection rate and minimal number of views. Nevertheless the problems of eigenspace classifiers still exist and the sampling of high dimensional continuous distributions increases the runtime.

In [9] the Condensation algorithm is applied for creating sample sets from the continuous probability distributions. After updating the particle distribution its density is evaluated by a kernel density estimator such as a Parzen estimation. For solving the problem of choosing optimal views the authors use reinforcement learning for viewpoint training, which allows selecting a sensing action from a continuous viewpoint space based on the learned action-value function. Advantages of this approach definitely lie in the independence of the classifier for the fusion process, its continuous viewpoint space and the automated training procedure. Still the offline training process takes already very long for a single degree of freedom camera. Also the alternation of sampling and fitting again parametric probability distributions is laborious.

III. Perception Architecture

This work focuses on model-based applications which means that object properties and appearances are initially well-known and unknown objects are not taken into account. It addresses the problems of object classification and pose determination for manipulation tasks. Figure 1 illustrates the scheme of the statistical framework which is applied to this active perception problem. The active system follows a sequential, iterative two-step process, including next best view planning and state estimation.

1) Optimal sensor parameter selection: During the first step the current state q is updated with an estimated observation $\hat{O}(a)$. The parameter a denotes a sensing action which can either be a new viewpoint or a change in the sensor settings such as zoom. The evaluation process targets on finding the sensor action which minimizes the uncertainty in the state estimate at most. The actuator accomplished the requested action.

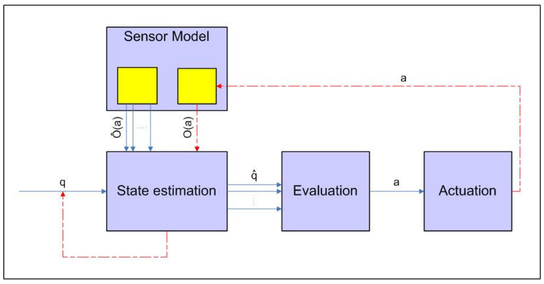


Fig. 1. Active object detection architecture: the bricks connected by the solid line depict the sensor parameter selection sequence, the dashed line concatenates the perception action stages and closes the perception cycle

2) Perception action: Step two executes the sensing task by performing the real observation $O(a)$ given the previously determined sensor settings a. This leads to a state update which will be used as input for the next parameter estimation step.

IV. Parametric Probabilistic Framework

This section explicates the state estimation process and constitutes the concepts of the state space and its probabilistic representation.

A. State space representation

A set of object models $(\Omega_1, ..., \Omega_k)$ establishes an object database Ω, containing all k different object classes, we finally want to detect. Besides the classification of an object type, the object's location and the object's pose are of great importance. The object position is expressed in the $m + n$-dimensional, continuous state space $\rho = \mathbb{R}^m \times \mathbb{S}^n$, where n translational components span the real infinite state space $\mathbb{R}$. As for periodic variables this infinte state space is inappropriate the periodic m-dimensional state space $\mathbb{S}$ is introduced which ranges from 0 to 2π and is very suitable for angular variables. Consequently the world state $q \in Q$ with $Q = \Omega \times \rho$ can be described as a set of tuples $q = (\Omega_i, \phi^T)^T$ with $\phi = (\phi_1, ..., \phi_m, \phi_{m+1}, ..., \phi_{m+n})^T$ where $(\phi_1, ..., \phi_m) \in \mathbb{R}^m$ and $(\phi_{m+1}, ..., \phi_{m+n}) \in \mathbb{S}^n$. In other words the entirety of all object-tuples builds up the world state. Since Q represents both discrete and continuous dimensions it will be further considered as a mixed state.

