Project Acronym:	GRASP
Project Type:	IP
Project Title:	Emergence of Cognitive Grasping through Introspection, Emulation and Surprise
Contract Number:	215821
Starting Date:	01-03-2008
Ending Date:	28-02-2012

Deliverable Number:	D31
Deliverable Title :	Demonstration on a humanoid robot platform with cognitive grasping capabilities
Type (Internal, Restricted, Public):	PU
Authors	T. Asfour, M. Do, S. Ulbrich, M. Przybylski, D. Kragic, J. Bohg, A. Morales, B. Leon, V. Kyrki, D. Burschka, A. Argyros, H. Deubel, M. Vincze and R. Dillmann
Contributing Partners	All

Contractual Date of Delivery to the EC:	28-02-2012
Actual Date of Delivery to the EC:	08-02-2012

Contents

Chapter 1

Executive Summary

This deliverable describes the demonstrations realized on the humanoid platforms ARMAR-IIIa and ARMAR-IIIb showing the successful integration of components developed within the GRASP project. In order to enhance the cognitive grasping capabilities of a robotic system operating in human-centered environments, various modules improving the robot's sensory capabilities have been integrated. In addition, systems allowing the extraction and the application of grasping knowledge through human observation have been evaluated on the humanoid platform. The results of the demonstrations are presented in this deliverable in the form of publications and videos.

The deliverable consists of the following demonstrations

- SVM-based grasp stability Detection which is adapted and enhanced for usage on the humanoid robot ARMAR-IIIb, in order to allow the robot to predict whether a grasp is stable or not, based on sensor data. Tactile sensor data and joint angle data from ARMAR were used to train the Grasp Stability Detector, the trained detector was then tested on ARMAR-IIIb and enhancements to the original detector were developed and tested, such as filters for dealing with oscillation in the detector's output.
- Blind grasping using the haptic capabilities of the humanoid robot ARMAR-IIIb in order empty a basket filled with different objects. By exploiting the haptic capabilities of the robot, contacts with an object are detected and based on the contact positions correction movements are performed. This groping action is performed recursively until the detected contacts indicate a stable grasp.
- Vision-based grasping of known objects. A 3D point cloud an object is recognized and its pose is estimated by means of a 3D model matching. Based on the corresponding object model grasp hypotheses are generated and fitted into the scene. The best grasp hypothesis is executed by the humanoid robot ARMAR-IIIa.
- Task-based grasp adaptation in which an unknown table scene is explored and the objects within this scene are segmented and categorized. Based on the determined category and a given task previously trained systems are used to infer the most suitable, stable grasp which satisfies the task-specific constraints defined on the object category.
- Robotic grasping system combining the advantages of simulation-based grasp planning with knowledge from human grasping examples. Human grasping data were used to rate grasp hypotheses from the Medial Axis grasp planner in such a way that the robot could select among the planned grasps the most human-like grasp for actual execution.
- Grasp recognition system in which the robot imitates the grasp performed by the human in front of him. A human performs a grasp without any markers on their arms or hands. The robot observes the arm movement and the grasp type, maps them to his embodiment and performs the grasp on a similar object situated in front of him.
- Grasp imitation framework using a continuous grasp representation based on virtual springs. Fingertip trajectories of human grasp are observed by using the stereo camera system of a humanoid robot. From the captured motion data the parameters needed for the instantiation of the grasp

representation are estimated. The resulting grasp is adapted to the current object of interest and executed on the humanoid platform ARMAR-IIIb.

- A benchmarking software framework that addresses the problem that algorithms for grasping and dexterous manipulation cannot be evaluated and/or qualitatively be compared at all sites and under the same conditions. It, hence, offers an environment for testing and evaluating different grasping and dexterous manipulation algorithms in pure simulation. Further, it features a library of domestic everyday objects models and a real-life scenario including a humanoid robot in a virtual kitchen.

