The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

# A Probabilistic Measurement Model for Local Interest Point Based 6 DOF Pose Estimation

Thilo Grundmann Robert Eidenberger Georg v. Wichert

***Abstract*— The ability to recognize objects and to localize them precisely is essential in all service robotic applications. One of the main challenges for service robots during operation lies in the handling of unavoidable uncertainties which originate from model and sensor inaccuracies and are characteristic for realistic application scenarios. Robustness under real world conditions can only be achieved when the dominant uncertainties are explicitly represented and purposefully managed by the robot's control system. We therefore adopt a probabilistic approach in which environment perception over time is regarded as a sequential estimation process and follow a Bayesian filtering methodology. Under these assumptions probabilistic models of the robot's perception systems play a decisive role.**

**In this paper we describe our object localization system which is based on local features and uses 3D models that are created in an off-line modeling process. A probabilistic model of the errors, which occur in the 6D localization based on local features, is directly derived from the pose reconstruction procedure. Experimental results from an household scenario illustrate the effectiveness of our approach.**

## I. Introduction

Precise object localization in all six cartesian dimensions is essential to all service robotic scenarios in which the robot interacts with and purposefully manipulates the environment. Because of the significant uncertainties, which cannot be avoided in realistic application scenarios, we expect that isolated sensor measurements rarely provide information precise enough for e.g. grasping an object in a cluttered scene. We therefore regard perception as a sequential estimation process and follow a Bayesian filtering methodology to fuse multiple sensor readings over time. Equation (1) describes the recursive estimation of the belief $Bel(x_t)$ at time $t$ over the world state $x_t$, $z_t$ denotes the system's sensor measurements and $u_t$ represents the system's actions that modify the overall world state $x_t$ (here $\eta$ is a normalization constant to be chosen so, that the resulting belief distribution integrates to 1).

$$Bel(x_t) = \eta\ p(z_t|x_t) \int p(x_t|u_t, x_{t-1}) Bel(x_{t-1}) dx_{t-1} \tag{1}$$

In this paper we present the details of the sensor model $p(z_t|x_t)$ used in this filtering process along with the detection methodology which allows for accurate 6D object pose estimation. We use Lowe's SIFT algorithm [1] for determining local, scale-invariant features in images. The application of the SIFT algorithm on both left and right images from a calibrated stereo camera with consecutive matching of the detected feature points from both images allows for the reconstruction of 6D object poses. Our appearance- and model-based approach consists of two separate stages: *Model generation* and *Object recognition and pose estimation*.

Thilo Grundmann, Robert Eidenberger and Georg v. Wichert are with Autonomous Systems, Corporate Technology, Siemens AG, D-80200 Munich, Germany


Fig. 1. A precise model of the errors that occur during 6D object localization is essential for many robotic tasks.

Model generation is an off-line process, where the object database is generated by extracting the essential information from training data. Object pose estimation at run-time is based on matching local features extracted from the current camera images with features from the object database. Based on the statistics of the matched features our system computes probability densities over object poses. The 6D pose is described by 3 translational and 3 rotational components and is formulated in the continuous domain.

The robust and accurate object recognition system presented here has been developed for the anthropomorphic service robot shown in figure 1. It is the basis for various other research in the area of object manipulation [2][3], perception planning [4] and physical object dependency analysis [5].

The remainder of the paper is organized as follows: Section II outlines current state of the art approaches to model-based object recognition, model generation and probabilistic models for stereo vision. In Section III our model generation procedure and the method for object 6D pose estimation will be described. Section IV explains the probabilistic sensor

model which is required for multi view fusion. Finally, in Section V, we report on experiments which demonstrate the proposed theoretical concepts on real data.

## II. Related work

Object recognition and localization has attracted interest of the scientific community for a long time. Early attempts restricted the pose dimension to one [6] or three [7][8].

Point correspondence from wire-frame model corners were used by Kragic [9] to find full 6D poses using a mono camera and POSIT [10]. This particular approach is restricted to objects that can be modeled by wire frames, but the underlying concept could be generalized by using interest points like SIFT [11] or SURF [12] that can be found numerously on textured objects.

All approaches to 6D pose estimation that use interest points belong to the class of model based recognition methods and therefore need a method to create the required models.

Gordan and Lowe [13] proposed a system that constructs 3D model of sift features from arbitrary object images using bundle adjustment off-line and determines the 6D pose through the use of RANSAC [14] and the Levenberg-Marquardt algorithm.

