

# General 3D Modelling of Novel Objects from a Single View

Zoltan-Csaba Marton, Dejan Pangercic, Nico Blodow, Jonathan Kleinhellefort, Michael Beetz  
Intelligent Autonomous Systems, Technische Universität München  
{marton, pangercic, blodow, kleinehe, beetz}@cs.tum.edu

**Abstract**—In this paper we present a method for building complete models for grasping from a single 3D snapshot of a scene composed of objects of daily use in human living environments. We employ fast shape estimation, probabilistic model fitting and verification methods capable of dealing with different kinds of symmetries, and combine these with a triangular mesh of the parts that have no other representation to model previously unseen objects of arbitrary shape. Our approach is enhanced by the information given by the geometric clues about different parts of objects which serve as prior information for the selection of the appropriate reconstruction method.

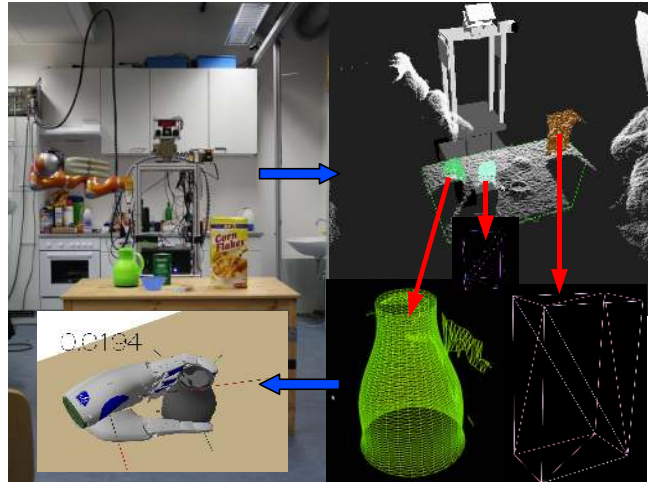
While we designed our system for grasping based on single view 3D data, its generality allows us to also use the combination of multiple views. We present two application scenarios that require complete geometric models: grasp planning and locating objects in camera images.

## I. INTRODUCTION

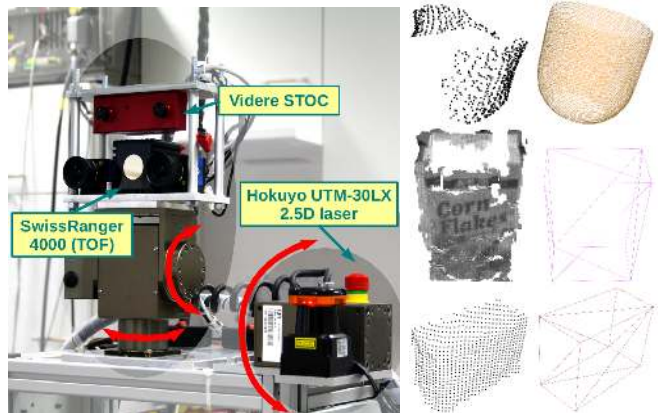
Autonomous robots performing tasks in human living environments have to manipulate many different objects, and in the course of their long term operation they will inevitably encounter novel objects. In this paper we consider the problem of estimating 3D models of objects for grasp planning from a single view of a 3D sensor. The biggest problem in this perception task is that the robot sees only the front part of the object and must make assumptions about the possible back side based on this data to avoid incorrect grasps. We tackle this problem by exploiting some of the structure that the environment and the task at hand give us, namely that many objects designed for daily use in human living environments have common symmetries, and that the objects are typically physically stable, which means that they are standing on planar surfaces. The reconstruction algorithms that we are investigating can use this structural information in order to obtain better reconstruction results (see Figure 1).

Our objective is to investigate this problem in the setting of a kitchen using small, compact and low cost devices, like the time-of-flight (TOF) cameras, stereo cameras and small laser scanners which have rather low resolution and accuracy. The reason for doing so lies in their flexibility and safety concerns which is a major drawback of e.g. SICK LMS 400 laser sensor which is highly accurate but not eye-safe. The pre-processing and feature estimation steps presented herein thus enable the fitting to work not only with the highly accurate 3D sensors, but also with the less precise Hokuyo laser, and to some extent with the TOF and stereo cameras.

