Fuzzy Multisensor Fusion for Autonomous Proactive Robot Perception*

Martin Weser, Sascha Jockel, Jianwei Zhang
TAMS - Technical Aspects of Multimodal Systems
Department of Informatics, Hamburg University
Vogt-Kölln-Strasse 30, 22527 Hamburg, Germany
{weser,jockel,zhang}@informatik.uni-hamburg.de

***Abstract*— Robot perception still lacks reliability in complex natural environments. A commonly used method to improve perception is to incorporate more sensors with different modalities. This leads to increased computational requirements due to the parallel processing of huge amounts of sensor data. Appropriate sensor fusion methods are needed if contradictory information is provided by different sensors. We propose a feature-based technique to fuse multimodal sensor data using fuzzy rules. Probabilistic methods are avoided by applying fuzzyfication at the feature level. We propose a higher information gain of the available sensors by utilizing robot actions to focus sensors on objects of interest. Therefore sensor readings, algorithms and robot actions are combined into feature detectors. A goal-directed activation of these feature detectors renders parallel processing of all sensor data unnecessary.**

***Index Terms*— Fuzzy behavior selection, autonomous robot, multimodal perception, active perception**

I. Introduction

Similar to other wearable devices e.g. cell phones, PDAs, etc. robots will be ubiquitous in all kinds of environments in the near future. They will operate in natural environments that are meant to be comfortable for people. It is desirable for the robots to adapt to the environment, not *vice versa*. In contrast to other devices they will carry out physical tasks autonomously. Therefore perception of the environment is essential.

Nowadays robots are still under development, and their perception is not very robust. Thus robots are still not able to react properly to changes in the environment. Furthermore, the level of reliability required of robots is particularly high. A success rate of 99.9% is in general not enough. Since robots will be deployed in recurring tasks, even small probabilities of failure have to be eliminated in order to provide reliable operation over long periods of time.

Our goal is to make robot perception more robust and time-efficient. Their robustness will be improved by combining several sensor modalities and adopting actions to enhance perception. We propose an approach of fuzzy integration of data from different sensor modalities. This leads to appropriate processing of conflicting information and to computational efficiency.

A. Scenario: Service robot in natural environment

The TAMS research group at the University of Hamburg operates a mobile service robot equipped with several sensors and actuators. The robot is shown in Fig. 1. It can already perform some high-level tasks like grasping objects from a table, using light switches, replacing rubbish bins, communicating with people etc. All of these tasks require knowledge about the position of the involved objects. To extract this knowledge, the robot has several sensors, e.g. a laser range scanner, force sensors, active cameras mounted on a pan-tilt unit (PTU) as well as an omnidirectional camera. Combining all sensory data for a common representation of objects is a difficult task and still challenging to the robotic research community.

To enable the robot to perceive the environment exhaustively, learning mechanisms must be used due to the huge number of objects. To decrease this number, we consider only objects that are semantically relevant in the service robot domain, i.e. that can be important for the robot. Due to the still-poor manipulative capabilities of mobile robots only large objects will be considered for direct manipulation. Additionally, reference objects like a table, a corridor, a window, a corner etc. will be addressed by the perception process. By means of restricting perceptible objects we avoid learning mechanisms. The incorporation of learning mechanisms in our approach will be addressed in future work (see Sec. VII).

The scenario given above is utilized to clarify our proposed theory of how to perceive real-world objects using raw sensory data. The existence and identity of certain objects will be represented in fuzzy terms in contrast to commonly used probabilistic methods.

In special circumstances, the robot is not able to perform its given tasks due to insufficient perception caused by e.g. dim lighting conditions, occlusion, high distance to objects of interest etc. We will improve the robustness and performance of robot perception. The robot will perceive complex objects while utilizing actions for perception.

B. Outline

The remainder of the paper is structured as follows. Section II-A reviews important mobile robot platforms with the focus on multisensor fusion and action-oriented perception. In Sec. III, the representation of objects based on unimodal features

*This work is funded by the DFG German Research Foundation (grant #1247) – International Research Training Group CINACS (Cross-modal Interactions in Natural and Artificial Cognitive Systems).