Azad et. al. presented a method stereo vision based [15] for full 6D pose retrieval of textured objects using classic SIFT interest points. The method requires the objects to possess flat surfaces for the stereo recognition and no empirical evaluation of the accuracy of the pose is given.

Recently Collet et al. presented an object localization system [16] using a monocular camera, based on SIFT features, which uses RANSAC and mean shift clustering to generate object pose hypotheses. They also describe an almost fully automatic process for the model generation and give some experiments with four objects where measured poses are evaluated against ground truth. The error is described by two histograms over the translational and rotational error.

Our localization approach is mostly comparable to the method of [16], using a comparable 3D model which is not restricted in the shape and is also based on SIFT features. Through the use of stereo image pairs and thus conceptually different methods for the pose determination a higher accuracy in pose measurement is achieved.

None of these localization methods have proposed a model to describe the pose errors that occur.

Cordes et. al. [17] have proposed a probabilistic model for 3D points which are constructed using a calibrated stereo triangulation set-up and use an improved Canny edge detector to find correspondences. In their approach, a model of the interest points is not available and therefore model uncertainties are not taken into account.

## III. Stereo sift 6D pose determination

Robotic manipulation in realistic scenes imposes high demands on the precision of 6D object localization, since objects will in general be close to each other, without large clearance between them.

For this reason we chose to use a stereo-based approach, which for obvious geometric reasons can be expected to deliver superior pose accuracy compared to monocular approaches [18].

In this section we first describe the model generation process and the measurement of the 6D pose using local features.

### A. Model Generation

Model generation aims at the acquisition of training data and its processing to generate object class models. It is essential to filter significant data and efficiently store it to enable reasonable processing times. The KIT object modeling center IOMOS [19] is used to acquire stereo images and corresponding camera poses.

The build process starts with computing the sift interest points (IPs) for each image and calculates 3D points by triangulation of IPs in each stereo image. Then matches over all images are determined and equivalence classes from these matching IPs found. At last each equivalence class is represented by one descriptor and one 3D location. All these equivalence class representatives together build up the model.

In detail the process comprises the following steps:

1) SIFT interest points calculation
   The SIFT interest points of each image form the base for the process. Each SIFT interest point $s^i = (u, v, s, o, d)$ consists of the 2D location $(u, v)$ in the image, its scale $s$ and its orientation $o$. The 128 dimensional descriptor is denoted by $d$. For each training image the set of interest points in the left camera image $S_l$ with $S_l = \{s_l^1, ...s_l^n\}$ is determined. The interest point set $S_r$ for the right camera image is acquired respectively.
2) Triangulation
   3D points are computed for corresponding IP's in the left and right images. For every IP in a left image $V_l$ we compute the epipolar line $L_e$ in the right image $V_r$ and determine the subset $S_r^e \subset \{s^i \in S_r | dist(s^i, L_e) < \varepsilon_E\}$ with $\varepsilon_E$ as the maximum epipolar distance. Then, sift descriptor matching is performed. It is important to do the epipolar examination before the sift matching step. This way the set of possible matches is constrained to a region in the image. IPs with similar descriptors from other parts are not considered and cannot distort the result. For each matched IP pair $(s_l^i, s_r^j)$ the corresponding 3D location and orientation are computed using triangulation [20] and transformed into the objects coordinate system to get the spatial feature representation:
   $$s_\# = (x, y, z, x', y', z', s, o, d). \quad (2)$$
   The first three elements $x$,$y$ and $z$ denote the translational coordinates, $x'$,$y'$ and $z'$ represent the direction from where the interest point is visible. Scale $s$, orientation $o$ and descriptor $d$ are equal to the parameters of the 2D interest point.
3) Equivalence Relation

The next step aims at partitioning this set of 3D points into subsets originating from the same physical point. This equivalence relation is seeded from IPs with fitting appearance and location and is completed by the transitive closure.

First candidates are found by appearance. The enormous amount of data with on average 850 3D locations for each training view, 280000 for each object, can be handled efficiently by means of a kd-tree. Using Euclidean distance in the sift descriptor space the nearest $n_N$- neighbors in the kd-tree are searched for each $s^i_\#$. $n_N$ depends on the sampling rate of the training data as sift descriptors are only invariant within a limited angular range and each face of the object is only seen from a certain range of camera positions. It was set to 150 in our case.

