Proceedings of the 2006 IEEE/RSJ
International Conference on Intelligent Robots and Systems
October 9 - 15, 2006, Beijing, China

# Towards Open-Ended 3D Rotation and Shift Invariant Object Detection for Robot Companions

Jens Kubacki & Winfried Baum
Robot Systems
Fraunhofer IPA
Stuttgart, Germany
{jbk, wmb}@ipa.fhg.de

*Abstract* - **Robot companions need to be able to constantly acquire knowledge about new objects for instance in order to detect them in the environment. This ability is necessary since it is hard to predict what objects the robot may face in the operation phase during development. This paper presents ideas and results on two topics. The first topic is on the design of an open-ended object detection system that uses scale invariant feature key-point descriptors that are trained with a one-class radial basis function support vector machine. Unlike using other classifier-based approaches our method does not assume the number classes to be known *a priori*. The method is shown to be stable against full 3D rotation of the object relative to the sensor. The second issue in this paper deals with a solution on how to provide new object information to the robot. A modern range imaging sensor in conjunction with a conventional color imaging sensor is used for a first figure-background separation. The experiments presented support the basic statements in this paper. Conclusions are drawn and future work is addressed.**

***Cognitive robot companion, object recognition, object detection, active learning, interactive learning, support-vector learning, shift-invariant feature transform, range imaging sensor, one-class SVM.***

## I. Introduction

We develop cognitive robot companions that are able to learn new skills and tasks in an active, open-ended way and to grow in their capacities in constant interaction and co-operation with humans and autonomously [1]. See figure 1 for an example of a modern robot companion.

One very important visual competence is ability to detect objects in the environment e.g. in order to grasp them. The problem is to estimate the position of an object to be grasped in the coordinate system of the robot. An assumption is that internal models representing these objects can not be pre-programmed to the robot but must be learned either through interaction with the human or autonomously through acting on the objects. The learning capability needs to be *open-ended*. Hence, there should be no limiting system parameters on the number of objects to be learned. Also, strategies for interactive and active autonomous learning should support open-endedness and keep the involvement or the "pedagogical effort" of the user in the role as tutor low. This might be an important usability aspect regarding future robot companions.

In this paper these two issues are addressed. First a method for recognition is proposed that uses shift invariant feature transform (SIFT) descriptors [2] that are learned with a one-class radial basis function support vector machine (RBF-SVM) [3] to allow for open-ended learning of object appearances in batch mode i.e. there is one distinct training period for each object in which all necessary appearances are provided. The one-class method allows that an object can be learned in the absence of training samples from other past or future objects. Hence, it is not necessary that the number of objects has to be known *a priory*. Training samples of known objects don't have to be kept in the system. This would be the case for a multi-class SVM scheme that is built up of several binary SVMs. Different approaches exist to build such a multi-class network [4].

Fig. 1 Example of a modern robot companion and related manipulation tasks: The Care-O-bot II includes a mobile platform, a 6 DOF manipulator arm, and a variety of sensors such as color cameras and laser range finders to enable the perception of objects and humans. Left: the robot picks up a bottle, right: the robot takes a juice box out of the fridge.

The second issue in this paper is a method that allows the robot to segment feature key-points from the background using a modern range imaging sensor [5] calibrated with a conventional color imaging sensor. The segmented key-points are fed into the one-class SVM to train the detection system.

In chapter II some references to related work are given. Chapter III describes the recognition method based on SIFT key-points and the one-class RBF-SVM and the method for a first figure-background separation. In chapter IV measurements, experiments and results are reported. Finally, chapter V summarizes the work, draws some conclusions and points to future work.