Chapter 2

Integration Activities

2.1 Stability detection

In this work, we show the integration of the grasp stability detection method (LUT) introduced in [BLJ+11] and [LKK10] on the robot ARMAR-IIIb. The method uses a support vector machine (SVM) to assess a grasp's stability based on the output of different sensors on ARMAR's hand. In order to train the SVM classifier for data from ARMAR's hand, we collected sensor data from several hundreds of sample grasps on a set of different household objects varying in size, shape, and weight (KIT). Sensor data collection was performed in the following way: A box containing the objects was placed in front of the robot. ARMAR's right hand was moved to a pre-pose near the object. We used varying pre-poses, including grasps from the top with vertical or tilted approach directions, grasps from the side as well as varying roll angles of the hand around its approach direction for all the cases described above. At the pre-pose, we started to record the tactile sensor data and the joint angle data. Then we moved the hand towards the object until the tactile sensors reported contact with the object and closed hand. After the pressure on the hand's actuators had stabilized, sensor data recording was stopped and we tried to lift the object, where we recorded if the grasp was successful or not. For the actual training of the SVM classifier, joint angle data were directly used, while in case of the tactile sensors, image moments were computed on the tactile sensor images which are related to the total pressure, the mean and the shape of the contact area. Further investigations were performed on the behavior of the classifier, where the effect of a mean filter and an exponential filter on oscillations in the classifier's output were studied. By applying these filters, the grasp stability detection can be run continuously during a grasp which is an improvement over [BLJ+11]. The continuous grasp stability detection enables faster decisions to be made on the stability of a grasp. Finally, two sorts of experiments were conducted. First, synthetic tests showed the classifier's behavior in terms of true positives, false positives, true negatives, false negatives. Second, validation tests were performed in a real world usage scenario, where the classifier was used to predict grasp stability on ARMAR-IIIb.

The results show that the integrated approach for grasp stability detection also works on complex robotic hands like the hand of ARMAR-IIIb.

Involved Partners: LUT. KIT

Leader and responsible: KIT

2.2 Haptic-based blind grasping

In the video **"BlindGrasping.wmv"**, we present a grasping method for grasping unknown objects with the help of the haptic sensors. The robot ARMAR-IIIb is equipped with tactile sensors in the fingertips and the palm of the hand and a 6 DoF force/torque sensor in the wrist. Furthermore, joint encoders and pressure sensors are mounted on the pneumatically actuated – hence compliant – hand, which can be used to detect forces exerted on the hand. The objects are located in a basket which is on a table in front of the robot. Basic image processing techniques are applied to locate the position of the basket and

to obtain an initial hypothesis on the position estimate of a single object. Once a hypothesis is found the following strategy is used for grasping the object:

The robot approaches the position estimated by computer vision from above until contact is detected which is indicated either by the tactile sensors, by forces measured with the 6 DoF force/torque sensor or the position encoders of the finger joints. If the contact position is located at one of the fingers, a correction movement is performed in the direction of the contact point. This is repeated until the hand is in a pose, which looks promising for a stable grasp. This is the case, when the only contact position detected lies in the palm of the hand and, hence, the fingers do not have contact with any obstacle. To grasp the object, the hand is closed and a grasp stability check is performed using the method described in section 2.1. In case of a stable grasp the robot tries to lift the object, while continuously checking the grasp stability. If the grasp appears to be unstable, the robot starts over.

After several trials, the robots accomplishes to grasp most of the objects. Due to the grasp stability detection the system knows if an object has been grasped successfully or if it has to try again.

Involved Partners: KIT

Leader and responsible: KIT

2.3 Vision-based grasping of known objects

In this demo, we present a robot emptying a box or cleaning a table with known objects. Initially, 3D point clouds of the objects are computed from stereo images using a stereo module (KIT).

Based on these point clouds, we apply an object recognition module which is presented in [CPB12] (TUM). The objects are recognized and their poses are estimated. The applied algorithm consists of two phases. The first one — the model preprocessing — is done offline. It is executed only once for each model and does not depend on the scenes in which the model instances have to be recognized. We assume that each object to be recognized is represented by a model consisting of a set of points with corresponding surface normals. The second phase is the online recognition and pose estimation which is executed on the range scan using the model representation computed in the offline phase.

The result of the recognition is a list of elements, where each element consists of both an ID and pose of the object. This information is used by the grasping module to generate a number of grasp hypotheses which are ranked regarding reachability, collision avoidance, and the resulting arm configuration provided by the inverse kinematics solver (distance to the joint angle boundaries). The most suitable grasp hypothesis is passed to the humanoid platform ARMAR-IIIa for execution. Using a Visual Servoing approach, the textured objects within the GRASP object set could be grasped successfully.

