

# Efficient 3D Object Detection by Fitting Superquadrics to Range Image Data for Robot's Object Manipulation

Georg Biegelbauer and Markus Vincze  
Automation and Control Institute, E376  
Vienna University of Technology, Austria  
{gb,vm}@acin.tuwien.ac.at

**Abstract**—Fast detection of objects in a home or office environment is relevant for robotic service and assistance applications. In this work we present the automatic localization of a wide variety of differently shaped objects scanned with a laser range sensor from one view in a cluttered setting. The daily-life objects are modeled using approximated Superquadrics, which can be obtained from showing the object or another modeling process. Detection is based on a hierarchical RANSAC search to obtain fast detection results and the voting of sorted quality-of-fit criteria. The probabilistic search starts from low resolution and refines hypotheses at increasingly higher resolution levels. Criteria for object shape and the relationship of object parts together with a ranking procedure and a ranked voting process result in a combined ranking of hypothesis using a minimum number of parameters. Experiments from cluttered table top scenes demonstrate the effectiveness and robustness of the approach, feasible for real world object localization and robot grasp planning.

## I. INTRODUCTION

Detecting and localizing objects is a fundamental task of robotic systems. The task is of great importance in industrial applications for the automation of part production and the ultimate goal to achieve a lot size of one or pure customization. In this work, we present the automatic localization of a wide variety of differently shaped objects scanned from one view in a cluttered setting. The resulting object pose determination enables a grasp planning of a desired robot arm.

The classical approach is to use intensity or color cameras to exploit the appearance of objects for the detection task. Because shape is not directly encoded, this problem is in general difficult or ill-posed [3]. However recent progress in invariant feature extraction is the basis to obtain first good results in realistic settings [6], [12], [16]. Although these approaches are rather fast, they do not work satisfactorily in cluttered scenes and “inherit the major problem of intensity-based systems, that is, dependency on lighting conditions” [11].

To overcome these problems with intensity images the 3D shape of the objects has to be directly recovered from range images. Multiple parametric models have been introduced for 3D object recovery, but the Superquadrics are perhaps the most popular for several reasons. The compact shape can be described with a small set of parameters ending up in a large variety of different basic shapes. The recovery of

Superquadrics has been well investigated and even global deformations can be easily adopted [21]. They can be used as volumetric part-based models desirable for robotic grasping operations. These advantages cannot be found in other geometric entities which predestines the Superquadric model for our application.

### A. Related Work

Solina et al. pioneered work in recovering single Superquadrics with global deformations in a single-view point cloud [21] and demonstrated that recovering a Superquadric from range data is sensitive to noise and outliers, which renders a stable object recognition difficult. To overcome this problem many approaches assume that full 3D data of the objects is available to estimate the complete set of model parameters [5], [7], [14]. On the other hand, full 360° views are difficult to obtain in practice and for object grasping it is sufficient to estimate pose parameters while the task constraint specifies shape and size parameters [20], [24].

Much progress has been made in the last decade in tackling the recognition problem by acquiring a single-view range image and interpreting the scene using Superquadrics. Leonardis et al. introduced the recover and select paradigm for segmenting a scene with simple geometric objects without occlusions [15]. This method aims at a full search with an open processing time incompatible to robotic applications. Subsequent work has improved this segmentation, e.g. Krivic et al. demonstrated the recovery of a known complex object in a scene using the connectivity information of the Superquadrics [13] handling sparse scene occlusions by using the redundancy information of the part connections. Tao et al. also has presented an improvement of Leonardis' segmentation [22] using random samples with the focus on speeding up the Superquadric fit in noisy range data. This approach comes close to filling the needs of robotic applications but only using undeformed Superquadrics and concentrating his work on detecting pipes in sonar range images. Taylor et al. first segments the image for describing the scene [23] but saves processing time with the limitation of detecting single geometric primitives disclaiming the flexibility of Superquadrics to describe objects.

Katsoulas proposes a novel object detection approach searching for box-like objects using parabolically deformable Superquadrics [11]. He weakens the bottleneck of the scene segmentation using a 3D edge detector and achieved some

This work is supported by the Austrian Science Foundation grants S9101-N04 and S9103-N04.

improvement in processing time, but this method cannot handle non box-like objects and scene occlusions.