1-4244-0259-X/06/$20.00 ©2006 IEEE

## II. Related Work

The work presented here relates to some earlier work in learning object appearances [6] specifically to some more recent work in learning object appearances with SVMs e.g. [7] and [8]. The proposed method in this paper differs from previous work in that SIFT key-points are used in conjunction with the one-class RBF-SVM. Newer approaches to robot or "cognitive" vision stress the physical embodiment [9, 10] e.g. to solve the figure-background problem [9] or to enable autonomous learning through actively exploring object properties. In this work a modern time-of-flight range imaging sensor [5] is used for first figure-background separation. There is other work existing on using SVM classifiers in combination with key-points [11]. The aim is to retrieve abstract class knowledge. The work presented here differs in that the classifier is object-specific. The goal is to find a specific object in the scene on the basis of its appearances that are extracted in the learning phase. The generalization ability of the classifier is used to index object views or key-point descriptors with acceptable error rates and model corpus sizes. The novelty is that a one-class machine is used that separates well between the current object and unseen structures. In this paper the object appearances are shown the robot by the human tutor. A more autonomous active learning strategy is described in [12].

## III. Proposed Methods

### *A. Recognition method: key-point learning and indexing with a one-class SVM*

If an object is to be detected in the scene with a color imaging sensor there are various difficulties involved, among them:

- Rotation and translation of the object
- Different backgrounds
- Change of lightening conditions
- Effects of the near vicinity (occlusions, shadows)
- Change of the object appearance due to deformations

To be able to cope with these difficulties there are different approaches suggested in the computer vision literature. So-called *interest points* or *feature points* or sometimes referred to as *feature key-points* stressing their use as a *key* to recognition find locations in the image that optimize two properties: a high repeatability and high information content [13]. The feature key-points developed in [2] achieve high repeatability through the detection of scale-space extrema (see [14] for an introduction) in the difference of Gaussians space calculated from the image. The high information content is obtained through a histogram of orientations. These feature key-points have been shown to be suited for object recognition [2] and robot navigation [15]. A second body of research deals with learning algorithms that can be used for e.g. visual character recognition. Support vector machines (SVM) can be used to index images containing object appearances [6, 7]. The interesting fact about these approaches is that the learning machine is trained with raw object images without any complex pre-processing. An example of a data base that is used in this context is introduced in [16].

We propose an approach to the detection of a specific object under the difficulties that are described above that makes use of the results of both research areas: using modern feature key-points and a new SVM learning algorithm. See figure 2 for a discussion of this synthesis. If the object is shifted in $x$, $y$ and $z$ or rotated in $\gamma$ then the SIFT key-points are expected to be stable. However, if the object rotates in $\alpha$ and $\beta$ then we can observe two effects:

- New key-points enter the scene that could not be seen before and other key-points disappear.
- The key-point descriptors change according to the local 3D surface structure of the object and due to self-occlusion.

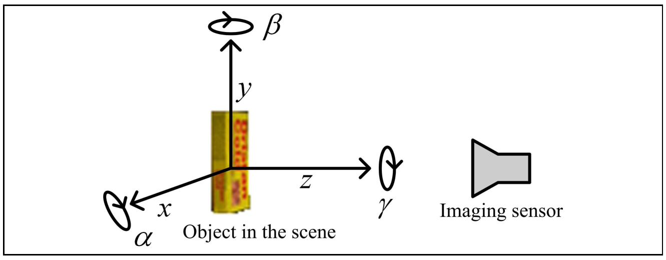


Fig. 2: An object placed in front of an imaging sensor and the coordinate frame attached. SIFT key-points are expected to be stable against shift in x and y and rotation around z. SVM-based detectors are typically stable against rotation around x and y.

It is expected that a precise model of how the feature key-point descriptors change while rotating in $\alpha$ and $\beta$ is difficult to obtain. The goal is that the SVM learning algorithm provides a good model of descriptor changes of a specific object in these cases. The suggested strategy is to learn pre-processed feature key-point descriptors with a one-class SVM. The pre-processing and the selection of the SVM model are described in the next two sub-paragraphs.

**Color pre-processing.** The SIFT filter is developed for typical gray-level images. To use it for color images we apply the SIFT filter to the value (brightness) channel of a HSV representation of the image. At the discrete point coordinates the hue and saturation values are extracted from the image and appended to the descriptor. The descriptors used here have 128 dimensions and the pre-processed descriptors 130 (i.e. including the two additional color components).