B. State estimation

Most approaches listed in Section II base on the widely-used Bayesian statistical framwork for state estimation. As well this approach updates the probabilistic posterior state distribution by taking into account prior knowledge and the observation likelihood. The goal is to conclude a state probability estimation for incorporating a new observation $O_t(a_t)$ at time t. Defining $p(q_t|O_{t-1}(a_{t-1}), ..., O_0(a_0))$ as the a priori probability distribution for previous sensor measurements $O_{t-1}(a_{t-1}), ..., O_0(a_t)$, the updated state distribution $p(q_{t+1})$ can be conducted using Bayes' rule:

$$\begin{aligned} p(q_{t+1}) : &= p(q|O_t(a_t), ..., O_0(a_0)) \\ &= \frac{P(O_t(a_t)|q)p(q|O_{t-1}(a_{t-1}), ..., O_0(a_0))}{P(O_t(a_t), ..., O_0(a_0))} \end{aligned} \tag{1}$$

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

$$\begin{aligned} &P(O_t(a_t), ..., O_0(a_0)) = \\ &\int_q P(O_t(a_t)|q)p(q|O_{t-1}(a_{t-1}), ..., O_0(a_0))dq. \end{aligned} \tag{2}$$

As the current observation $O_t(a_t)$ can be considered as the detection of a set of N features $F = \{f_1(a_t), ..., f_N\}$ it can be rewritten as

$$O_t(a_t) = \{f_1(a_t), ..., f_N(a_t)\}. \tag{3}$$

Substituting $O_t(a_t)$ in the likelihood distribution $P(O_t(a_t)|q)$ with Equation (3) leads to a complex expression whose calculation becomes computationally very expensive for a large number of feature. Instead from assuming all features to be conditionally independent which states a valid approximation the likelihood distribution can be computed by applying the naive Bayes assumption:

$$P(O_t(a_t)|q) = \prod_i^N P(f_i|q) = \prod_f P(O_t^f(a_t)|q). \tag{4}$$

While Equation (1) is used for a state update given the real observation from Equation (3), the estimated posterior distribution $p(\hat{q}_{t+1})$ for a given estimated observation $\hat{O}(a_t)$ is calculated analogously finally resulting in the posterior estimate

$$p(\hat{q}_{t+1}) = \frac{\left[\prod_f P(\hat{O}_t^f(a_t)|q)\right]p(q_t)}{\int_q \left[\prod_f P(\hat{O}_t^f(a_t)|q)\right]p(q_t)dq}, \tag{5}$$

where $p(q_t)$ denotes the prior distribution.

C. Probabilistic representation

For determining an appropriate probability density distribution for the representation of the world state the scene properties have to be examinated. Because of the purpose to model various object hypotheses or even multiple objects within the same probability distribution, all single peaked densities are inappropriate. Thus, alternatives like particle distributions and mixtures of Gaussians are the only feasible and due to the high dimensional state space we decided on using the multivariate Gaussian mixture distribution. Equation (6) depicts it as the weighted sum over normal distributions, given by their mean vectors μ_k and covariance matrices Σ_k. π_k denotes the weight of the kth component. [10]

$$p(q) = \sum_{k=1}^K \pi_k \mathcal{N}(q|\mu_k, \Sigma_k) \tag{6}$$

D. Bayesian framework with Gaussians mixtures

When treating the posterior probability in the Bayes theorem as the a priori probability of the sequencing process of sequential decision making, both distributions must be of the same type. Referring to conjugate priors [11] for the parametric probability density multiplication of normal distributions we show that the multiplication of likelihood and prior, both of the type of mixtures of Gaussians, can be done simply by a componentwise parameter update to receive the posterior Gaussian mixture distribution:

$$\sum_{k=1}^{K_1 K_2} \pi_k \mathcal{N}(q|\mu_k, \Sigma_k) = \sum_{i=1}^{K_1} \pi_i \mathcal{N}(q|\mu_i, \Sigma_i) \sum_{j=1}^{K_2} \pi_j \mathcal{N}(q|\mu_j, \Sigma_j) \tag{7}$$

The resulting hyperparameters μ_k, Σ_k, π_k of each component are computed from the modes of the multiplicands i and j.