978-1-4244-1819-0/08/$25.00©2008 IEEE

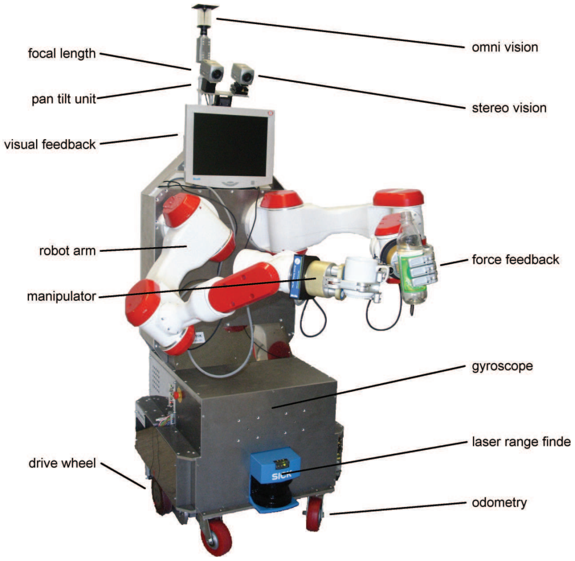


Fig. 1. Mobile service robot TASER with its various sensor and actor devices.

is given. The fuzzy sensor fusion approach proposed here is given in Sec. IV. We provide the general idea of how to utilize action for perception and how to choose appropriate behavior in Sec. V. Experiments and results are described in Sec. VI. Finally, we discuss drawbacks and opportunities of the proposed system in Sec. VII.

II. Multisensor Fusion in Mobile Robots

In mobile service robots a wide variety of sensors are used. Information provided by different sensor modalities may differ in the type of information, in accuracy, robustness etc. Sensors are usually used for certain tasks only. There are two reasons for this: Some tasks require highly specialized information in a certain domain and thus sensors are developed particularly with regard to these tasks. The sensors are almost exclusively used for these dedicated tasks. The other reason is that few labs possess several sensor systems and thus researchers develop algorithms that use sensor data from one domain only.

In recent years it has become widely accepted that crossmodal integration greatly increases the perceptual performance of both natural and robotic systems. Cognitive scientists tackle the question of how natural systems integrate different modalities. Researchers in the robotics society try to apply these findings in artificial systems and respectively develop new methods for cross-modal integration.

A. Categorization of sensor fusion methods

In general, multimodal integration is done for two reasons [1]. *Sensory combination* describes interactions between sensory signals that are not redundant. That means crossmodal integration leads to increased information compared to single modalities. By contrast, *sensory integration* describes interactions between redundant signals. This leads to enhanced robustness and reliability of the derived information. In the following, some major robot platforms are reviewed under the aspect of sensor fusion.

B. Examples of Multimodal Robot Platforms

A complete service robot platform was developed by the Centre for Autonomous Systems, Royal Institute of Technology, Sweden [2]. Applications that can be carried out by the robot involve exclusively one sensor modality, e.g. following a person using monocular vision, exploring the environment using sonar, docking into corners using a laser range scanner. The only task that uses a combination of multiple sensor modalities is self localization [3].

The University of Bielefeld developed the mobile service robot BIRON [4] that is equipped with several sensors. Multisensor fusion is performed for people tracking where laser range data and a color-based face detector in camera images are used [5]. This method belongs to sensory combination according to Ernst's categorization of sensor fusion [1]. Sensory integration is realized in robot communication tasks [6], [7]: Audio data are used for speech recognition and gesture detection for enhanced communication skills. These two types of data provide complementary data and thus lead to increased information.