We present the methods we employ, and we evaluate them on scans of various objects coming from different sensors. We rely on the detection of planar and arbitrary rotational



**Fig. 1:** Novel object modeling in a household environment using 3D sensors. Our Kuka-based mobile manipulation platform (top-left) acquires one-shot point clouds of typical kitchen objects (top-right) and computes respective mesh models for them (bottom-right). Models are provided to the grasp planner for grasp computation and evaluation (bottom-left).



**Fig. 2:** Left: some of the sensors our robot is equipped with. Middle: raw data from SICK LMS400, Videre STOC and SwissRanger 4000, respectively (from top to bottom). Right: the reconstructed model of objects scanned using different sensors.

symmetries to complete the data from a single scan with the back side, and use the full object models for computing valid grasps even if the back of an object has not been visible during the acquisition phase (see Figure 2).

We have already shown before that exploiting rotational [1] and planar symmetries [2] is especially useful for modeling of many types of objects which are common in the indoor environments, and that hybrid models containing shape primitives and triangulated surfaces are flexible enough to represent arbitrary objects [3]. The prerequisite for these

works were, however, highly accurate 3D measurements, which as argued earlier are not always available.

The main contributions of this paper are:

- a framework for single view grasping that is general enough to deal with noisy sensing devices,
- robust reconstruction method that approximates the underlying surface, its normals and the physical radius,
- improved model fitting and verification methods that take the corrected surface parameters into account and detect probable symmetries.

## II. RELATED WORK

Multiple sensors were used for solving similar tasks, like cameras [4], [5], stereo cameras [6], [7], 3D sensors [8], and also their combinations to speed up or improve results.

A vision-based grasping system which segments objects on a table and constructs triangular meshes for them is presented in [9]. While the presented method is general and works for many objects, it creates complicated models for certain objects, which could be simplified through the usage of geometric primitives. A simplification of the modeling problem is used in Grasp-It [10], where geometric shape primitives are used to model each object as a combination of spheres, cylinders, cones or boxes.

In purely computer vision based approaches, image features [11] are used to find matches between parts of a scene and a database of object images. These approaches rely on a database of segmented images and since no knowledge about the 3D information is known, the system can easily make mistakes and return false positives (e.g. when an object contains an image of another). Some of the solutions adopted consist in creating complete 3D models offline for the targeted objects and finding feature spaces to match the partial views with models in the database [12]. Another approach to obtain 3D information directly from camera images is to project CAD models from a database to the image and search for good matches in the edges domain, as in [4] for example. While this is a more direct method, it is still dependent on a database of CAD models.

Available models of complex objects are decomposed into superquadric parts in [13] and [14], and these models are matched to a point cloud. This however needs a database of models, and moreover, their decomposition into superquadric components, which is often difficult to obtain. A sample consensus based approach for model decomposition is presented in [15], where a set of 3D geometric primitives (planes, spheres, cylinders, cones and tori) are fit to noisy point clouds. Since the point clouds presented there are complete, the authors do not need to reconstruct the missing parts.

In [16] the authors describe a method for detecting and verifying symmetries in point clouds obtained from a single viewpoint, and they project the existing points according to the detected symmetry to obtain the back side. However, using our methods we were able to reconstruct surfaces by approximating them with shape equations and thus generate a complete model, which can be meshed with the required density or sparseness according to the speed and accuracy requirements of our applications.