$$\begin{aligned} \Sigma_k &= (\Sigma_i^{-1} + \Sigma_j^{-1})^{-1} \\ \mu_k &= \Sigma_k(\Sigma_i^{-1}\mu_i + \Sigma_j^{-1}\mu_j) \\ \pi_k &= \frac{\pi_i \pi_j}{(2\pi)^{D/2}} \sqrt{\frac{|\Sigma_k|}{|\Sigma_i||\Sigma_j|}} \\ &\quad \exp(1/2(\mu_k^T \Sigma_k^{-1}\mu_k - \mu_i^T \Sigma_i^{-1}\mu_i - \mu_j^T \Sigma_j^{-1}\mu_j)) \end{aligned} \tag{8}$$

$||$ indicates the determinant of a matrix, D denotes the dimension of the continuous state space. The evidence term which can be easily computed by calculating the sum of the weights, as the integral over a Gaussian distribution equals one, normalizes the distribution.

The number of components increases for each multiplication as it equals the product of the multiplicand modes. As there is no need to take along unimportant mixtures and to keep the number of components small two mixture reduction strategies are applied:

1) Merging mixtures: Components with similar mean-vectors can be unified according to following equations [12]:

$$\begin{aligned} \mu_c &= \sum_{k=1}^K \pi_k \mu_k \\ \Sigma_c &= \sum_{k=1}^K \pi_k(\Sigma_k + (\mu_k - \mu_c)(\mu_k - \mu_c)^T) \\ \pi_c &= \sum_{k=1}^K \pi_k \end{aligned} \tag{9}$$

2) Dropping mixtures: Mixtures with insignificantly low weights can be dropped as they do not distinctly influence the probability distribution. The weights of the remaining components have to be readjusted.

V. Sensor parameter selection

For selecting the best sensing action the estimated posterior probability distributions have to be evaluated regarding its uncertainty. The more a distribution peaks the more precise it determines the object's class and pose. The proposed sensor parameter selection process bases on information theoretic rating approaches for probability distributions:

A. Information theoretic posterior evaluation

The information theoretic concept for measuring the uncertainty for continuous probability distribution is the differential entropy

$$h(q) = -\int_q p(q) \log p(q) dq. \tag{10}$$

As some properties of Shannon's entropy do not apply for the differential entropy in the following its suitability is discussed. From the parametric entropy equation for a multivariate normal distribution

$$h(q) = \frac{1}{2} \log{(2\pi exp)^D} |\Sigma| dq. \tag{11}$$

it can be easily seen that for the covariance matrix determinant $|\Sigma| < 1/(2\pi exp)^D$ the entropy can become negative. This bases on the reason that the probability density is a probability per unit length and consequently depends on the measure length. Thus the differential entropy is not invariant to coordinate transforms such as scaling. For translation and rotation invariance applies. As not the absolute entropy but the scale invariant entropy relation among the viewpoint dependent posterior distribution estimations is significant for selecting the best viewpoint, the differential entropy qualifies as a satisfying measure.

The computation of the entropy both numerically or by sampling from the probability density distribution is costly in matters of processing time for multivariate Gaussian mixtures. Therefore the upper and lower bounds of the differential entropy are determined according to [12]. The relevance of the upperbound approximation

$$h(\hat{q}_{t+1}) = \sum_c \frac{1}{2} log[(2\pi \exp)^D |\Sigma_c|] \tag{12}$$

for sensor parameter selection, where Σ_c signifies the merged covariances of all components for the object class c (Equation (9)), will be explained in the following section.

B. Sequential decision making

The prospective sensor action a_{t+1} is chosen from the lowest upper bound entropy estimate

$$a_{t+1} = \operatorname*{argmin}_a h(\hat{q}_{t+1}(a)). \tag{13}$$

Due to the minima function the state estimation will be improvement in every step, but the optimality of the found sequence of actions cannot be guaranteed.

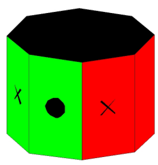
As already mentioned in Section I, the planning and sensing action steps are iterated until a stop criterion is reached. Ideally the detection precision fulfills the recognition requirements, when the sum of the variances over all object classes falls below a threshold. Setting this threshold greater than $1/(2\pi exp)^D$ guarantees positive entropies as the algorithm will terminate before. Due to the effect that equal measurements sharpen the posterior distribution, previously selected viewpoint are being penalized to avoid several sequencing observation with the same parameter set.