The service robot TASER used within this work can also perform several tasks that make use of combining sensor modalities. For people tracking, both camera and laser range scanners are used [8]. Approaches for learning by demonstration use a highly developed image algorithm as well as speech commands [9]. At least two sensor modalities are involved in manipulation tasks: Force sensors at the tool to evaluate e.g. grasps and other sensor(s) to determine the target object's position before a grasp. In current applications image retrieval algorithms combined with proprioceptive data from the robot arm [10] are fused to determine positions of objects.

III. Perception based on Unimodal Features

Multimodal integration is often done on a low level, i.e. raw data vectors are concatenated to form higher-dimensional vectors if different sensor modalities share a common workspace. Data fusion on the information level is usually done if objects can be sensed simultaneously in different sensor spaces. Information about the object of interest are transformed to a common coordinate space, usally a robot-centered cartesian or a polar coordinate space, and fused subsequently.

To be flexible regarding the above-mentioned fusing paradigm we adopt a feature-based object representation. Objects are defined via sets of features $O_k \hat{=} \{f_i | i = 1 \dots N_k\}$ where O_k is an object and f_i are features with the following definition: Features are subsets of the sensory data stream that are capable of being differentiated. The degree of complexity of a feature is not specified at this point since our system is thus extendible to new algorithms that may provide features that are not detectable yet. Features are usually unimodal in perception since crossmodal integration is done by combining features from different modalities. A feature may be multimodal in terms of combining action and sensing. Force trajectories measured by the robot hand, for example, provide useful information only if the hand is moved.

Some detectors provide features that over-estimate occurrences of the requested object, others provide features underestimating them. Some gain from previous assumptions i.e. they confirm / dismiss object evidences derived from other features.

IV. Fuzzy Multimodal Integration of Features

Since all sensors are susceptible to errors, no artificial system can ever be certain of its internal knowledge of the world. Therefore, normally probabilistic representations are used to represent the environment. That leads to an enormous rise in computational complexity because probability values have to be assigned to all possible locations and configurations for each known object.

A. Hybrid Fuzzy / Analog Object Representation

To avoid the above-mentioned computational problems, we adopt a fuzzy representation of certainty about the existence and identity of objects and their defining features. An object / feature can be definitely existent, doubtful or impossible. No probability for the identity and existence of objects is used. The configuration, i.e. the exact geometric shape, position, size etc. is represented in continuous space as usual.

The distinction in identity and geometric configuration is also evident in findings in neuroscience. It has been shown that sensory information in the human brain is processed by taking the what and where dichotomy. The so-called *what path* processes the affiliation of sensations to real-world objects and the identity of abstract objects. The *where path* processes geometric locations and motion trajectories [11].

An evidence for a non-probabilistic pass-fail representation of objects is the phenomenon of bistable images provided by cognitive scientists. It is well known that images that allow ambiguous interpretations are perceived bistable by the human brain i.e. only one interpretation is perceived at a time [12]. The article investigates the influence of context on the decision of the brain, to select one of two interpretations. A simultaneous perception of different interpretations is unknown.

B. Fuzzification

We assume a system that is goal-driven, i.e. only objects of interest are perceived. It is impossible that features attract interest themselves, only features requested by objects of interest are detected. Thus there is no need to deduce objects bottom up from sets of detected features. In turn, feature detectors are only applied if required by an object of interest. The existence of a certain object implicitly defines the object's identity. Thus a representation of identity in addition to the existence is unnecessary.

Three fuzzy states are defined and applied to objects as well as perceived features:

- assured A feature (object) is definitely existent and its identity is certainly known. The state has no relation to knowledge about exact configuration which is continuous and may be probabilistic.
- doubtful Strong evidences for a feature (object) can be measured but uncertainty is still there.
- impossible The existence of a feature (object) is impossible. If one of an object's features belongs to this state the object usually also belongs to this state.