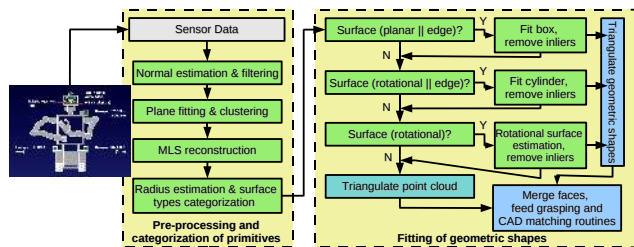


Fig. 3: Automatic geometric model generation pipeline

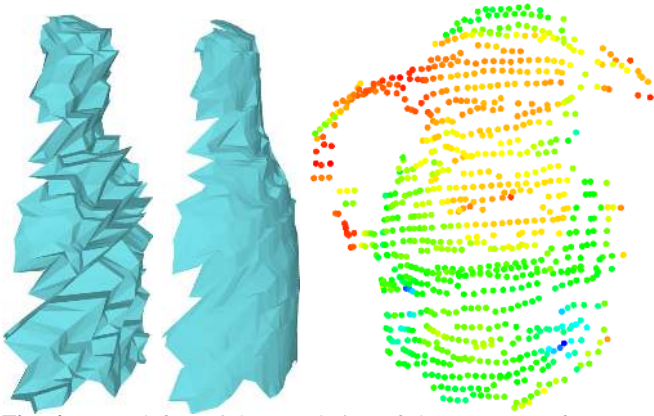
## III. SYSTEM OVERVIEW

Figure 3 depicts the necessary steps for the generation of the models from the raw point cloud data acquired from four different sensing modalities on our robot: two different laser scanners (Sick LMS400 and Hokuyo UTM 30LX), TOF camera (SwissRanger 4000) and stereo on chip camera (Videre STOC). The pipeline is implemented in ROS (Robot Operating System - [www.ros.org](http://www.ros.org)), as a collection of modules – “nodes”, depicted as green or blue blocks in Figure 3. The data flow between nodes can either occur through shared memory, when optimal performance is required, or over the network, which is convenient for decentralized processing.

As described in our previous works, e.g. [17], we proceed by first removing the erroneous measurements based on jump-edge detection, and estimating the normals for each surface voxel. In the next step we restrict the search space by detecting horizontal regions of interest (ROI) (e.g. tables, counter tops, shelves) using an efficient variant of the RANSAC algorithm [18]. We expect these places to support object candidates (point clusters) for modeling. Point clusters are extracted by testing whether the projection of the points fall inside the bounds of the previously extracted ROI (we rely on a physical separation between the objects for segmentation).

These point clusters representing object hypotheses are processed as follows: first, the points in sparse regions are removed based on statistical analysis, then they are smoothed, and at each point the surface normal and the minimal curve radius are estimated. The data is then passed into a hierarchical loop which consists of the following steps: i) fit boxes and cylinders first to detect the simplest symmetries (if any), ii) estimate whether the remaining data contains a surface of revolution and fit a contour, and iii) triangulate the remaining free form surfaces.

We use a Sample Consensus (SaC) approach to fit geometric models (boxes, cylinders and rotational shapes) because they scale well with the different amounts of data coming from various sensors without adjustment. The accuracy and robustness of our methods are increased further by the previously mentioned surface estimation step, as well as by taking into account the accurate normal estimates provided by it, and the point-level surface radius estimation results. After each fit we validate the model by computing the plausibility of the supposed shape to produce the obtained measurements. For this we maintain and update a voxelized labeling of our environment that keeps track of occupied, free and occluded space as described in [1]. For objects and



**Fig. 4:** From left to right: Rendering of the raw scan of a teapot obtained from the Hokuyo UTM-30LX and the points projected onto the estimated underlying surface. Radius estimation results for a scan of a teapot – values are color-coded from red (radius of 0) to blue (radius of  $0.1m$  or higher). Please note that for visualization purposes the values were capped at  $0.1m$ , as planar regions have an infinite or very large radius.

parts that do not present the symmetries we are looking for, triangulation is the only alternative as no information about possible objects is known to the system.

In the final step we triangulate all fitted models, merge the triangulation results, and publish the model for applications like force-closure grasping and CAD model-based object detection in 2D images as discussed in Section VI.

As the TOF and stereo cameras proved to be too inaccurate to reconstruct objects that are more complicated than boxes reliably (as in Figure 2), and the SICK laser is very accurate, we focused solely on the processing of the data coming from the Hokuyo laser sensor.