VI. Experimental Results

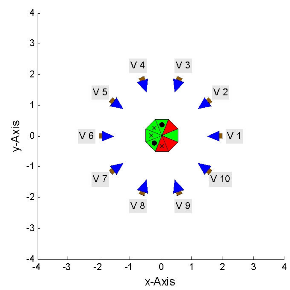
In two experiments the state estimation and sensor action selection process will be explained in detail. While the first one is based on a synthetic object in order to clearly show the functionality of the approach the second experiment was performed with real sensors and objects.

A. Experiment with generated sensor model

The first experimental setting aims on detecting the artificial octagon shown in Figure 2a). The object is positioned in the three-dimensional state space $\mathbb{R}^3$, where two dimensions describe the location in the xy-plane, the third provides information about the object's rotation. The periodicity of the angular component is taken care of by choosing an appropriate working point for correct processing. In addition to the state q in the world coordinate frame the types of sensing actions have to be defined. In this experiment a stands for a change of the sensor's viewpoint, where the state space of the sensor action equals $\mathbb{R}^3$. Figure 2b) illustrates the object's real world pose at $\phi = (0, 0, 180°)$ and an arrangement of 10 viewpoints which can be chosen from to quickly and accurately determine the correct object pose.



(a) Octagon object



(b) Viewpoints and object arrangement; the lateral features are drawn on the top of the object for better comprehension

Fig. 2. Experimental setting

The generated database and sensor model allow very precise control, monitoring and verification of the active recognition process. We introduce two different types of sensors. A color detector distinguishes between red (dark) and green (light) and a shape detector which is able to recognize point and cross features. Both sensors work as binary feature detectors meaning they can either see the feature or not. Seeing a feature signifies that the feature occurs within the sensor's vision range which is modeled by a geometric cone with an aperture angle of 90 degrees. The actual detection also depends on the feature properties, resulting in a likelihood of detecting the feature. Each feature is attached to the object's geometric mantle implying an areal pose. Its area of visibility is described in ways of the viewing perspective and distance. Additionally to this physical properties we assume noise effects on the sensor activity inferring that the feature might not be seen despite its presence. Based on these assumptions for each sensing viewpoint the likelihood $P(f(a)|q)$ of a positive detection of an object at a certain pose has to be determined from true and false positive feature incidences. At this point we assume the likelihood distribution over all states q to be equivalent to the distribution of the field of view of the feature. Latter can be computed by sampling the state space, quantifying all object poses from where the feature is visible and fitting a multivariate mixture of Gaussians into this uniform space. For this purpose expectation-maximization approaches [13] [14] provide feasible results. In order to find the feature likelihood this normalized mixture distribution has to be raised accordingly by adjusting the component's weights. As the database and all features are initially known all feature distributions can be assigned in an offline training process. Figure 3 displays the covariance ellipsoid of the likelihood distribution of a point feature, modeled by a single Gaussian mixture, which covers 97 per cent of the probability mass.

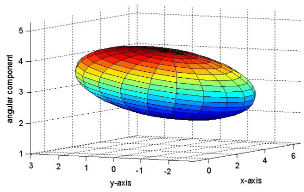
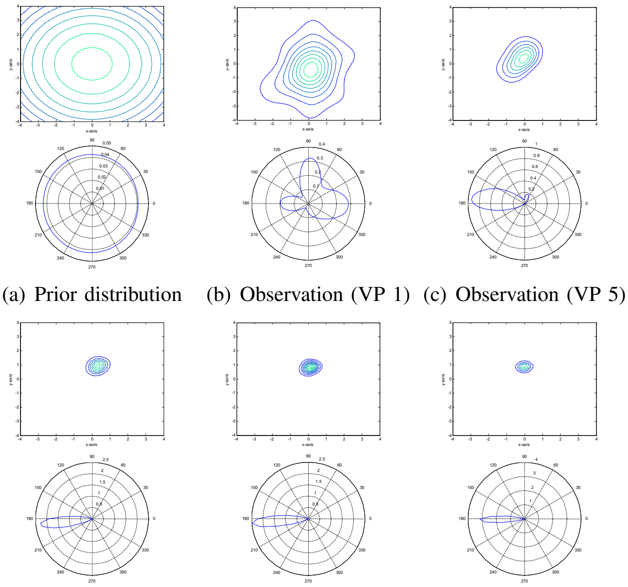