**SVM selection.** For multi-class separation using binary SVMs there are different strategies possible [4] that for instance combine several SVMs that each separate between two classes. In such multi-class systems the number of classes must be known *a priory* since the number of connected SVMs depends on them.

In contrast to such conventional classifiers one-class methods try to describe the distribution of only one class in hyperspace. An algorithm based on the SVM idea is published in [3].

Consider the set of "one-class only" training data:

$$x_1,\dots,x_l \in R^n \tag{1}$$

where $n$ is the dimension of the input space and $l$ is the number of samples (points) in this space. The one-class SVM algorithm finds a linear decision function

$$f(x) = \mathrm{sgn}((\omega \cdot \Phi(x)) - \rho) \tag{2}$$

where $\omega$ is the weight vector and $\rho$ the offset. This function separates between the origin (zero) vector and the samples expressed in feature space using kernel function

$$k(x, y) = (\Phi(x) \cdot \Phi(y)) . \tag{3}$$

This decision function separates the data from the origin with maximal margin and can cope with "outliers" through solving the optimization problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu l}\sum_i \xi_i - \rho \\ &\text{subject to } (\omega \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i ,\ \xi_i > 0 . \end{aligned} \tag{4}$$

Here $\xi_i$ is the distance of an outlier to the hyper-plane and $\nu$ is a parameter that controls the number of outliers tolerated. In this paper we use the radial basis function (RBF) kernel:

$$k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} . \tag{5}$$

This kernel function was used in experiments on character recognition [17] and also in experiments on machine fault detection reported in [18]. The RBF kernel has a number of advantages compared to other kernels stated in [21] and is suggested as the first choice to start with when encountering new learning problems. Also there exists already a strategy for selecting the kernel parameter $\sigma$ such that low error rates can be achieved. Additionally, this kernel function can make suited non-separable problems separable through increasing $\sigma$ and it is the only parameter to be optimized. The parameter $\nu$ bounds the tolerated errors. In the setting here it is seen to depend on the complexity of a specific object that is to be learned. The parameter $\sigma$ controls the kernel with. Therefore, this parameter effects all object models that are expressed in kernel space. That leads to the idea that $\sigma$ should be optimized by the developer whereas $\nu$ should be found dynamically (i.e. object dependent). The proposed method to find a good $\nu$ is a variant of the strategy presented in [18] that constantly increases $\sigma$ until the rate of positive true responses $r_{pt}$ exceeds $1-\nu$. In our approach, $\nu$ is constantly increased until the same criterion holds for a fixed $\sigma$. Furthermore, the data is pre-clustered. The first sample "opens a bin", the next samples fall either into an existing bin in the case that the Euclidian distance to the first bin sample is below a distance threshold or they open a new bin. Experiments show that the false positive responses $r_{fp}$ can be reduced significantly using this simple pre-clustering technique. In prediction mode several SVM models are effectively ORed together.

### *B. First figure-background separation based on range segmentation*

To acquire feature key-points that belong to the object a first figure-background problem needs to be solved. For this a range imaging sensor [5] is used. Figure 3 depicts the updated sensor head of the robot shown in figure 1. The range imaging sensor is mounted next to a conventional color imaging sensor.

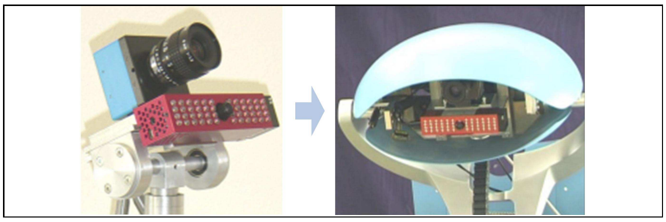

Fig. 3: Left: the time-of-flight range imaging sensor and the conventional color imaging sensor, right: both sensors mounted on the robot Care-O-bot II for the experiments that are reported in this paper.