#### IV. 3D MODELLING PIPELINE

In this section we will detail the different processing steps, then in the next section we will present the experimental results and application scenarios, and conclude in Section VII, giving insights on our future work.

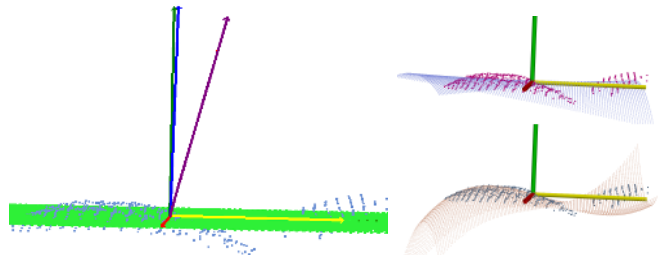
##### A. Data Pre-processing

Since the data coming from the sensors is noisy, we are using a MLS algorithm described in our previous work [19] to correct it (see Figure 4). At each point, after a tangent plane has been robustly identified, we approximate the point’s neighborhood using a height function  $f$  relative to this plane, in the form of a  $3^{rd}$  order bi-variate polynomial defined in a local coordinate system:

$$f_{(u,v)} = c_0 + c_1u + c_2v + c_3uv + c_4u^2 + c_5v^2 + c_6u^2v + c_7uv^2 + c_8u^3 + c_9v^3$$

where  $u$  and  $v$  are coordinates in the local coordinate system lying on the tangent plane. To obtain these ten unknown coefficients  $c_i$ , we perform a direct weighted least squares minimization, and project the point onto the obtained surface.

By choosing the query point to be at the origin of the local coordinate system ( $\vec{U} \perp \vec{V} \perp \vec{N}$ , with  $\vec{U}$  and  $\vec{V}$  in the plane, and  $\vec{N}$  parallel to its normal), we can easily compute the normal  $\vec{n}$  of the estimated surface. We compute the two



**Fig. 5:** Left: surface normal estimated by PCA (green)  $2^{nd}$  order (blue) and  $3^{rd}$  order (purple) polynomials. Right top and bottom: the polynomials fitted to the neighborhood. Please note that for better visibility high quality data was used for this example.

partial derivatives at  $(0, 0)$  which are  $c_1$  and  $c_2$  (identifying the two tangents) and their cross product:

$$\vec{n} = (\vec{U} + c_1\vec{N}) \times (\vec{V} + c_2\vec{N}) \quad (1)$$

Since the minimum radius estimator and the reconstruction modules work robustly with varying densities and their accuracy is influenced only by the correctness of the estimated normals, using this approach significantly improves their success rate even without re-sampling (see Figure 5).

##### B. Estimating The Radius of The Underlying Surface

While estimated surface normals are important, they alone tell little about the surface’s type. The estimation of the curvature (using the ratio of eigenvalues of a neighborhood’s covariance matrix) on the other hand neglects surface normals and we found it to be too inaccurate. To differentiate surface types we used variants of Point Feature Histograms [20] in a classification framework. It is however a relatively time consuming process even with the latest (less descriptive) Fast Point Feature Histograms [21], and local surface type classification is not always needed. What we are actually interested in is the approximated radius of the fitting curves to a local neighborhood. This value with *physical meaning* can be tied directly to the underlying surface without the need for classification, and we consider this very important.

We compute this feature (Radius-based Surface Descriptor, or RSD) starting similarly as in the case of spin images with a local support [22], with the added advantage that we consider the normals of the neighboring points directly, and that we can extract values that intuitively describe the local surface.

If we look at the case of the sphere, for each point all the circles that fit to its neighborhood have the same radius, namely the radius  $r$  of the sphere itself. For each of these circles we can write the following relation between the distance  $d$  of a point on the sphere from the original point and the angle  $\alpha$  between these two points’ (undirected) normals:

$$d_{(\alpha)} = \sqrt{2}r\sqrt{1 - \cos(\alpha)} \quad (2)$$

From this we can see that given the distance and the angle between two point normals one can approximate the radius of the circle the points are on.