Fig. 3. Covariance ellipsoid of the normalized probability density distribution of $P(f(a)|q)$

To compute the posterior probability as described in Section IV the set of expected features for an observation from a certain viewpoint is estimated from the prior probability distribution. Initially the prior is assumed to be equally distributed and is modeled as a single Gaussian mixture component with large variances. The top plot of Figure 4 a) shows the contours of the probability distribution of the xy-dimensions integrated over the angular dimension. The bottom polar plot represents the angular probability distribution integrated over the other dimensions. Starting from this initial prior a sequence of observations is accomplished, each from a viewpoint which was selected in order to reduce the uncertainty in the state distribution at most. Figure 5 illustrates the consecutive sensing actions, Figures 4 b) to f) represent the related posterior distributions. From seeing the green (light) feature from viewpoint 1 not that much information can be gathered as this feature emerges quite often on the object's surface. When incorporating a measurement from viewpoint 5 seeing the cross and the green (light) feature reduces the uncertainty especially for the angular prediction a lot. Further observations from the viewpoints 8, 3 and 4 sharpen the object's location and also contribute a little to determine the orientation more precisely. The algorithm terminates with the peak probability distribution of the object pose at $\phi = (0.21, -0.01, 181.15°)$.

Relative to the coarse likelihood functions for each feature, this result is very satisfying. To get meaningful information



(a) Prior distribution (b) Observation (VP 1) (c) Observation (VP 5)

(d) Observation (VP 8) (e) Observation (VP 3) (f) Observation (VP 4)

Fig. 4. Sequence of posterior distributions based on the chosen sensing actions; the top plot shows the contours of the probability distribution in xy-dimensions, the polar plot highlights the angular distribution.

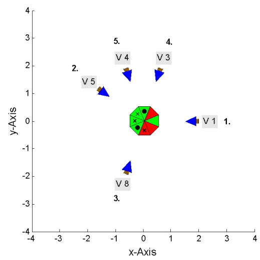


Fig. 5. Sample viewpoint sequence

of the performance of the algorithm several scenarios using different initial object poses (within an area of $1m^2$ and randomly oriented) and starting viewpoints have been tested repeatedly. Table I compares the results of this approach with a random viewpoint selection strategy. As termination criterion angular variances of the merged mixtures (Equation (9)) of less than $17.5°$ and $27.5°$ was chosen. While the average errors for the pose which was calculated from the difference of the maximum of the posterior distribution and the real pose are similar, the proposed approach requires less numbers of steps to reach the target accuracy than the random strategy. The estimation accuracies of on average $0.45m$ and $5.81°$ are very satisfying regarding the average likelihood variances of around $4.5m$ and $125°$ for one feature. It can be clearly seen, that the later perception actions only reduce the uncertainty in the state estimation. The processing time on a $1.6GHz$ AMD Dual-Core processor is less than $20ms$.

B. Experiment with real data

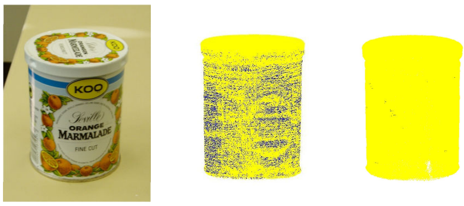
The goal of this experiment is to locate the jam tin illustrated in figure 6a) in a setting as before, using a three-dimensional state space for the object pose and 10 circularly