It is assumed that all sensor preprocessing modules provide membership values for each feature in each fuzzy set. In practice we apply the s-function

$$S(x,a,\varphi) = \begin{cases} 0 & x \leq a-\varphi \\ 2\left(\frac{x-a+\varphi}{2\varphi}\right)^2 & a-\varphi < y \leq a \\ 1-2\left(\frac{x-a+\varphi}{2\varphi}\right)^2 & a < x \leq a+\varphi \\ 1 & x \geq a+\varphi \end{cases} \qquad (1)$$

to a measurement value x that is proprietary to the underlying algorithm of the considered feature. The parameter φ describes the slope of the fuzzy sets boundary whereas the position of the boundary is defined by a. The result of the s-function provides membership values for the features to the fuzzy sets (assured, doubtful, impossible).

C. Inference and Composition Rules

The perceived unimodal features are combined into multimodal objects using fuzzy rules. Fuzzy inference rules are thus the underlying mechanism to integrate information from different sensor modalities. The state of object O_k^t at time t is defined by the set of fuzzy rules

$$\text{if } (f_1^t \wedge f_2^t \wedge \ldots) \text{ then } O_k^t \in s \qquad (2)$$

where s is one of the states (assured, doubtful, impossible) and f_i^t are memberships of the features to certain sets. Some of the rules are trivial, e.g. if all unimodal features belong to assured then O_k^t belongs to assured as well. Different qualities of feature detectors as described in Sec. III are reflected in the rules. The rules to combine unimodal features are defined manually to date. One of our future goals is to apply learning mechanisms to automatically derive inference rules.

Table 2 exemplarily shows a rule matrix to combine features for doors from two different sensor modalities. In practice, more complex rule matrices are used if more than two feature detectors come into operation.

V. Action-oriented Perception and behavior Selection

Usually perception happens in the service of action. This is reflected in the often quoted action-perception cycle. In turn, perception can be highly improved if actions are utilized as

Camera \ Laser	impossible	doubtful	assured
impossible	impossible	impossible	doubtful
doubtful	impossible	doubtful	assured
assured	impossible	assured	assured

Fig. 2. Unimodal features are integrated to multimodal perceived objects using membership to fuzzy sets. The underlying fuzzy rules can be visualized in a table as shown here.

described in [13]. Reasons for the need for action-oriented perception are

- resolution of sensors decreases quadraticly with distance between sensor and sensed object
- sensors may require action e.g. force sensors
- sensors are often mounted in a way that their work spaces do not overlap to provide a broad coverage of the environment

Furthermore, natural environments are cluttered and objects may be occluded. A physical approach as well as realignment of sensors relative to assumed objects of interest are promising actions to solve the above-mentioned difficulties.

However, robot actions are time and energy consuming and thus should be carried out only if necessary. Since these actions are in the service of perception, they are used only if the identity of an object of interest is not certain. Actions for better perception are selected according to the current situation, i.e. the robot task and the robot's internal knowledge about the world. We realise this by attaching actions or circumstances to features that have to be fulfilled during feature detection. Circumstances are e.g. relative positions of the robot to the evidence of an object of interest.

VI. Experiments and Results

The following two examples describe the general operations carried out while perceiving an object that is requested by the overlaying deliberative system. A short description of the used feature detectors shows that highly elaborated algorithms are not required if actions as well as multimodal information are combined. To evaluate our system, we recorded laser range scans as well as omni-directional images at 114 positions in the TAMS robotic laboratory at the University of Hamburg. The positions are chosen on a grid with 45cm edgelength and according to the circumstances of the environment. The positions of 5 tables and 3 doors have been annotated manually to provide ground truth. On closer examination, the reader will identify 10 tables in Fig. 5 but 5 of them cannot be detected by the robot due to occlusion and to cramped confines. The following subsections explain the unimodal feature detectors and results of the multimodal combination.

A. *Find table in natural environments*

The table experiment utilizes three features that each use a different sensor modality. A first estimation of table instances is given by a fast detector using laser scans. Two features that utilize robot action can be used to confirm the available target candidates as described later. The associated action-sensing behavior is only carried out if needed to reach the desired state `assured` for the object `table`.