In the case of an ideal plane, this estimated radius will always be infinite with all neighbors, since they have parallel normals. A point on a cylinder is on multiple circles (ellipses actually), and the radius estimated with different neighbors





**Fig. 6:** Radius-based Surface Descriptors (RSD): 2D histograms of normal angle difference and distance from a reference point to its neighbors (examples are shown from left to right for histograms computed for points on a plane, sphere, corner, edge and three cylinders, where the last two are computed for different points of the corrected teapot scan presented in Figure 4) – from these histograms the physical surface radii can be read as the slopes of the different lines going through the lower left corners.

will vary between the minimum radius (the radius of the cylinder) and the maximum radius (infinity). For corners and edges the estimated radius changes similarly as for spheres and cylinders (see Figure 6).

Given a point on a surface along with its neighbors, this minimum and maximum radius can be estimated using the model in Equation 2 by solving the equation system for  $r$  given the maximum and minimum angles for different distance intervals. While estimating both of them would be required for differentiating between spheres and cylinders for example, for our application the minimal radius (defined by the maximal angle) is enough. To make the estimation easier, we can exploit the fact that  $\alpha \in [0, \pi/2]$  and the Taylor decomposition of Equation 2 is simple:

$$d_{(\alpha)} = r\alpha + r\alpha^3/24 + O(\alpha^5) \quad (3)$$

where  $O(\alpha^5)$  indicates the existence of elements with order 5 and upwards. Thus we can assume  $d = r\alpha$  which renders the problem of finding  $r$  to a simple regression.

This feature is easy to compute, while still being very descriptive<sup>1</sup> and does not require consistently oriented normals. Because it is a continuous value that estimates the real minimal metric radius of the curve each points lies on, it can be used for example as a prior when sampling points to fit different surfaces.

In our case, for fitting boxes points with very high estimated minimal radius are preferred, while for rotational models they are avoided along with points having a very small radius. We incorporated this preference in the random sampling step for each model fitting algorithm. The inliers are also weighted by how well their radius fits the model.

### C. Box and Cylinder Fitting

The method used for deciding on when to use box or cylinder fitting is based on our previous work on footprint analysis [2], but now we also rely on the 3D normals to robustly detect a circular or a rectangular footprint, and also during the 3D model fitting and validation step. This enables our method to make the best choice regarding what model it should choose, a robustly estimated oriented bounding box, a cylinder, an arbitrary rotational object, or fall back to the triangulation if none of the models fit the points well enough.

If enough points have minimum radius that was capped, meaning that they most probably lie on planar surfaces, we

<sup>1</sup> using both the minimal and maximal angle along with gradients it can produce different signatures for spheres, cylinders, planes, cones, corners, edges, and surfaces of high normal variance (noise, handles, stems, etc.)

set out to find the best fitting box to the cluster. Unlike in [2], we fit a rectangular model directly to the points having normals perpendicular to the up axis (as we assume boxes to be standing on one of their sides), using the preferential sampling described before. For each sample, we accept those points as verifying the model that have parallel or perpendicular normals (weighted by the size of their minimal radius to lower the effect of false positives), and maximize the number of inliers in a RANSAC loop. This direct approach clearly outperforms the line detection and merging and/or box detection based on PCA analysis. Since we assume a correct segmentation and noise removal is performed, after the box’s orientation is found, its sizes are set to encompass the cluster, and the final number of inliers is checked against the model.

In order to detect cylinders, we also use a SaC approach which is based on the observation that on a cylinder surface, all normals are orthogonal to the cylinder axis, and intersect it. We consider the two lines defined by two sample points and their corresponding normals as two skew lines, and the shortest connecting line segment as the axis. Determining the radius is then a matter of computing the distance of one of the sample points to the axis. By enforcing an upright position results are more robust, but it is not mandatory.