Involved Partners: TUM. KIT

Leader and responsible: KIT

2.4 Task-based grasp adaptation

In this work, we show how an object can be grasp under consideration of a given task represented by task-specific constraints defined on object categories. Given an unknown scene which contains several object instances of different categories, first, a scene exploration using a growing neural gas approach is performed to obtain interest points indicating possible objects locations. By applying a segmentation algorithm based on Markov Random Fields at these interest points, the objects are segmented and the corresponding point clouds are calculated. For the object categorization, two complementary categorization systems are used: 1) shape-based categorization using the segmented point clouds as input (TUW) and 2) appearance-based categorization using segmented images as input (KTH).

As described in [WV11], the shape-based categorization module exploits the similarity between the depth image given by the object's point cloud and a 3D model. In order to compare both, synthetic depth images of the 3D model are generated by sampling from different view points around the model. To determine the similarity between these depth images a combination of multiple shape descriptors is used consisting of following:

- D2 shape distribution descriptor on multiple resolutions which encodes a histogram of distances between randomly sampled points of the point cloud
- Moment invariants in 3D which inferred from the Hu moments invariants in 2D
- Voxel Based Spherical Harmonics descriptor containing a histogram which is calculated for 32 concentric spheres and frequency bands

The categorization result consists of two confidence measures whereas the first confidence measure is calculated offline and represents a bias calculated for each descriptor and each category. The second measure is computed online by comparing the multiple shape descriptors of the point cloud and the synthetic depth images of the 3D model.

The appearance-based categorization module as introduced in [MSK12] represents an object using a 2D shape descriptor in the form of a Histogram of Gradient and a SIFT feature map. For each descriptor a Support Vector Machine is trained with data originating from an object database containing various views of the object category prototypes. Given a view of a current object instance, the combined classification result provides the object category. The final category is determined through cue integration of the two categorization systems mentioned above.

Since an object category is represented by a set of object prototypes, in an offline step grasp candidates are trained on each prototype using the Medial axis planner in order to generate grasps for a category (KIT). To grasp a current object instance, once the object category is determined, an object category prototype matching the scene best is transformed and aligned within the scene to estimate the object's pose. The same transformation is applied to the corresponding grasp candidates.

In a further step, task constraints are inferred limiting the selection of grasp candidates which are most suitable to accomplish a certain task which can be e.g., pouring, hand over or tool use and is given by the user. To learn these task constraints, as introduced in [SHKK10], for a specific task a corresponding dataset consisting of feature vectors which include object features (size, convexity), action features (approach vector, grasp configuration), and manually defined constraints (freeo object volume, grasp stability) which are defined on the object as well as on the action features is generated. In order to encode the statistical dependencies between object, action and constraints features a Bayesian network is trained from which the constraints on the grasp action can be determined based on the given task and object information. Based the grasp constraints each grasp candidate is rated resulting in a ranked set of grasp candidates which is checked for reachability and collision avoidance (UJI). The best ranked, reachable grasp candidate is executed on the humanoid platform ARMAR-IIIa.

In order to guarantee a stable grasp execution a Visual Servoing approach using an appearance-based object tracking method based on textural SIFT features is applied. Before approaching the target object, the robot looks at the object and a set SIFT reference features is extracted from this initial view. During the approach phase, these features are tracked to obtain a current update of the object's position. For our experiments on the humanoid platform, we evaluated the resulting system on three object categories (car, bottle, and mugs) and four tasks (playing, pouring, hand over, and dishwashing). The object set consisted of at least two object instances of each object categories with different measurements. The results of this work are presented in the video **"TaskBasedGraspAdaptation.wmv"**.

Involved Partners: KTH. UJI, TUW, KIT

Leader and responsible: KIT

2.5 Human-inspired grasping

In this work, we demonstrate the human-inspired selection of grasp hypotheses for execution on a humanoid robot. The novelty of this work is the rating of the generated grasp hypotheses according to their human-likeness and the selection of a grasp for execution which resembles a human grasp the most. This is achieved by evaluating grasp data recorded by LMU.