All these approaches use segmentation to analyse the scene. It is however argued that a bottom up segmentation is in conflict with purposeful object detection. The detection of a known object does not require segmentation. Moreover, segmentation wastes valuable processing time as shown in the experimental comparison of different range image segmentation algorithms [10]. One of the main contributions of our work achieving fast object detection is the lack of necessity for any segmentation step.

### B. Problem Description and Method of Resolution

The approach we propose exploits a robot mounted laser range sensor, which scans the scene of interest. The acquired range data is sparse due to the single-view scan, exhibiting the typical laser and camera shadows, and the objects are only partly visible in cluttered scenes typically encountered in realistic scenarios. Detecting an object in this range data is difficult and the task given to the robot is to find and locate an object specified by the user within a few seconds. The benefit of our approach is that we are able to handle these problems with the object's occlusions and intersections, because of the purposeful approach to search for a given predefined object model.

The second problem to deal with is the single-view range image. The information of the objects' rear is not available. What can be exploited is the symmetry of the real-world objects. Furthermore, keeping the set of model parameters low, it is not possible to model complex shaped object with a single Superquadric or a set of Superquadrics with global deformations. Therefore, the model describes the approximated overall shape of the object where the details in shape are not present in the sparse range data. Example objects, which are commonly used in a human-robot interaction and which can all be handled with the proposed approach, are e.g. pen, cup, bowl, book, ball, computer mouse, or tools.

The contribution is a rapid and reliable detection and pose determination of known objects. It is shown that a hierarchical search in the sense of a coarse to a fine processing (sub-scaling the range data) increases computational efficiency and saves valuable processing time [4]. Clever generation and verification of pose hypotheses is the key for efficient object detection [25]. Hence, the proposed approach, achieving fast results, is embedded in a hierarchical structure starting with a RANSAC-based object search to find pose hypotheses followed by a pose refinement. In the second step these hypotheses are verified executing a ranked voting process [18] over the sorted quality-of-fit criteria (the measure of fit, the number of interior points and the number of points on the Superquadrics surface) to robustly select the final object pose. Moreover, we shown that searching for a tapered Superquadric with fixed size and shape makes the Levenberg-Marquardt minimization process [17] less complex, less time consuming and more stable. Using a model description with several Superquadrics extends the ranked voting process with two additional parameters describing the relationship

between the Superquadrics, the main axis steradian and the center distance.

The paper is structured by starting with a short introduction to Superquadrics in Section II. Section III presents the approach in detail and Section IV give experimental results to evaluate the proposed method. Section V gives the final conclusion.

## II. THE SUPERQUADRICS MODEL

Superquadrics are a family of parametric shapes which were first introduced in computer graphics by Barr [1] in 1981. Superquadrics can be classified into Superellipsoid, Supertoroid and Superhyperboloid with one and two parts. In this work we focus on the Superellipsoid which is useful for a volumetric part-based object description, because they are compact in shape and have a closed surface.

The implicit form to describe a Superquadric is given by

$$F(\mathbf{x}) = \left( \left( \frac{x}{a_1} \right)^{\frac{2}{\varepsilon_2}} + \left( \frac{y}{a_2} \right)^{\frac{2}{\varepsilon_2}} \right)^{\frac{\varepsilon_2}{\varepsilon_1}} + \left( \frac{z}{a_3} \right)^{\frac{2}{\varepsilon_1}}, \quad (1)$$

where  $\varepsilon_1$  and  $\varepsilon_2$  are the shape parameters, ranging from 0.1 to 1, and  $a_1$ ,  $a_2$  and  $a_3$  are the scale parameters along the  $x$ ,  $y$  and  $z$ -axis of the Superquadric. We refer to equation (1) as inside-outside function  $F$  of a given point  $\mathbf{x}$ . If  $F(\mathbf{x}) = 1$  the point is on the Superquadric surface. If  $F(\mathbf{x}) < 1$  the point is inside the Superquadric and visa versa. To cover most of the object part shapes global deformations of the Superquadrics are required which were also introduced by Barr [2]. We use global tapering along the Superquadric's  $z$ -axis. The taper transformation and its coefficients  $k_x$  and  $k_y$  are defined in [21] with a range of  $-1 < k_x, k_y < 1$ . A tapered Superquadric in general position  $(p_x, p_y, p_z)$  and orientation  $(\phi, \theta, \psi)$  has the ability to represent shapes of a sphere, cuboid, cylinder, pyramid, cone, wedge and all shapes in between with the parameter set  $\Lambda \{a_1, a_2, a_3, \varepsilon_1, \varepsilon_2, \phi, \theta, \psi, p_x, p_y, p_z, k_x, k_y\}$ . Given a set of points  $\mathbf{x}_k$  the parameter set  $\Lambda$  is recovered applying the Levenberg-Marquardt algorithm [17] in a least-squares minimization. Extending the inside-outside function to the parameter set  $\Lambda$  the following expression must be minimized