### D. Estimation of Rotational Surfaces

To reconstruct almost arbitrary surfaces of revolution, we employ a RANSAC-based two-step approach as described in [1]. In the first step, a rotation axis is estimated from sample points by minimizing a function over the line-to-line distances between the axis and the lines defined by each sample point and its corresponding normal. This is based on the observation that for a rotational object, a line constructed from a point and corresponding normal intersects the symmetry axis. The contour line is then estimated in the second step (see below).

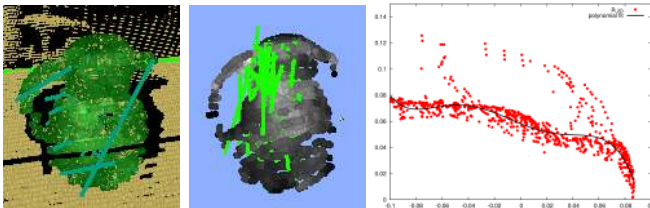
Let  $\langle \mathbf{a}, \vec{\mathbf{a}} \rangle$  denote the axis, defined by a point  $\mathbf{a}$  and a direction vector  $\vec{\mathbf{a}}$  and let  $\langle \mathbf{p}_i, \vec{\mathbf{n}}_i \rangle$  denote the line defined by the  $i$ th sample point and its corresponding normal vector. Then, we minimize the following function over  $\mathbf{a}$  and  $\vec{\mathbf{a}}$ :

$$\sum_{i=0}^m d_{l,l}(\langle \mathbf{a}, \vec{\mathbf{a}} \rangle, \langle \mathbf{p}_i, \vec{\mathbf{n}}_i \rangle)^2, \quad (4)$$

where  $d_{l,l}$  stands for the line-to-line distance. This can be solved using a non-linear optimizer like Levenberg-Marquardt.

Once an axis has been found, the original sample points are transformed into a 2D coordinate system such that the rotation axis coincides with the  $X$  axis. Every point  $\mathbf{p}_i$  is projected onto a point  $\mathbf{p}_{i,2D}$  whose  $x$  coordinate is defined as its position along the rotation axis, and whose  $y$  coordinate represents the point-to-line distance between  $\mathbf{p}_i$  and  $\langle \mathbf{a}, \vec{\mathbf{a}} \rangle$ .

We then employ a polynomial fitting step based on least-squares minimization to fit a preliminary contour line to the projected (2D) sample points. This contour line is then used to determine which points are inliers to the rotational model, and the polynomial is refitted using these inliers. This can



**Fig. 7:** The axis estimation step fails consecutively if surface points and normals contain too much noise (left). Re-sampling the points results in usable axis estimates (middle). Fitting a polynomial function to generate the contour curve given the axis (right).

be repeated until changing the polynomial coefficients does not increase the number of inliers. An example of this 2D problem is shown in Figure 7-right for the teapot data-set used earlier. Note that the contour curve is not influenced by the outliers that come from the handle, for example.

The data from the less accurate sensor is polluted by more noise, so the axis estimation step as presented in [1] does not work properly out of the box. The minimization function 4 becomes less smooth and the optimization gets “stuck” in local optima most of the time, as can be seen in Figure 7-left. We alleviated this problem by performing the rotational estimation on reconstructed point cloud data, which has much less noise and more accurate normals (see Figure 7-middle).

## V. DISCUSSION AND EXPERIMENTAL RESULTS

We evaluated our algorithm using different sensors on multiple household objects (see Figure 2 and 8 for a few examples). The objects were scanned from multiple angles thus presenting various (incomplete) symmetries. When using the pure stereo (without e.g. projected light patterns) and the TOF camera only the boxes were recognized consistently (see Figure 2), the other shapes fell back to triangulation because of the acute sensor inaccuracies. These can be somewhat improved by using a projector together with the stereo camera and by keeping the TOF camera as perpendicular to the surface to be reconstructed as possible. On Hokuyo data it however performed remarkably well (see Figure 8) despite the high noise levels in the  $1 - 3\text{cm}$  range.