With this sensor setting a *direct calibration* [20] can be used to obtain the color information of 3D points in the environment. The direct calibration is given by the following equations:

$$x_R = p_3 x_C + p_2 \tag{6}$$
$$y_R = p_1 y_C + p_0 \tag{7}$$

where $(x_R, y_R)$ are the coordinates in the range imaging sensor, $(x_C, y_C)$ are the coordinates in the color imaging sensor and $p_0 \dots p_3$ are parameters to be estimated. Note, that this simplified method is not suitable for arbitrary settings.

The direct calibration can be used to cut out those pixels or feature key-points that occur in a certain region of 3D space. We call this *range segmentation*. To estimate the location of this region in space there are different scenarios possible. Figure 4 illustrates two of them.

In an *interactive* mode the object is shown to the robot by the human tutor and the location of the region surrounding the object is assumed to be available through a body-part tracking module. Such modules utilizing range imaging sensor are becoming available. If the human is detected in the environment [21] a cylinder model is fitted into the corresponding point cloud [22] and then tracked. The readings of the estimated kinematics model include the arm end-position in which also the object occurs if it is grasped. Current work focuses on improving body-tracking such that is robust in the presence of grasped objects.

In the case of the *active* autonomous strategy either a learning table is assumed to be known to the robot in terms of position or geometry [11] or the object is passed to the robot. Simple object passing can be realized using the local sensory systems integrated in the gripper (see [23] for a description of the sensors). If the object is placed on the learning table it can be segmented by only considering the 3D space above the table and fitting a 3D model sufficient for a first grab [11]. In both cases finally the robot possesses the object and can rotate it in order to produce new samples of appearances.

Fig. 4: Illustration of the basic components for the interactive and active object learning scenarios. In the interactive mode the human holds the object; in the active mode the robot holds it. In both cases the rough location of the region surrounding the object in 3D space is available (see text). The object is then rotated in order to produce sample appearances.

In interactive and active learning there are disturbing visual components close to the object stemming from the human's hand or from the robot's gripper. The feature key-point acquisition needs to be robust to this. One method to cope with this is to train the key-points of the human hand or the robot's gripper first and then remove corresponding feature key-points by indexing them with the SVM method. In the experiment in section IV we did not explicitly remove these feature points.

## IV. Experiments And Results

### *A. Statistical Measurements on SIFT key-points in the COIL object data base*

First statistical measurements on the distribution of SIFT key-points over the COIL data base [16] were performed in order to evaluate if these key-points can represent all objects of the data base. See figure 5 for examples of object images.

Fig. 5: Examples of images of the COIL data base [16]. Top row: Views of object no. 1 at 0°, 30°, 60°, and 90°. Bottom row: Object no. 2 to object no. 5 all at 0°.

The COIL data base contains 7200 images of 100 diverse objects. There are 72 views per object, respectively. The objects were placed on a turning table that was rotated about 5° from view to view. All images were pre-processed such that there is uniform black background and the image size is the smallest enclosing square of the current object view. All images have the same size of 128x128 pixels.

In figure 6 the distribution of SIFT feature key-points over the value channels of all COIL images is shown. The software available at [24] is used for key-point extraction with the default parameters. The COIL data base is used in JPEG format.

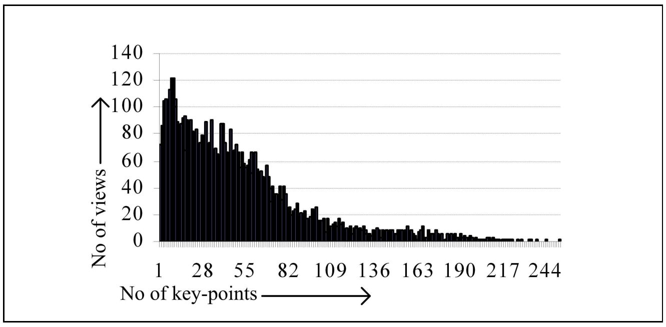


Fig. 6: The distribution of SIFT feature key-points over the value (grey) channels of the images of the COIL data base.

In average there are about 56 feature key-points in the value (brightness) channel per view of a COIL object. The lowest number of key-points is one (29 entries) and the largest number is 263. In average a complete object consists of 4026 key-points (without any grouping). The complete data base contains 402,556 key-points. Since there are only 29 images containing one key-point it is assumed that all objects can be represented using the SIFT key-points.