Then the candidates are checked for their spacial fit by calculating the Euclidean distance in the image. If this distance, which resembles the expected reprojection error, exceeds a threshold the candidate is rejected. In rare cases it happens that two IP's from the same view are in one equivalence class. These IP's are removed. A standard connected component algorithm is used to compute the transitive closure.

4) Subdivision and Representatives

We then seek a simple representation for each equivalence class above a minimal size $v_{min}$. Very small classes (eg. $v_{min}$=4) are discarded to suppress IPs of low value for the recognition process and noise.

When evaluating classes one finds the locations clustering well, but descriptors to spread considerably. This is not surprising since we do see most points from a wide angle range were the sift descriptor cannot be assumed to be invariant. Instead of more complex density models we favor a very simple representative, which is a simple mean for location and normalized mean for descriptor. To make this simple model suitable we sacrifice the simple relation, where one class corresponds to one physical point, and split classes with k-means until they can be represented as spheres. This considerably simplifies and speeds up the recognition process.

The full model for one object needs about 5% of storage of the initial sift features. This enables fast recognition since big databases can be held in RAM completely.

### *B. 6D Object Localization*

The 6D localization process consists of the following steps:

1) Extract IPs and compute SIFT feature vectors.

For each of the stereo images $V_l$ and $V_r$ a corresponding set of interest points is calculated $S_l = \{s^1_l, ..., s^n_l\}$ and $S_r = \{s^1_r, ..., s^m_r\}$.

2) Find correspondence to object models (Fig. 2).

For all elements $s^i_l \in S_l$ we search up to $p_{mm}$ multiple matches $c^k = \{i, j\}$ with a 3D feature from the model database $s^j_\# \in M$. The criterion for a match is that the Euclidean distance in descriptor space is below an absolute threshold $p_{tm} = 0.3$.

To speed up the search for matches, the descriptors from the database $M$ are structured in a kd-tree using the ANN library. To increase the performance, the nearest neighbor search is approximated.

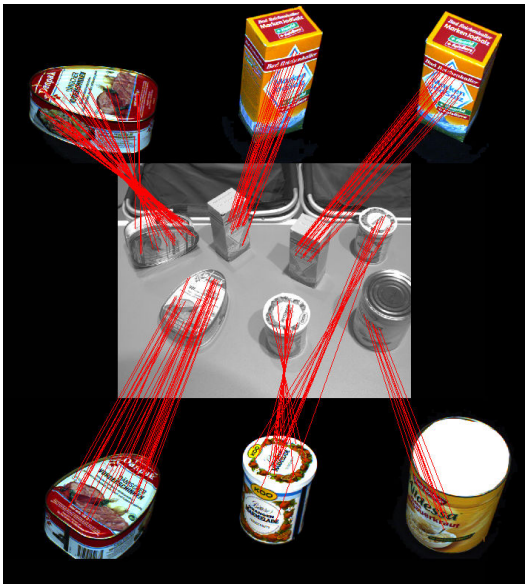


Fig. 2. Recognition principle: finding matches between interest points $s_\# \in M$ from given models and interest points detected in one image $s \in S_l$ (step 2)

3) Stereo reconstruction of 3D interest point locations

For all matchings $c^k = \{i, j\}$, we try to find multiple corresponding interest points $s^h_r \in S_r$ on the right interest point set $S_r$. The epipolar constraint is used in the same manner as described in Section III-A, but after the epipolar spatial restriction, a relative multi match is used. This has to be done to account for the classic situation where multiple instances of the same object class are placed side by side on a board, leading to multiple similar features on the epipolar line $L_e$. After this procedure we obtain a set of $l$ 3D sift points $S_\# = \left\{s^1_\#, ..., s^l_\#\right\}$ by triangulation.

4) Clustering

To account for scenes with large numbers of identical objects and to deal with the high number of erroneous 3D sift features, we construct initial pose estimates $P = \left\{p^1, ..., p^n\right\}$ from $S^t_\#$. This is done by randomly choosing non collinear triplets of 3D interest points from $S_\#$ and check whether their mutual Euclidean distances match those in the model database.

Within the 6D space of the initial pose estimates $P$, qt-clustering [21] is performed to find consistent 6D pose estimates. The clusters $c$ consist of 6D poses $p^i$ each of which corresponds to a triplet of 3D interest points. That way each cluster describes a set of 3D interest points $S^c_\# \subset S_\#$.