The performed evaluation was threefold: i) by manually inspecting the generated models, ii) by calculating the grasp points on the generated models (see Section VI-A) and iii) by using the generated models in the application of shape-based matching of objects in 2D images (see Section VI-B). The invisible side of the objects played an important role especially in the second case (please see the video attachment). Slight inaccuracies in generated models do not pose a problem while carrying out the actual grasping given our manipulator with force-compliant arms. Important for the third type of evaluation was to generate models with as few triangles as possible since the shape matching approaches require significant pre-processing that is polynomially increasing with the number of faces in the model. Our most complex object models, those of the teapots, were generated with roughly 600 faces, and were afterwards successfully located in images (see Figure 8-bottom), thus warranting the use of our 3D data based method for generating the models.

Two important difficulties have to be considered though. One is the problem of having only a single side of a box visible (or not the complete top). In this case it is impossible to judge its extent towards the back reliably. This case can be detected however, and signaled to the robot control program to be accounted for. The other one is that most 3D sensors have problems obtaining reliable distance measurements from thin handles, which typically get curved towards the back, or completely averaged into the background. Most of the time however there is some information present (please see Teapot1 in Figure 8) to aid both applications in giving correct results (please see Figure 1 for the computed grasp).

## VI. APPLICATION SCENARIOS

### A. Force-Closure Grasping

We imported the triangulated meshes together with the located table into a grasp planning software [23], which searches for grasps that are likely to be successful. The planner randomly chooses parameters (for example the approach direction) for the grasp, and then tries to execute it in a kinematic simulation. By detecting collisions between the triangulated mesh of the object and a CAD model of the hand, the planner can find the contact points.

From the contacts, the grasp quality is statically estimated using an algorithm by Ferrari and Canny [24], which – under some simplifying assumptions – scores those grasps higher that are able to resist higher force and torque (as presented in Fig. 8 and the video attachment). A score of zero indicates that the grasp is not force-closed. A selection of the best grasps found by this method for the DLR Hand II can be seen in Fig. 8. Applying those grasps on the real robot is part of our future research agenda. We have however used our models for grasping on the PR2 robot (please see the video attachment).

### B. CAD Model-based Object Detection

We perform 3D position retrieval of objects on 2D images using state-of-the-art 3D shape model matching technique that simulates the 2D appearance of the objects in a shape model generation phase (see [4]). The routine gets object candidates (clusters of points) projected onto the corresponding 2D image from point cloud data and then uses a priori built CAD model of objects to perform the actual detection. The strength of our approach is in that the CAD models must not be modeled manually but can rather be generated by the robot itself (see Figure 8-bottom).

## VII. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a system that can generate reasonable hypotheses of complete models of previously unseen objects, and is precise enough for using it in grasping objects of significant variation from a single view. We have tested our method on different sensors, and due to the accurate surface reconstruction and estimation steps, the fitting methods were able to correctly reconstruct the object models from low resolution and/or noisy data, given reasonable accuracy – as it is the case even with small lasers.



**Fig. 8:** Automatic modeling applied to various objects. **1st row:** objects on kitchen table, **2nd row:** radius estimation results on the reconstructed point cloud (for the interpretation of the colors please see Figure 4), **3th row:** generated geometric models, **4th row:** applications to grasping and matching in 2D images

As the presented methods work on multi-view point clouds as well, they can be run on complete scans of objects in order to obtain a geometric reconstruction automatically. We plan to explore this application scenario as well, and devise ways to evaluate the quality of the matches. Because of the randomized element in the model fitting step, the results have slight variations, and the quality of the matches can not fully be judged based on the number of inliers. In the future we will also work on the integration of the computed surface parameters into grasp planning to exploit the symmetries of the objects in order to generalize successful grasps, and to integrate active verification of the hypothesized back sides of objects using a next-best-view algorithm.

As the models generated for arbitrary rotational objects contain several hundred triangles at least, the matching of these models in camera images is not fast enough for practical use. Simplifying the triangular mesh and employing adaptive re-sampling [19] could alleviate the problem.

**Acknowledgements** – This work was supported by the DFG cluster of excellence CoTeSys (Cognition for Technical Systems) and by Willow Garage, Menlo Park, CA.