### *B. Experiments in complete image and key-point learning with a one-class RBF-SVM*

The first experiment conducted is aimed to show that the one-class RBF-SVM with the proposed parameter search strategy and pre-clustering can be used to classify complete COIL images as done in [7] with a classifier network. The images were converted into 32x32 sized gray-level images using the load and resize functions from openCV [26]. The normalized 1024 dimensional vectors of all COIL objects were learned with the one-class SVM from [25]. The parameter $\nu$ was iteratively increased in 0.002 steps starting from 0.05 for each object and $\sigma$ was set to $\sigma = 9$ . The true positive response of the one class SVM is about 88% and the false positive response 8%. In average 18.66 support vectors are needed from the initial 72 training views to describe the data.

This shows that the one-class SVM is well suited for complete image indexing. Even though the rates are worse than those in [7] it is seen positively that the one-class machine can be used with its advantages described. The rates were further improved through pre-clustering. The distance threshold was set to 5.0 and groups with less than 2 entries omitted. In average there were 5.69 groups per object. The true positive response rate achieved was 85% and the false positive response was only 1%. The average number of support vectors was 26.75. Therefore, it is possible to decrease the false positive response rate with pre-clustering. However, then the model sizes increase. These tests were performed over all 100 objects of the data base (i.e. 4950 comparisons).

In the next experiment the normalized 130 dimensional feature key-point descriptors were fed into the SVM including the two components for hue and saturation. For RGB to HSV conversion also the openCV library was used. The key-point descriptors were classified and the complete images. The images were judged to belong to the object of interest if they contained more than 40% of key-points that index the object. The step width to increase $\nu$ was chosen to be 0.05 to save processing time and $\sigma$ was set to $\sigma = 0.6$. Each of the first five objects was tested against all other 49 objects from the first fifty objects of the data base. For the discussion we use $r_{tpf}$ and $r_{fpf}$ to denote the true and false positive responses on feature key-points and $r_{tpi}$ and $r_{fpi}$ meaning the true and false complete image responses. The results are shown in table 1. The true positive responses are between 55% and 70%. They could be increased by selecting a lager $\sigma$. But then also the false positive response rates increase. For a real improvement one has to select a smaller step width for the $\nu$ search.

TABLE I
RECOGNITION RATES FOR THE FIRST FIVE COIL OBJECTS AGAINST THE FIRST FIFTY OBECTS

| Learned obj. no. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| no. key-pts | 6923 | 476 | 3425 | 1304 | 4641 |
| no. groups | 48 | 11 | 44 | 36 | 57 |
| no. SVs | 3812 | 240 | 1833 | 970 | 2481 |
| $r_{tpf}$ | 68% | 70% | 67% | 55% | 70% |
| $r_{fpf}$ | 12% | 1% | 8% | 4% | 11% |
| $r_{tpi}$ | 100% | 94% | 100% | 90% | 100% |
| $r_{fpi}$ | 4% | 0% | 1% | 0% | 3% |

The rates on the classification of the complete coil images are acceptable with more than 96% in average for the true responses and less than 2% in average for the false positive responses. The latter number is a reliable estimate since 49x72=3528 images were used to measure it per object. A typical COIL image is processed in about 0.1 second on a 3 GHz Pentium (without the time needed for feature key-point extraction).

The experiments show that key-point indexing with one-class RBF-SVMs is still possible invariant to the rotation of the turn table. This in turn leads to the conclusion that a complete invariant recognition system is possible to design accounting for the 3D rotation and shift cases shown in figure 2. However, the number of support vectors is about 41% out of the original feature key-points per object. Though there might be better parameters for the SVM model it seems that the distribution of the descriptors is quite difficult to learn. For the complete image tests the rate of the support vectors is only 26% (18.66 out of 72) yielding a combined error of 20% of the views. The key-point descriptors are designed to be highly distinctive whereas the SVM tries to generalize. Therefore, we conducted new experiments where only the scale-space extrema detection of the SIFT method was used and the descriptors were simply fixed-sized image patches with their center at the key-point's location. The three RGB channels of the images were processed separately. Hence, the SVM was provided with much more information to find discriminative aspects automatically. With only 6.5% of support vectors out of all key-points a true positive response rate of 77% and a false positive response rate of 35% were achieved over the first five objects of the data base. This does not seem very good but in future we want to investigate this further since these descriptors are much faster to compute. Maybe a better kernel will enhance the results.