TABLE I

COMPARISON OF THE PROPOSED APPROACH WITH A RANDOM VIEWPOINT SELECTION STRATEGY FOR 1000 TRIALS AND TERMINATION CRITERIA FOR THE ANGULAR VARIANCES OF $\sigma_{3,3} = 17.5°$ AND $\sigma_{3,3} = 27.5°$

strategy	mean no. of views	average translational error (in m)	average angular error (in °)
random	8.03	0.44	5.48
planned	6.92	0.47	5.93
random	6.08	0.45	5.96
planned	4.88	0.44	6.01

aligned viewpoints. The artificial sensor model is replaced



(a) (b) (c)

Fig. 6. (a) shows the jam tin object, (b) the feature locations on the object geometry of the object model and (c) illustrates observed features and similar ones.

by real assembly. A RGB camera with a resolution of 1388×1038 pixels is mounted on a robot.

The detection is based on SIFT features [15]. From 36 images which are taken in 10 degree longitudinal steps from all around the jam tin 31500 features are acquired. A comparison of the feature descriptor vectors by taking the Euclidean distance shows that several features are similar in their characteristics. The histogram in Figure 7 shows the correlation of these feature types with the numbers of features belonging to them.

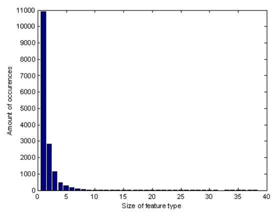


Fig. 7. Histogram showing how many feature types of a certain size are stored in the database

To examine the strength of our approach using SIFT features we assume for this experiment that the object is occluded by other objects and the viewing conditions are bad. To simulate this effect the least discriminant features, which are seen in a single view only, are suppressed in our database. For the remaining features the likelihood distributions are calculated from the views they are seen in and the scale descriptors.

In the running experiment real measurements, containing very few features due to bad viewing conditions, are performed. Figure 6c) displays the observed features and all database features similar to those. From the corresponding feature likelihoods the posterior distribution is updated and the prospected sensing action is estimated.

The encountered result in terms of numbers of views supports the result of the synthetic experiment. So does the computational time, which mainly depends on the amount of viewpoints, the number of features and the number of mixtures.

VII. CONCLUSION

This paper describes a fast method for state estimation and viewpoint selection, which is suited for high dimensional state spaces and entirely computes the probability distributions and uncertainty measures via hyperparameters. The method uses Gaussian mixtures as parametric representations of the density function. As shown, the Bayesian update can be done parametric and the entropy based decision is made by applying a upper bound Gaussian mixture approximation. Since, the computational time is very short, the algorithm can be applied to real, everyday environments.

ACKNOWLEDGMENT

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. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[2] L. Paletta and A. Pinz, "Active object recognition by view integration and reinforcement learning," vol. 31, pp. 71–86, 2000.

[3] T. Arbel and F. P. Ferrie, "Viewpoint selection by navigation through entropy maps," in *Proceedings of the 7th International Conference on Computer Vision*, vol. 1, 1999, pp. 248–254.

[4] C. Laporte, R. Brooks, and T. Arbel, "A fast discriminant approach to active object recognition and pose estimation," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, 2004, pp. 91–94.

[5] B. Schiele and J. L. Crowley, "Where to look next and what to look," in *In Proceedings of the Conference on Intelligent Robots and Systems*, 1996, pp. 1249–1255.

[6] ——, "Transinformation for active object recognition," in *International Conference on Computer Vision*, 1998, pp. 249–254.

[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] H. Murase and S. Nayar, "Visual learning and recognition of 3-d objects from appearance," *Int'l J. Computer Vision*, vol. 14, pp. 5–24, 1995.

[9] 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.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[11] D. Fink, "A compendium of conjugate priors," Tech. Rep., 1995.

[12] M. A. Carreira-Perpinan, "Mode-finding for mixtures of gaussian distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1318–1323, 2000.

[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistics Society*, vol. 39, pp. 1–38, 1977.

[14] S. Roberts, D. Husmeier, I. Rezek, and W. Penny, "Bayesian approaches to gaussian mixture modelling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1133–1142, 1998.

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