1) Fast detection of table candidates in Laser range scans: Possible table legs appear in range scans as local distance minima and are thus easy to detect. These are grouped to form table evidences if relative positions fulfill two criteria: a) the distance between them is within a certain range, b) the angle between them is close to orthogonal. The tolerance in angle and size are the only parameters that influence the quality of results regarding detection rate and number of false alarms. Thus we evaluated the influence of different parameter sets on the result of the feature detector. Fig. 3 shows the number of correctly detected tables using different parameters.

In further experiments, 120mm tolerance in size and 0.1rad tolerance in angles are used. This results in 2 correct and 2.4 false tables detected per scan on average. The detection of 2 correct tables out of 5 possible ones results from occlusion and measurement errors. Fig. 4 shows the floor plan of the laboratory with one laser range scan and resulting table evidences.

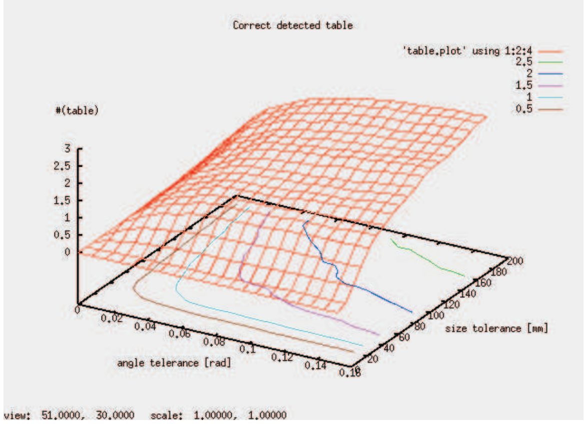


Fig. 3. Average number of correct tables detected by the feature detector using laser range scans with different parameter combinations.

2) Active Visual Exploration of Table Candidates: To confirm or dismiss a table candidate, the robot drives to a position relative to the table and focuses the camera mounted on a pan tilt unit on a corner of the table. Images captured from similar relative positions allow simple image processing algorithms to judge the correctness of the previous assumption. In the current implementation an appearance-based method is used. Fig. 5 shows pictures captured autonomously by the robot.

3) Touching the tabletop: The ability to sense the tabletop by touching is used to confirm existence of the table if it is still not certain. The sensor trajectory of the manipulator while moving the arm vertically downwards will be used to distinguish hard surfaces like a table from soft surfaces like a sofa, an easy chair, and from no surface at all. Since the execution of this feature takes a lot of time, it is only used if

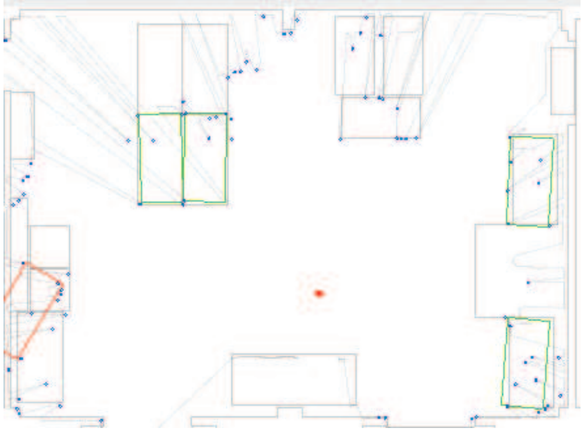


Fig. 4. Example of four correct (green) and one false (red) detected table. This is the result of the unimodal feature detector using laser range scans.

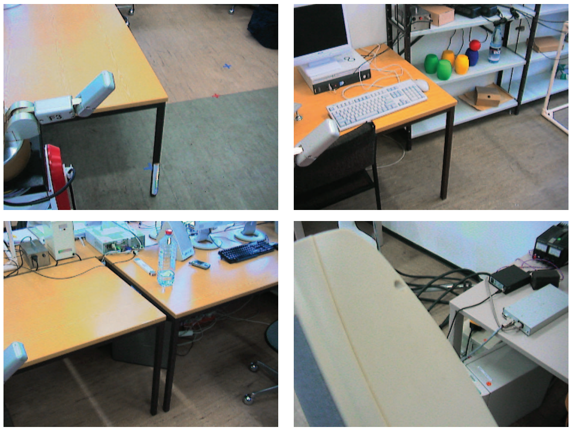


Fig. 5. The first three examples show confirmed table evidences, the last evidence has been dismissed.