In a first step, we generate grasp hypotheses for the test objects using the Medial Axis grasp planner, which exploits an object's local symmetry properties for grasp planning (see [PAD11]). Making use of the Medial Axis Transform shape descriptor, the grasp planner extracts grasp center points, hand approach directions and hand orientation vectors from the object geometry that have a high probability to result

in a successful grasp. The grasp hypotheses are tested for force-closure. In a second step, we use human grasping data for rating the grasp hypotheses generated by the grasp planner. Human grasping data were collected using a Polhemus Liberty electromagnetic motion tracking system, where the sensors were attached to the test subjects' fingernails in order to obtain fingertip trajectories of the grasping process. The rating procedure works as follows: For each of the grasps generated by the grasp planner, which are represented by the wrist pose and the hand joint angles, we calculate the grasp points on the object using the forward kinematics of the hand. For each grasp the sum of absolute differences between grasp point locations of the planned grasp and the mean of grasp locations of the observed human grasps serves as a measure of human-likeness. The grasps are now ranked with respect to this measure and the best rated grasp from the set of reachable grasps is chosen for execution on the robot.

The accompanying video **"Human-inspired-grasping.wmv"** shows the rating process and the execution of the selected grasp for two example objects. First the scene with the robot in front of the object is shown where the grasp hypotheses from the grasp planner are visualized. Then the results of the rating process are shown, where the individual ratings for all grasps as well as the average fingertip end positions from the human grasps are displayed. Finally, the best ranked grasp hypotheses which is actually reachable is displayed and executed by the robot.

This work resulted in a joint paper of KIT and LMU published at Humanoids 2011 (see [PAD11]).

Involved Partners: LMU, KIT

Leader and responsible: KIT

2.6 Grasp recognition and mapping

The video **GraspRecognition.wmv"** shows the work presented in [DRK+09] where human teaches a robot what grasp type is to be applied to grasp a particular object. For this purpose the robot observes how the human performs the action, paying special attention to how the arm is moved, which grasp type is performed and which object is grasped. The arm movement is observed with the wideangle stereo pair in ARMAR, and processed with an upper body tracker (see [AAD08]). Once the human hand is at the target object, the grasp type is observed with one foveal camera in ARMAR, compared with a database of hand poses and classified as a particular grasp type (see [RKK10]). For this demo, we restricted the grasp types to be just one of the following three: power grasp from top, power grasp from the side, and pinch grasp. Finally, the object recognition is based on comparison with different views generated from an object database (see [AAD09]). All this information is used to generate an approach movement with the Master Motor Map interface, and apply a grasp (defined as the correspondent grasp to the one performed by the human) to the recognized object. The system runs in real time, and the human does not need to wear any markers or special devices.

Involved Partners: KTH. KIT

Leader and responsible: KIT

2.7 Grasp imitation using a continuous grasp representation

In this work, we evaluated our grasp imitation framework in order to acquire novel grasping skills from human observation. Based on captured motion data of a human grasp, our goal is to generate a generalized representation allowing the synthesis of the demonstrated grasping procedure on a humanoid platform in a continuous manner which involves all three phases of grasping: preshape, approach, and enclose. To represent a grasp in a continuous way allowing object specific adaptation, we exploited a grasp representation in task-space based on virtual springs between the fingertips of the grasping hand (see [DAD11b]). The instantiation of the grasp representation is accomplished by parameterizing the spring constants of the virtual springs. By means of a central force towards a target configuration consisting of the supposed contact points, the resulting dynamical system is modulated resulting in a continuous grasping movement of the fingertips. For the estimation of the spring constants, an estimation procedure combining global and local optimization algorithms has been implemented which allows the estimation of parameters of a dynamical system from noisy data. To accomplish the mapping between the hand of the observed human subject and the robot hand, the estimation is performed using the grasp representation which has been

adapted to the measurements of the human hand. The adaption is done by equating the virtual spring lengths with the distances between the fingertips in the initial pose. For the mapping on our robot, a grasp representation adapted to the robot hand is instantiated using the spring constants which has been estimated from the human demonstration. The instantiated grasp representation encodes a specific grasp type for a prototype of an object category such as a cylinder or a box. To grasp a specific object instance the configuration of contact points determined for the prototype objects has to be scaled to match the current object instance. To take into account a task-specific adaptation of the grasp, Dynamic Movement Primitives have been used enabling the robot to approach an object from different start poses towards variable object target poses. The execution of a grasp is performed by moving the robot's hand towards a designated target pose whereas the acceleration during the approach movement is used to modulate the grasp representation.