$$\min \sum_{k=0}^n (F^{\varepsilon_1}(\mathbf{x}_k; \Lambda) - 1)^2. \quad (2)$$

The exponent  $\varepsilon_1$  in equation (2) is necessary for optimal minimization to have a correct distance measure independent to the shape parameter. Note the missing factor  $\sqrt{a_1 \cdot a_2 \cdot a_3}$  for a volume minimizing recovery as described in [21]. In our object detection approach the shape, scale and taper parameters are given by the target object and the pose (position and orientation) parameters are estimated.

## III. OBJECT DETECTION

Any task-based system needs to first learn about the relevant object. This is achieved by specifying a simplified Superquadric model of the wanted object or a unique part of it. The size, shape and taper parameter of the parameter set  $\Lambda$

and potentially the Superquadric relations are set according to the best object description. A standard method to describe this model is to view the target object alone and fit the object to the data.

We assume for our approach that the object to be detected is in the scanned area. Otherwise the object, which is most similar to the searched object will be selected. Fig. 1 outlines the detection approach.

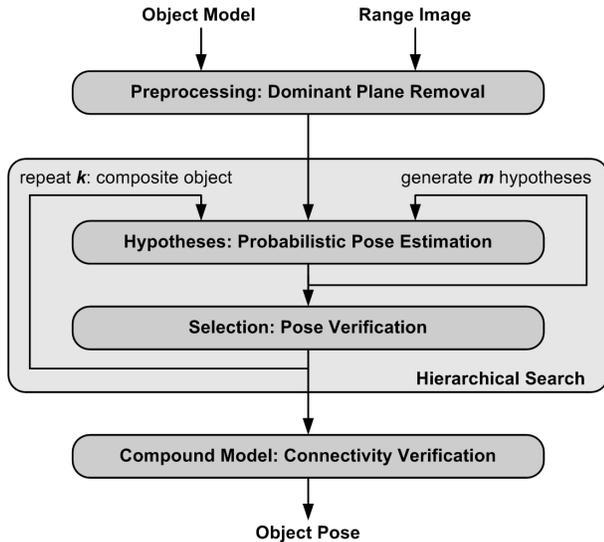


Fig. 1. Flow chart of the proposed object detection approach.

The system receives the object model and the scene represented by an unstructured point cloud scanned from one-view. The preprocessing removes the dominant plane from the range image which is in most cases the table or the ground plane. This information is not necessary for detection.

In order to achieve fast detection results a probabilistic approach is used to generate pose hypotheses. For keeping the computational effort low the search process is structured in a two-level hierarchy.

First the low-level search (Probabilistic Pose Estimation) is RANSAC-based [8] with  $n$  samples on sub-scaled raw data to speed up the Superquadric recovery using the Levenberg-Marquardt [17] minimization. The best fit of the low-level search is iteratively refined. To ensure a true positive detection result  $m$  hypotheses are computed, which indicates how often the object could appear in the view of the given scene.

Secondly, a high-level selection (Pose Verification) uses the full resolution image to verify the object existence. To achieve this a voting [18] for the pose hypotheses is proposed considering three constraints: the measurement of fit, the number of points on the Superquadric surface and the number of the Superquadric's interior points.

If the model describes a composite object ( $k > 1$ ) the pose verification has to be extended to the compound model pose (Connectivity Verification). An additional voting of the Superquadrics relationships, the main axis steradian and the center distances, is used. Detailed information about the processing steps is given in the following sections.