*C. Experiments on key-point learning through range segmentation*

Next, we conducted an experiment related to the first figure-background separation that can be realized using the range imaging sensor in order to allow for the scenarios presented in section III. Two objects (shown in figure 7) were trained for a proof-of-concept of the range segmentation and the indexing method. The human supervisor showed the front-side of the two objects to the robot from 20 different arbitrary viewpoints by rotating them slightly. There were no pre-assumptions other than the specified range interval. The feature key-points were extracted using the range segmentation method. Only the gray-level channel was used. For detection the positively indexed features were grouped in clusters and the center position (vector mean) of the largest cluster gave the detection result.

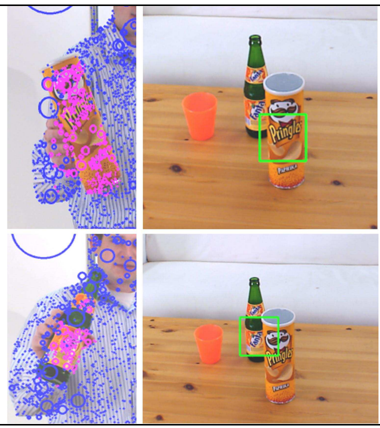

Fig. 7 Top left and bottom left: the two objects and SIFT key-points of the background (in blue) and segmented key-points (in magenta) using the proposed range segmentation. Top right and bottom right: The two objects successfully detected in the scene denoted with a green box.

For a small set of scene shots the detector worked successfully. Yet this has to be scientifically evaluated with a new object data base including range images. With the calibration to the range imaging sensor an approximate object position in 3D co-ordinates relative to the robot is available.

## V. SUMMARY, CONCLUSIONS AND OUTLOOK

We have shown that the one-class RBF-SVM can learn to index key-points. It has been argued elsewhere (e.g. [2]) that the SIFT key-points (and also other modern key-point types) are stable against shift and orthogonal rotation of object in front of the imaging sensor. Rotation around the remaining axes uncovers new feature key-points on the object and other key-points disappear. During rotation around these axes the descriptors behave in a way that is difficult to express analytically. Our method learns the key-point descriptor models using the one-class SVM with a radial basis function kernel including the parameter optimization strategy and pre-clustering. Key-point descriptors can be indexed correctly with relative high probability. However, the number of support vectors is relatively large. Typical COIL images can be classified with only a small error of less than 2% related to the more important false positive response rates and more than 96% considering the true positive response rates. It is not stated that these are the final "best" numbers since a finer search for $\nu$ , a better selection of $\sigma$ or a change of the kernel can enhance the results. There is no information on the number of classes needed *a priory* but the nice properties of SVM generalization can be used and the size of the model corpus grows at most linear with the number of objects. The processing time for the specific object of interest is independent of it. The two scenarios for acquiring key-points of relevant objects interactively and actively are based on a first figure-background separation that uses a modern time-of-flight range imaging sensor to segment key-points. All these approaches have to our knowledge not been explored before.

There are future plans for "feature-sharing" between objects if they contain similar feature points. Another future activity will be to study the removal of the hand and gripper key-points in the vicinity of the object after range segmentation. Finally, the key-point descriptors will be changed in order to provide more information to the SVM and to reduce the percentage of support vectors needed.

## ACKNOWLEDGEMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ("The Cognitive Companion") and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

## REFERENCES

[1] R. Chatila, "The cognitive robot companion and the European beyond robotics initiative," 6th EAJ International Symposium *"Living with Robots,"* 2004

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.

[3] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R.C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 8, pp. 1443–1471, 2001.