In order to generate motion data of human grasp demonstrations, methods for capturing fingertip movements have been implemented (see [DAD11a]). To design the human-machine interface for the user as intuitive as possible and, therefore, to realize a grasp imitation process in an online fashion, the tracking of the human fingertips is performed in a markerless manner using the stereo camera system of the humanoid. In order to track circular image features exposed by the fingertips, a method combining particle filter and mean shift algorithms has been implemented allowing robust fingertip tracking at approximately 25 fps.

As it can be seen in the video **"VirtualSpringRepresentation.wmv"**, the resulting framework has been successfully evaluated on four different grasp types including a power, tripod, pinch, and a lateral grasp.

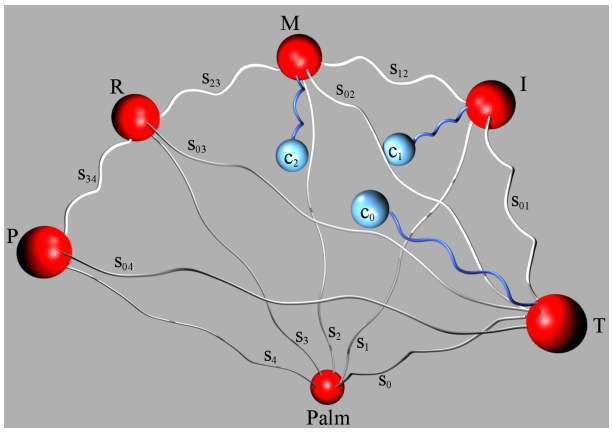

Figure 2.1: Grasp representation for a tripod grasp with three contact points.

Figure 2.2: Left: Hough space visualization. Center: Contour for the tracking. Right: Result image.

Involved Partners: KIT

Leader and responsible: KIT

2.8 Benchmarking environment

In this work, we present the new benchmarking suite – a software framework that has been developed at KIT as a part of the *OpenGRASP toolkit* which has been introduced in [UKA+11]. This software addresses the problem that algorithms for grasping and dexterous manipulation cannot be evaluated and/or qualitatively be compared at all sites and under the same conditions if the laboratories have different (or any) humanoid robots, manipulable and graspable objects, and environments.

This *OpenGRASP Benchmark Environment* offers an environment for testing and evaluating different grasping and dexterous manipulation algorithms in pure simulation, which aims at clear and reproducible qualitative evaluation in any laboratory. It offers all necessary mechanisms for the implementation of performance metrics and benchmark routines for the different aspects of the topic, which have the chance

to be accepted by a wide community. By providing a consistent environment, individual benchmarks can be combined in order to evaluate complete high-level tasks. Therefore, the software framework has a very extendable structure in order to be able to include a wider range of benchmarks defined by the community of robotics researchers. The benchmark suite comes with an expendable list of individual benchmarks that serve as a guidance for the community-driven development.

In addition to the software development framework, the benchmark environment features a great library of domestic everyday objects models that integrate well into the integrated real-life scenario featuring a fully employable model of the humanoid ARMAR-III acting in a virtual kitchen (see Fig. 2.4).

An introduction to the OpenGRASP benchmarking environment is given in the video **"BenchmarkingEnvironment.wmv"**. This video demonstration provides an introduction to the presented software framework focusing on the usage of the interface and the already implemented benchmarks. The benchmarks shown were created using the framework and give an impression on how new algorithms for grasping and dexterous manipulation are integrated. They cover the topics of collision free motion planning with *Rapidly-exploring Random Trees (RRT)* (see Fig. 2.3) and grasp planning with the *Medial Axes Planner*.

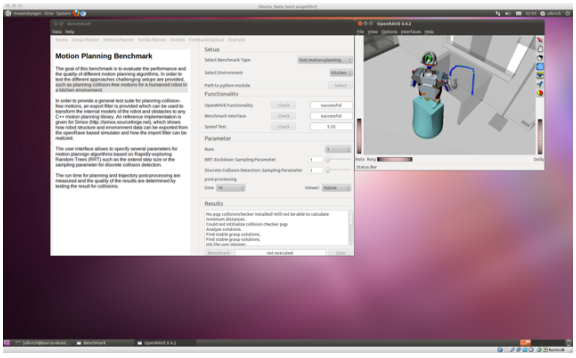

Figure 2.3: Screenshot of the graphical user interface of the motion planing benchmark.