### A. Preprocessing: Dominant Plane Removal

Most of the range images of a table scene consist of the table plane. However, this plane is not needed for the object detection and, so much the worse, it slows down the object detection process and raises the likelihood of false detections. Hence, the first step is to detect and remove the raw data points associated with the ground plane. Furthermore, the plane can be used as reference for the robot. We define the ground plane as the dominant plane in the range image associating more raw data points than in the rest of the area.

Finding the dominant plane is achieved by fitting local plane patches and extending them to verify a global plane. We radically decrease the resolution of the range image to 100 seed points equally distributed over the area. Each point is now a seed for fitting the plane patch to the points in the close neighborhood. The normal vector of the surface patch corresponds to the vector  $\mathbf{n}$  which is determined by the eigenvalue problem  $C\mathbf{n} = \lambda_{min}\mathbf{n}$ , where  $\lambda_{min}$  is the smallest eigenvalue and  $C$  is the covariance matrix of the surface patch point set. The plane point is the mean of the surface patch point set. Each plane hypothesis is now verified against the other 99 points by calculating the normal distances to the plane. A point belongs to the plane if the distance is smaller than the median of all normal distances. The plane hypothesis with the most supporting points is the dominant plane. If it contains more raw data points than in the rest of the area the raw data points of this plane are removed, otherwise no ground plane is detected.

### B. Hierarchical Search

The computational bottle neck of processing is the iterative Levenberg-Marquardt algorithm minimizing six variables simultaneously (position and orientation of the Superquadric). This time consuming algorithm is executed as often samples are computed. Its processing time depends on the number of points to be fitted and on the number of iterations to converge the optimization.

The classical RANSAC algorithm is a robust and reliable method to detect an object. Using the hierarchical structure we show that we achieve more detection robustness in less time than  $n \cdot m$  RANSAC samples. The reason being is that our method works on different scaling levels applying different evaluation criteria. Detailed information about the hierarchical processing – finding the hypotheses and selection – is given in the following sections.

An important minor detail for every Superquadric fit minimization should be mentioned. While recovering a set of points the initial  $z$ -axis orientation of the Superquadric must be assigned to a local coordinate system build with extracted from the central moments of the point cloud to be fitted. The solution proposed here is to execute three estimations (sub-sampled Superquadric fits with less iterations) where the  $z$ -axis is aligned with the three central moment directions and to then choose the best fit. This eliminates the shape ambiguities in a robust way.

1) *Hypotheses: Probabilistic Pose Estimation:* For finding first hypotheses in the low-level search the RANSAC algorithm is exploited. The number of hypotheses  $m$  that will be sampled depends on the ratio of the Superquadric size and the range image dimensions to allow for cover the whole scene. Each sample calculation is started by picking a random point in the raw data set and using a set of random points (30 points) in the seed point's neighborhood within a radius  $r = \min(a_1, a_2, a_3)$ , the smallest parameter for the Superquadric size. Using this point set, a Superquadric is recovered, where the algebraic distance  $\mathcal{M}$  is the measure of fit [9].  $\mathcal{M}$  is the mean of the algebraic distance calculated with the inside-outside function of equation (1), that is,

$$\mathcal{M} = \sqrt{\frac{1}{n} \cdot \sum_{k=0}^n (F^{\varepsilon_1}(\mathbf{x}_k; \Lambda) - 1)^2}. \quad (3)$$

The pose estimation with the smallest value of  $\mathcal{M}$  is then used for a refinement step. This is necessary because the pose hypotheses found by the RANSAC algorithm are not well aligned with the raw data points. The refinement is an iterative process on the local full resolution data points to achieve a better adjustment of the Superquadric and the iteration is stopped if no more improvement of the measure of fit  $\mathcal{M}$  can be achieved. The result of the probable pose estimation step are  $m$  object pose hypotheses.

2) *Selection: Pose Verification:* The found and refined pose hypotheses from the low-level search are now verified using voting with additional criteria exploiting full resolution data. The measure of fit  $\mathcal{M}$  is not a sufficient criteria for a correct object detection. Shape and size ambiguities in the scene may lead to false detections because it is a fit to a local surface patch. Additional information is needed to verify the pose hypotheses. Exploiting the fact that the Superquadric object model represents the surface of the entire object no raw data points of the single-view range image should be located inside the recovered Superquadric. And secondly, the probability that a recovered Superquadric represents the object searched increases with the number of raw data points located on the Superquadric's surface.