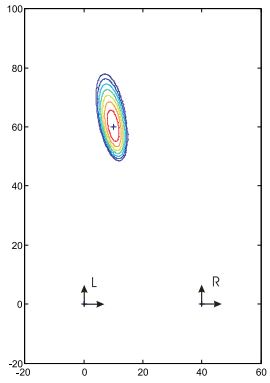



Fig. 3. Contour plot (bird's eye view) of the sensor model $p(z|x)$ as a function of $x$ (2D state without rotation) for a single feature point detected in both images. L and R are the two camera coordinate systems.

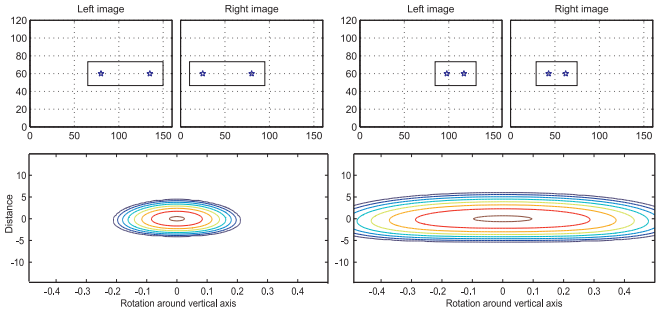



Fig. 4. Comparison of synthetic sensor model results for two objects of different width, represented each by two 3D feature points. Both objects are at the same distance from the stereo camera system. The upper plots show simulated camera images, the lower plots show the measurement uncertainties for both situations. Here we plot $p(z|x)$ as a function of $x$ in the distance vs. vertical rotation space. As expected, the rotational component of the smaller object's pose-estimate has a significantly higher uncertainty.

The resulting 3D interest points from each cluster $c$ form a specific, spatially consistent object pose hypotheses. A 6D pose estimate could be computed by e.g. a least squares fit of the cluster point cloud $S^c_\#$ to the corresponding 3D database sifts $M^c \in M$ (Fig. 2). The 2D interest points that were used to generate the matched 3D interest points $S^c_\#$ and the database points $M^c$ form the input data for the probabilistic sensor model described in Section IV.

## IV. Probabilistic Sensor Model

As already pointed out in Section I, the sensor model $p(z_t|x_t)$ plays an important role in the sequential estimation process that models our system's perceptual activity[1]. It is the basis for the fusion of sensor readings over time into a consistent model of the robot's environment.

Assuming correct correspondences between image and model interest points, as provided by the method described in section III-B, the dominant sources of pose estimation errors are the localization of the interest points in both images and the uncertainty of the 3D interest points in the model database.

For an arbitrary object pose hypothesis $x$ in 6D camera coordinates $C_V$ the projected image coordinates $s_{V_n}$ in the image $V \in V_l \cup V_r$ can be computed assuming a standard pin hole camera model $h_p(f, T_{C_V}(x), M^c_n)$ as

$$s'_{V_n} = \begin{bmatrix} s'_{V_x} \\ s'_{V_y} \\ s'_{V_z} \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} T_{C_V}(x) M^c_n \quad (3)$$

$$s_{V_n} = \begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} \frac{s'_{V_x}}{s'_{V_z}} \\ \frac{s'_{V_y}}{s'_{V_z}} \end{bmatrix} = h_p(f, T_{C_V}(x), M^c_n) \quad (4)$$

Here $f$ is the focal length of the camera and $T_{C_V}$ is the homogeneous transformation from object to camera coordinates of the left and right images. For the set of N interest point correspondences $\{S^c_{\#n}, M^c_n\}$, $n \in [1, N]$ and an arbitrary object pose hypothesis $x$ in 6D camera coordinates we compute the observation density $p(z|x)$ assuming a mutually uncorrelated detection of the interest points $s_{l_n}$ and $s_{r_n}$.

$$p(z|x) = \prod_1^N p(s_{l_n}|x) p(s_{r_n}|x) \quad (5)$$

We assume the location uncertainty of the 3D model database points $M^c_n$ to be Gaussian with a covariance $\Sigma^n_M$ and the detection of the interest point locations $s_{V_n}$ also to be affected by Gaussian error with a covariance $\Sigma^n_s$. $\Sigma^n_M$ is determined during step 4 of the model generation process, see Section III-A. $\Sigma^n_s$ has been empirically determined. We linearize the camera model to project the Gaussian model error $\Sigma^n_M$ at the interest point location $M^c_n$ into the image planes of the two cameras. Under these assumptions $s_{V_n}$ is approximately normally distributed according to