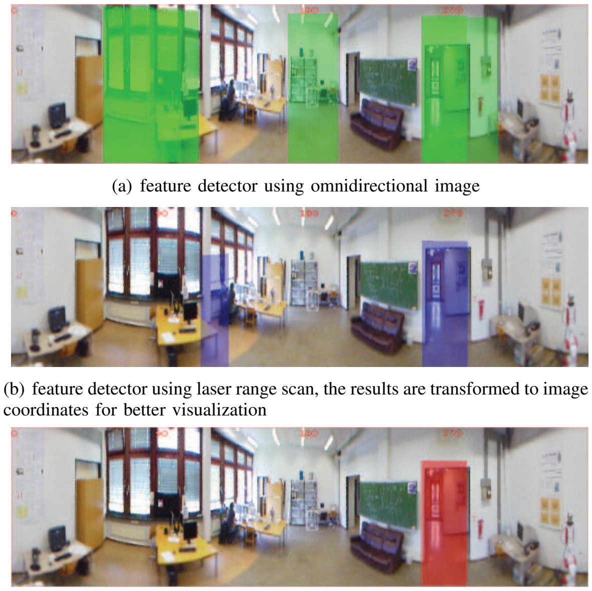
the first two did not lead to unique results i.e. `assured / impossible`.

4) Results: The complete perception procedure is hard to evaluate because each experiment takes 2 - 15 minutes, depending on the number of table evidences initially assumed by the laser scans. Resulting tables that satisfied all three features during our experiments where 100% correct. That means that no false positives occured at all. The number of tables that are not found depends on the choice of parameters (see Fig. 3).

B. Find doors and decide their current state

For the object `door`, features in three different modalities are used: two to detect doors and the third to decide whether the door is open or closed.

1) Door Detection: Simple features in laser range scans and omni-directional images are used to estimate door candidates. The detection method in laser range scans searches for line segments that show a sufficiently long gap that is not caused by occlusion. In omni-directional images, pairs of sufficiently long horizontal edges are considered as potential door candidates. As shown in Fig. 6 the separate results of each feature detector are not reliable. The overall result of a fuzzy multimodal combination, however, renders higher elaborated methods unnecessary. Fig. 6 shows door candidates from both features as well as the combined result using fuzzy rules. The combined multimodal result shows only doors belonging to `assured`. Doors attributive to this set are 91% correct compared to the manually defined real doors, which is good at first glance. However, if we consider real applications (e.g. using the doorknob) this is not reliable enough and will soon be further developed as described in Sec. VII.



(a) feature detector using omnidirectional image

(b) feature detector using laser range scan, the results are transformed to image coordinates for better visualization

(c) integrated multimodal result

Fig. 6. Estimated door candidates from single sensors are not reliable. Multimodal integration compensates the detection errors in most cases. The door to the left of the blackboard is partly occluded and cannot be detected.

2) Open or closed doors: To decide whether the door latch is closed or not, the robot can touch the vertical surface of the previously detected door. The measured force trajectory will be different if the latch is not closed and the door can be pushed open. This can be measured in turn by laser range scans. Although the latter is not implemented yet, our system is able to detect doors correctly and decide the state of the door.

VII. Conclusion and Future Work

The overall results show that the proposed system is capable of reliably perceiving complex objects. The fuzzy integration of object features from multiple sensor modalities provide a computationally efficient method that is easily extendible to more objects. The paradigm to perceive objects intentionally allows the utilization of robot action for perception. The reversal of the dependency of action on perception is one of the main contributions of this work. Action is used to improve perception rather than perception being used to afford action.