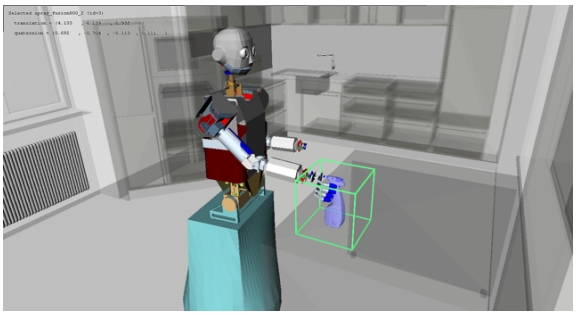

Figure 2.4: The scenario featuring a detailed model of the humanoid robot ARMAR-III in a virtual kitchen environment.

Involved Partners: KIT

Leader and responsible: KIT

References

[AAD08] P. Azad, T. Asfour, and R. Dillmann. Robust Real-time Stereo-based Markerless Human Motion Capture. In *IEEE/RAS International Conference on Humanoid Robots*, pages 700–707, Daejeon, Korea, December 2008.

[AAD09] P. Azad, T. Asfour, and R. Dillmann. Accurate Shape-based 6-DoF Pose Estimation of Single-colored Objects. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2690–2695, St. Louis, USA, October 2009.

[BLJ+11] Y. Bekiroglu, J. Laaksonen, J.A. Jorgensen, V. Kyrki, and D. Kragic. Assessing Grasp Stability Based on Learning and Haptic Data. *IEEE Transactions on Robotics*, 27(3):616–629, 2011.

[CPB12] Sven Parusel Kai Krieger Chavdar Papazov, Sami Haddadin and Darius Burschka. Rigid 3D Geometry Matching for Grasping of Known Objects in Cluttered Scenes. *International Journal of Robotics Research*, 2012. to appear.

[DAD11a] M. Do, T. Asfour, and R. Dillmann. Particle Filter-Based Fingertip Tracking with Circular Hough Transform Features. In *Accepted at IAPR Machine Vision Applications (MVA'11)*, Nara, Japan, June 2011.

[DAD11b] M. Do, T. Asfour, and R. Dillmann. Towards a Unifying Grasp Representation for Imitation Learning on Humanoid Robots. In *IEEE International Conference on Robotics and Automation*, pages 482–488, Shanghai, China, May 2011.

[DRK+09] M. Do, J. Romero, H. Kjellström, P. Azad, T. Asfour, D. Kragic, and R. Dillmann. Grasp Recognition and Mapping on Humanoid Robots. In *IEEE/RAS International Conference on Humanoid Robots*, Paris, France, December 2009.

[LKK10] J. Laaksonen, V. Kyrki, and D. Kragic. Evaluation of feature representation and machine learning methods in grasp stability learning. In *10th IEEE-RAS International Conference on Humanoid Robots*, pages 112–117, 2010.

[MSK12] M. Madry, D. Song, , and D. Kragic. From Object Categories to Grasp Transfer Using Probabilistic Reasoning. In *IEEE International Conference on Robotics and Automation*, Saint Paul, USA, May 2012.

[PAD11] M. Przybylski, T. Asfour, and R. Dillmann. Planning grasps for robotic hands using a novel object representation based on the medial axis transform. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.

[RKK10] J. Romero, H. Kjellström, and D. Kragic. Hands in action: Realtime 3D reconstruction of hands in interaction with objects. In *IEEE International Conference on Robotics and Automation*, pages 458–463, Anchorage, USA, May 2010.

[SHKK10] D. Song, K. Huebner, V. Kyrki, and D. Kragic. Learning Task Constraints for Robot Grasping using Graphical Models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, October 2010.

[UKA+11] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann. The opengrasp benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, USA, September 2011.

[WV11] W. Wohlkinger and M. Vincze. Shape-Based Depth Image to 3D odel Matching and Classification with Inter-View Similarity. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, USA, September 2011.