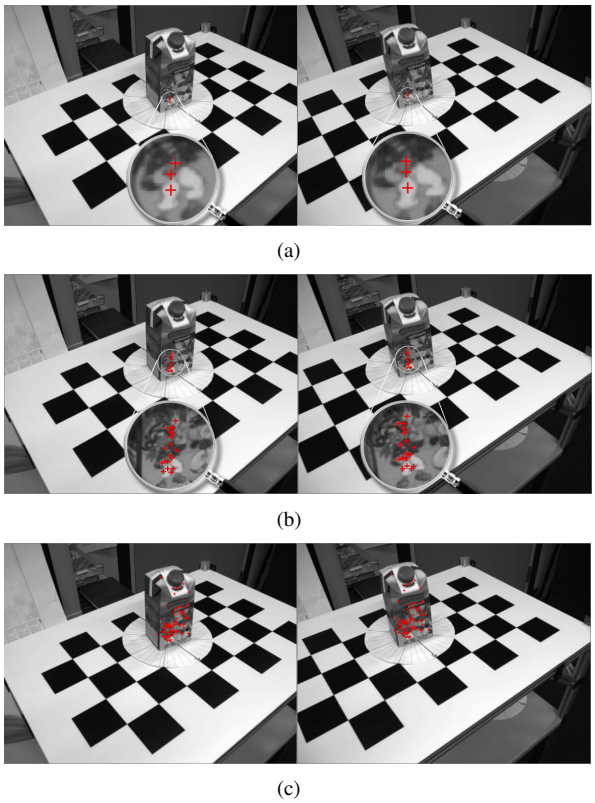
$$p(s_{V_n}|x) = \mathcal{N}(h_p(f, T_{C_V}(x), M^c_n), \Sigma^n_s + J_{h_p} \Sigma^n_M J^T_{h_p}) \quad (6)$$

$$J_{h_p} = \partial h_p(x) / \partial M^c_n \quad (7)$$

where $J_{h_p}$ is the Jacobian of the perspective projection (4) with respect to the 3D interest point locations $M^c_n$. This is used in (5). Figure 3 illustrates the result for a simple synthetic example.

The sensor model captures the dominant uncertainties of the pose reconstruction process based on local point features. Especially the effect of the geometric configuration of the detected 3D feature points on the estimated object orientation is correctly covered as shown in Fig. 4 using a synthetic example. Narrowly spaced feature points lead to significantly increased angular uncertainty in the pose estimate, compared to spatially wider distributed feature points. Narrow or partially occluded objects frequently cause such problems in our household scenarios (see Fig. 1) leading to significant pose-estimation errors in reality. If such uncertainties are correctly represented, they can be compensated by performing additional sensor measurements [4] in order to reach the precision required e.g. for grasping objects in cluttered scenes.

[1]While in general the world state comprises the poses of several objects, we consider only a single object in this section, and use $x$ to refer to its 6D pose vector, also leaving out the temporal index $t$ for notational simplicity.



(a)

(b)

(c)

Fig. 5. Stereo images of one scene with different sets of sift points (red crosses) plotted in. (a): 3 stereo interest points are used. (b): 18 stereo interest points are used. (c): 90 stereo interest points.

## V. Experimental results

The vision system on our experimental robot (Fig.1) consists of two AVT Pike F-145C cameras with a resolution of $1388 \times 1038$ pixels each, equipped with 8.5mm lenses and mounted with a disparity of roughly $0.12m$. Precise intrinsic and stereo calibration of the cameras is essential for our algorithms, so they were carried out with the Camera Calibration Toolbox for MATLAB, using about 60 stereo image pairs of a custom made highly planar checker board calibration pattern.

To use the probabilistic model within this stereo application we determine the two covariances $\Sigma_M^n$ and $\Sigma_s^n$ that parameterize our measurement model $p(z|x)$: We assume that the probability distribution is the same for all 3D interest points in the model, and we compute $\Sigma_M$ from 18743 interest points clusters that form the complete model of one of our test objects. The Gaussian error in the SIFT feature point localization $\Sigma_s$ is set to have a standard deviation of $0.25$ pixel in both directions and was empirically determined.