While working with expensive and damage-susceptible hardware in natural environments, reliability and safety is the most important issue. The results (91% correct door detection, 100% correct table detection) seem to be satisfactory but actually they are not. Robots will work autonomously and unsupervised over long periods of time, so a small chance of failure always exists. Thus we will improve our algorithms (unimodal features as well as fusion methods) and expand the test scenarios for better system performance.

Only few research groups possess mobile robot platforms equipped with various sensor modalities and manipulators. A direct comparison of results is almost impossible due to different hardware and software environments and different approaches to certain problems. An overview of the results of other research groups is given in Sec. II-A.

One major problem of the proposed system is the explicit definition of objects and detection skills. We only applied a set of simple features so far. An implementation of highly elaborated recognition algorithms especially in camera images will improve the overall performance significantly. Furthermore, the automatic generation of unimodal features as well as learning of feature subsets that describe an object will be considered. However, the small corpus of training data as well as the impossibility of unsupervised autonomous learning prevent the application of conventional machine learning algorithms. This turns learning of action-oriented robot perception into a particularly difficult problem. Nevertheless, this problem has to be addressed to enable the robot to act in environments that were not taken into account by the developer.

References

[1] M. O. Ernst and H. H. Blthoff. Merging the senses into a robust percept. *Trends Cognitive Science*, 8(4):162–169, April 2004.

[2] Magnus Andersson, Anders Orebck, Matthias Lindstrm, and Henrik I. Christensen. Isr: An intelligent service robot. In *Sensor Based Intelligent Robots*. Springer Berlin / Heidelberg, 1999.

[3] S. Rutgersson and B. Eriksson. Identification and sensor fusion to improve dead reckoning control of a mobile robot. In *Mechatronics*, Skövde, Sweden, September 1998.

[4] A. Haasch, S. Hohenner, S. Huewel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. A. Fink, J. Fritsch, B. Wrede, , and G. Sagerer. Biron: The bielefeld robot companion. In *Int. Workshop on Advances in Service Robots, Stuttgart, Germany*, 2004.

[5] M. Kleinehagenbrock, S. Lang, J. Fritsch, F. Lmker, G. A. Fink, and G. Sagerer. Person tracking with a mobile robot based on multi-modal anchoring. In *IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN)*, pages 423–429, 2002.

[6] J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer. Multi-modal anchoring for human-robot-interaction. *Robotics and Autonomous Systems, Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*, 43(2–3):133–147, 2003.

[7] Shuyin Li, Axel Haasch, Britta Wrede, Jannik Fritsch, and Gerhard Sagerer. Human-style interaction with a robot for cooperative learning of scene objects. In *Int. Conf. on Multimodal interfaces*, pages 151–158, New York, NY, USA, 2005. ACM Press.

[8] M. Weser, D. Westhoff, M. Hueser, and J. Zhang. Real-time fusion of multimodal tracking data and generalization of motion patterns for trajectory prediction. In *Int. Conf. on Information Acquisition*, 2006.

[9] Markus Hueser, Tim Baier-Lenstein, and Jianwei Zhang. Learning of demonstrated grasping skills by stereoscoopic tracking of human hand configuration. In *IEEE Int. Conf. on Robotics and Automation (ICRA), Orlando, Florida, USA*, 2006.

[10] Jianwei Zhang and Bernd Roessler. Self-valuing learning and generalization with application in visually guided grasping of complex objects. *Robotics and Autonomous Systems*, 47:117–127, 2004.

[11] Josef P. Rauschecker. Cortical processing of complex sounds. *Current opinion in neurobiology*, 8(4):516–21, 1998.

[12] Rashmi Sundareswara and Paul R. Schrater. A perceptual inference model for bistability. *Journal of Vision*, 7(9):803–803, 6 2007.

[13] M. Weser and J. Zhang. Proactive multimodal perception for feature based anchoring of complex objects. In *Int. Conf. on Robotics and Biomimetics*, 2007.