Let  $X_n$  be a set of raw data points in the neighborhood of the recovered Superquadric, that are points within the radius  $r = 1.2 \cdot \max(a_1, a_2, a_3)$  of the Superquadrics' center. We can define the evaluation criteria  $\mathcal{I}$  for the number of the Superquadric's interior points and  $\mathcal{S}$  for the Superquadric's surface points by

$$\mathcal{I} = \left\{ \sum_{k=0}^n \mathbf{x}_k \in X_n \mid F^{\varepsilon_1}(\mathbf{x}_k; \Lambda) < 1 - \gamma \right\}, \quad (4)$$

$$\mathcal{S} = \left\{ \sum_{k=0}^n \mathbf{x}_k \in X_n \mid 1 - \gamma \leq F^{\varepsilon_1}(\mathbf{x}_k; \Lambda) < 1 + \gamma \right\}, \quad (5)$$

where  $\gamma$  can be selected depending on the application specification required in all expressions. It can be seen as model approximation tolerance, where for all experiments presented below  $\gamma$  is set to 0.2 allowing a 20% deviation.

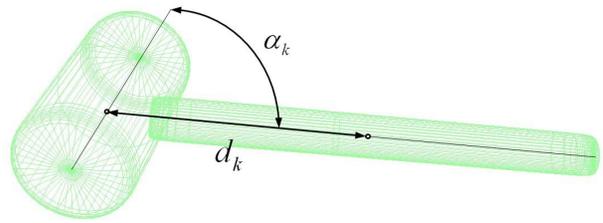


Fig. 2. The main axis steradian  $\alpha_k$  and the distance  $d_k$  between the Superquadric centers describe the connectivity of a composite object.

A ranked voting using these three criteria  $\mathcal{M}$ ,  $\mathcal{I}$  and  $\mathcal{S}$  is the key to achieving a robust hypothesis selection. The voting process is performed using the sorted ranks of  $\mathcal{M}$ ,  $\mathcal{S}$  and  $\mathcal{I}$ . The hypothesis with the lowest sum of all ranks is selected. This procedure does not need any parameters but uses the natural sorting, and combines three criteria, which individually describe only specific object characteristics but which together give a total constraint to select the best hypothesis.

To summarize, this hierarchical two-level search achieves a fast and robust detection result especially in cluttered scenes. Because of fitting the object model to local surface patches and verify them globally within the verification step, disconnected surface patches can be associated to one part. This enables the robust detection of partly occluded objects.

### C. Compound Model: Connectivity Verification

A composite object using a model description with two or three Superquadrics needs an extension in the voting process to include the relationship between the Superquadrics. Krivic et al. uses two vectors for the joint positions and a  $ZYZ$  rotation [13] to describe the linked Superquadrics. For our purpose we show that it is sufficient to use a more efficient connectivity description despite the fact that the description is not unique. We found that the steradian between the two neighboring Superquadric's  $z$ -axes ( $\mathcal{A}$ ) and the distance between the Superquadric's centers ( $\mathcal{D}$ ) is sufficient. This reduces the connection parameters from 6 to 2. These two evaluation criteria are exploited to increase the detection robustness using again voting [18]. Defining the vector  $\mathbf{a}$  as Superquadric  $z$ -axis and the point  $\mathbf{x}_c$  as Superquadric center brings us to the following criteria,

$$\mathcal{A} = \left| \alpha_k - \arccos \left( \frac{\mathbf{a}_i \cdot \mathbf{a}_j}{|\mathbf{a}_i| \cdot |\mathbf{a}_j|} \right) \right|, \quad (6)$$

$$\mathcal{D} = \left| d_k - \sqrt{(x_{c_i} - x_{c_j})^2 + (y_{c_i} - y_{c_j})^2 + (z_{c_i} - z_{c_j})^2} \right|. \quad (7)$$

The evaluation criteria  $\mathcal{A}$  and  $\mathcal{D}$  are the absolute differences between the permutations of all  $m$  low level fits and the nominal angle  $\alpha_k$  and the distance  $d_k$  specified in the model (see Fig. 2). Again the sorted ranks  $\mathcal{A}$  and  $\mathcal{D}$  for the final voting are calculated and stored in a list. The lowest sum of the ranks of each evaluation criteria  $\mathcal{A}$  and  $\mathcal{D}$  with their associated result  $\mathcal{M}$ ,  $\mathcal{I}$  and  $\mathcal{S}$  represents the parameter set  $\Lambda_k$  of the object detected.