#### REFERENCES

- [1] N. Blodow, R. B. Rusu, Z. C. Marton, and M. Beetz, "Partial View Modeling and Validation in 3D Laser Scans for Grasping," in *9th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Paris, France, December 7-10 2009.
- [2] Z. C. Marton, L. Goron, R. B. Rusu, and M. Beetz, "Reconstruction and Verification of 3D Object Models for Grasping," in *Proceedings of the 14th International Symposium on Robotics Research (ISRR09)*, Lucerne, Switzerland, August 31 - September 3, 2009.
- [3] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Close-range Scene Segmentation and Reconstruction of 3D Point Cloud Maps for Mobile Manipulation in Human Environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, October 11-15 2009.
- [4] M. Ulrich, C. Wiedemann, and C. Steger, "Cad-based recognition of 3d objects in monocular images," in *International Conference on Robotics and Automation*, 2009, pp. 1191–1198.
- [5] A. Coates, P. Baumstarck, Q. V. Le, and A. Y. Ng, "Scalable learning for object detection with gpu hardware," in *IROS*, 2009, pp. 4287–4293.
- [6] M. Fritz, T. Darrell, M. Black, G. Bradski, and S. Karayev, "An additive latent feature model for transparent object recognition," in *NIPS*, S. for Oral Presentation, Ed., 12/2009 2009.
- [7] U. Hillenbrand, "Pose clustering from stereo data," in *Proceedings VISAPP International Workshop on Robotic Perception – RoboPerc*, 2008.
- [8] B. Steder, G. Grisetti, M. Van Loock, and W. Burgard, "Robust on-line model-based object detection from range images," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, October 2009.
- [9] M. Richtsfeld and M. Vincze, "Grasping of Unknown Objects from a Table Top," in *Workshop on Vision in Action: Efficient strategies for cognitive agents in complex environments*, 2008.
- [10] A. Miller and P. K. Allen, "Graspit!: A Versatile Simulator for Robotic Grasping," *IEEE Robotics and Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.
- [11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004.
- [12] A. Collet, D. Berenson, S. S. Srinivasa, and D. Ferguson, "Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009.
- [13] G. Biegelbauer and M. Vincze, "Efficient 3D Object Detection by Fitting Superquadrics to Range Image Data for Robot's Object Manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, Rome, Italy, 2007.
- [14] Y. Zhang, A. Koschan, and M. Abidi, "Superquadric Representation of Automotive Parts Applying Part Decomposition," *Journal of Electronic Imaging, Special Issue on Quality Control by Artificial Vision*, Vol. 13, No. 3, pp. 411–417, 2004.
- [15] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for Point-Cloud Shape Detection," *Computer Graphics Forum*, vol. 26, no. 2, pp. 214–226, June 2007.
- [16] S. Thrun and B. Wegbreit, "Shape from symmetry," in *Proceedings of the International Conference on Computer Vision (ICCV)*. Beijing, China: IEEE, 2005.
- [17] R. B. Rusu, Z. C. Marton, N. Blodow, A. Holzbach, and M. Beetz, "Model-based and Learned Semantic Object Labeling in 3D Point Cloud Maps of Kitchen Environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, October 11-15 2009.
- [18] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, no. 24/6, 1981.
- [19] Z. C. Marton, R. B. Rusu, and M. Beetz, "On Fast Surface Reconstruction Methods for Large and Noisy Datasets," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 12-17, 2009.
- [20] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, "Learning Informative Point Classes for the Acquisition of Object Model Maps," in *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Hanoi, Vietnam, December 17-20, 2008.
- [21] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 12-17, 2009.
- [22] A. E. Johnson, "Spin-images: A representation for 3-d surface matching," Tech. Rep., 1997.
- [23] J. Kleinhelfort, "Implementation of a robotic simulator for grasp planning," Dec. 2009. [Online]. Available: <http://www9.in.tum.de/~kleinehe/sep-grasp-planning.pdf>
- [24] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, May 1992.