To show the influence of different numbers and spatial constellations of interest points on the estimated pose distribution we chose to use the same scene three times (Fig. 5) and artificially varied the number of interest points used for estimating the object pose. Since only the number and spatial distribution of the selected features was varied, whereas the images were identical, this allows for a direct comparison of the results.

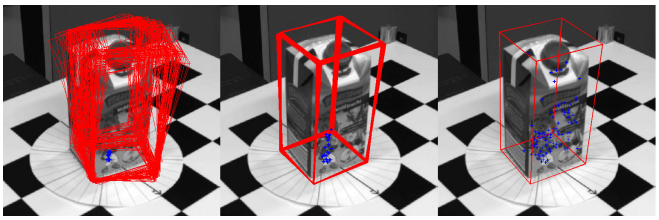


Fig. 6. Samples, drawn from $p(z|x)$ for all three sets of interest points. It is clearly visible, that the pose estimation accuracy increases with a growing number of interest points used and their wider spatial distribution.

We chose three different sets of interest points: the first consisting of only 3 stereo points that lie in a very small area (Fig. 5a), the second consists of 18 stereo points still covering a limited spatial region of the object (Fig. 5b). The third set consists of 90 interest points that are spread over the whole surface of the object (Fig. 5c).

In Fig. 6 the resulting measurement model $p(z|x)$ for all three situations is visualized by drawing samples from the model and plotting the corresponding bounding box of the object for each of the samples into the image.

The results show object pose distributions as expected, the more restricted the number and the area of the interest points is, the higher is the pose uncertainty. The second example shows, as a consequence of the uncertainty which is caused by the locality of the interest points in the lower area of the object, that the estimated orientation error in this situation is still significant, which is clearly visible in Fig. 6 (center). In contrast, the rather wide spread of the feature points across all of the object in the third example leads to a very precise localization of the object as indicated by the peaked distribution of the object pose estimate.

Figure 6 displays, how the sensor model captures the actual pose measurement uncertainty when artificially varying the number and spatial distribution of the interest points available for pose estimation. In contrast, Fig. 7 shows a realistic result for a complex scene with several, partially occluded objects.

The measurement uncertainty visualized by the overlaid object bounding boxes – each of which corresponds to a single sample drawn from the measurement model – varies across object instances as a consequence of the number of detected interest points available for matching. For some objects, overexposure prohibits the detection of most interest points and therefore greatly increases the pose estimation uncertainty, which is correctly captured by the measurement model presented in this paper. On the contrary, those objects with proper exposure like the salt box in the image center, where many interest points were detected and successfully matched to the object model, allow for rather precise pose estimates. While in in case of the scene in Fig. 7 the limited image quality resulted in higher uncertainty, similar effects will be caused by occlusions as well as larger object-camera distances both confining the detection of the feature points needed for pose estimation.

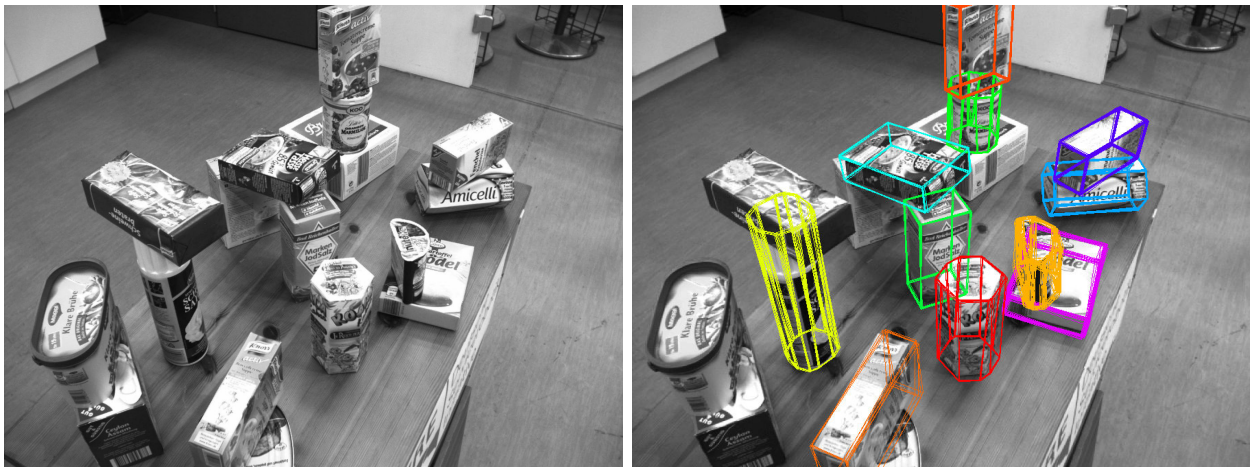


Fig. 7. A complex scene composed of several household objects in various configurations. In the left image samples drawn from the observation model $p(z|x)$ were overlaid onto the source image (colors indicate object class IDs; not relevant here).