[4] K.-B. Duan and S. Sathiya Keerthi, "Which is the best multiclass SVM method? An empirical study," *Multiple Classifier Systems*, pp. 278-285, 2005. http://guppy.mpe.nus.edu.sg/~mpessk/papers/nips03.pdf

[5] T. Oggier, et al., "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution SwissRangerTM," *Proceedings of the SPIE*, vol. 5249, no. 65, 2003.

[6] A. Pope and D. Lowe, "Learning appearance models for object recognition," In *International Workshop on Object Representation for Computer Vision*, Cambridge, England, 1996 http://citeseer.ist.psu.edu/pope96learning.html

[7] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 6, pp. 637-646, 1998. http://citeseer.ist.psu.edu/pontil98support.html

[8] D. Roobaert, "Pedagogical support vector learning: a pure learning approach to object recognition," Stockholm, 2001. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-3166 (2006-02-21)

[9] P. Fitzpatrick, "First contact: an active vision approach to segmentation," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vagas, Nevada, October 27 - 31, 2003.

[10] H. I. Christensen, "Cognitive (vision) systems," *ERCIM News*, pages 17-18, April 2003.

[11] G. Csurka, C. Bray, C. Dance and L. Fan, "Visual categorization with bags of keypoints," *ECCV International Workshop on Statistical Learning in Computer Vision*, Prague, 2004.

[12] J. Kubacki, B. Giesler and C. Parlitz, "Active autonomous object modelling for recognition and manipulation: Towards a unified object model and learning cycle," *Autonome Mobile Systeme 2005*, pp. 227-233, Springer-Verlag, 2005

[13] C. Schmid, R. Mohr and C. Bauckhage, "Comparing and evaluating interest points," *International Conference on Computer Vision*, Bombay, 1998. http://citeseer.ist.psu.edu/schmid98comparing.html

[14] T. Lindeberg "Scale-space theory: A framework for handling image structures at multiple scales," *Proc. CERN School of Computing*, Egmond aan Zee, Netherlands, 1996.

[15] S. Se, D. Lowe, and J. Little, "Local and global localization for mobile robots using visual landmarks," In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), pp. 414-420, Maui, Hawaii, 2001. http://citeseer.ist.psu.edu/se01local.html

[16] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-100)," Technical Report CUCS-006-96, Columbia University, 1996.

[17] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no 7, pp. 1443–1471, 2001.

[18] R. Unnthorsson, “Model selection in one-class ν-SVMs Using RBF Kernels.” *Proceedings of COMADEM 2003*, 16th Conference on Condition Monitoring and Diagnostic management, Vaxjo University, Sweden, 2003.

[19] C.-W. Hsu, C.-C. Chang, C.-J Lin, "A practical guide to support vector classification," Department of Computer Science & Information Engineering, National Taiwan University, Taiwan, 2003. www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[20] J. Kubacki & K. Pfeiffer, "Using range imaging sensors with color imaging sensors in cognitive robot companions: A new and simple calibration technique based on particle swarm optimization," In: Proc. *1st Range Imaging Day*, pp. 43-58, Zürich, 2005.

[21] J. Kubacki & Ulrich Reiser, "Mirrobot: fast detection of body-parts in the scale-space of range images with an application to human motion copying," to appear in the *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication,* RO-MAN 2006, United Kingdom, 2006.

[22] S. Knoop, S. Vacek, and R. Dillmann, "Modeling joint constraints for an articulated 3D human body model with artificial correspondences in ICP," *International Conference on Humanoid Robots* (Humanoids 2005), Tsukuba , Japan, 2005.

[23] M. Hans, "Eine modulare Kontrollarchitektur für den Hol- und Bringdienst von Roboterassistenten", Dissertation, Fraunhofer IPA, 2005.

[24] SIFT software available at: http://www.cs.ubc.ca/~lowe/keypoints/

[25] C.-C. Chang and C.-J. Lin, "LIBSVM -- A Library for Support Vector Machines.", 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[26] OpenCV Vision Library: http://sourceforge.net/projects/opencvlibrary/