## VI. Summary

Interest point based methods have become widely used in the field of 6D object localization. In general, their accuracy is high, however it varies strongly with the number and spatial distribution of the interest points used for pose reconstruction. Under realistic assumptions for actual applications this can and will be an issue for real systems, due to varying environmental conditions influencing sensor measurements and other real-world effects, like e.g. occlusions in cluttered scenes.

Our work aims at an overall perception architecture explicitly representing the actual uncertainties in a probabilistic framework. This enables a task-oriented assessment of the quality of the available knowledge and allows for the active refinement of the system's belief state over time [4]. Sensor models accurately modeling the actual measurement uncertainty are key in this context.

We presented such a measurement model for interest point based pose estimation, that explicitly models the actual sources of error. We used Gaussian models for the error in the 2D interest point localization as well as in the 3D models of the objects which are known to the system.

## References


[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] 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.

[3] Z. Xue, J. Marius Zoellner, and R. Dillmann, "Grasp planning: Find the contact points," in *IEEE International Conference on Robotics and Biomimetics*, 2007.

[4] R. Eidenberger, T. Grundmann, and R. Zoellner, "Probabilistic action planning for active scene modeling in continuous high-dimensional domains," *IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, 2009*, 2009.

[5] T. Grundmann, R. Eidenberger, and R. Zoellner, "Local dependency analysis in probabilistic scene estimation," in *ISMA 2008. 5th International Symposium on Mechatronics and Its Applications*, Amman, Jordan, 2008, pp. 1–6.

[6] S. Nayar, S. Nene, and H. Murase, "Real-time 100 object recognition system," in *Proc. IEEE International Conference on Robotics and Automation*, vol. 3, 22–28 April 1996, pp. 2321–2325.

[7] J. Zhang, R. Schmidt, and A. Knoll, "Appearance-based visual learning in a neuro-fuzzy model for fine-positioning of manipulators," in *ICRA*, 1999.

[8] J. A. Walter and B. Arnrich, "Gabor filters for object localization and robot grasping," in *ICPR*, 2000, pp. 4124–4127.

[9] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Seoul, Republic of Korea, 2001, pp. 2460–2465.

[10] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision, Springer Netherlands*, vol. Volume 15, Numbers 1-2, pp. 123–141, 1995.

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

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," vol. 110, no. 3. New York, NY, USA: Elsevier Science Inc., 2008, pp. 346–359.

[13] I. Gordon and D. G. Lowe, "What and where: 3d object recognition with accurate pose," *Toward Category-Level Object Recognition*, vol. Lecture Notes in Computer Science, pp. 67–82, 2006.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[15] P. Azad, T. Asfour, and R. Dillmann, "Stereo-based 6d object localization for grasping with humanoid robot systems," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, San Diego, CA, USA, 2007.

[16] A. Collet, D. Berenson, S. Srinivasa, and D. Ferguson, "Object recognition and full pose registration from a single image for robotic manipulation," in *ICRA 09*, 2009.

[17] K. Cordes, P. Mikulastik, A. Vais, and J. Ostermann, "Extrinsic calibration of a stereo camera system using a 3d cad model considering the uncertainties of estimated feature points," in *The 6th European Conference on Visual Media Production (CVMP)*, 2009.

[18] P. Azad, T. Asfour, and R. Dillmann, "Stereo-based vs. monocular 6-dof pose estimation using point features: A quantitative comparison," in *Autonome Mobile Systeme 2009*, ser. Informatik aktuell. Karlsruhe: Springer, 2009.

[19] Z. Xue, A. Kasper, J. Zoellner, and R. Dillmann, "An automatic grasp planning system for service robots," in *14th International Conference on Advanced Robotics (ICAR)*, June 22nd - 26th, 2009.

[20] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[21] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data: Identification and analysis of coexpressed genes," *Genome Res.*, vol. 9, no. 11, pp. 1106–1115, November 1999.