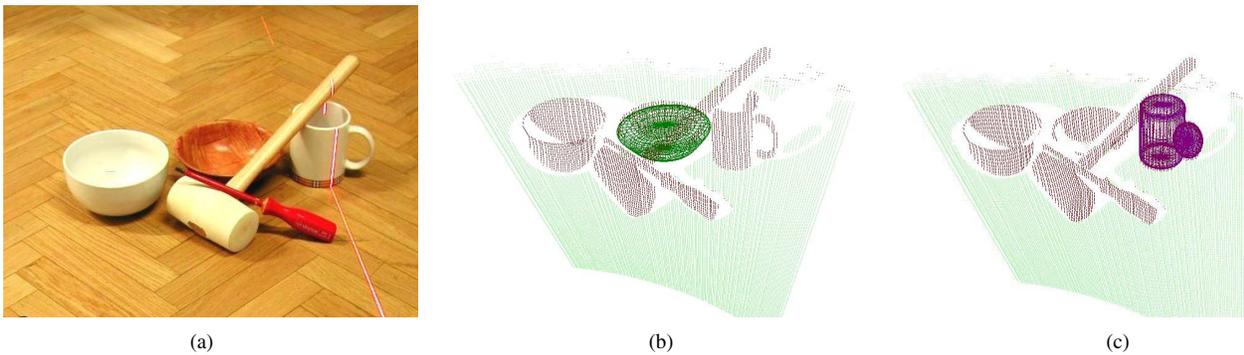


Fig. 3. (a) Intensity image of a typical scene consisting of a ceramic and a wooden bowl, a cup, a screwdriver and a rubber-mallet. Detection results are shown of the ceramic bowl (b) and the cup (c) modeled with two Superquadrics.

#### IV. EXPERIMENTAL EVALUATION

The experimental set-up consists of a triangulation laser range sensor from IVP mounted on a pan-tilt unit from AMTEC. The range image is acquired by sweeping the laser plane over the scene (compare Fig. 3). The resolution of the sensor is about  $2\text{mm}$  at a distance of  $1\text{m}$ . To obtain reasonable image quality (see Fig. 3(a)), about 150 scan lines are required for such a scene. For the experiments we used a Pentium IV 2.0 GHz PC. The algorithm is implemented in C++ and for displaying the results the Visualization Tool Kit (VTK)<sup>1</sup> is used. All experiments are carried out using  $n = 10$  initial RANSAC trials and 20 minimization iterations for each Superquadrics fit.

Fig. 3 shows a typical occluded arrangement of objects. The scanned range image consist 37708 points and the first experiment investigates the detection behavior in case of similar objects present in one scene. Fig. 3(b) shows the correctly detected wooden bowl with a similar white ceramic bowl to the right of it. The detection took 2.4 seconds and note the heavily occluded bowl and the broken surface patches. The second experiment demonstrates the accuracy of detecting a cup in the same scene modeled with two Superquadrics (the relations are defined in Tab. I). The cup is also partly occluded but the detection needed 27.3 seconds because of finding ( $m = 100!$ ) the small handle in the scene. However note, the sparse data available from the cup (especially of the handle) and the robust detection of it in 6 degrees of freedom suitable for robotic grasping tasks. Beside that, Salganicoff et al. showed that for a grasp point planning the Superquadrics' size, shape and pose parameter are sufficient. Detecting the cup in 5 degrees of freedom (only the cup-body) speeds up the processing time significantly (3.1 seconds). Tab. I summarizes the model and algorithm parameter of the bowl and the cup. Summarizing the first two experiments, the results show that due to the hierarchical RANSAC approach the processing time is nearly independent of the range image size (number of points), even on noisy data, but it depends on the number of Superquadrics and the model size related to the scene dimensions.

<sup>1</sup>Freely available open source software. (<http://public.kitware.com/vtk>)

The last experiment investigates the object detection in a cluttered scene in terms of detection probability using a composite object model exploiting the Superquadric relationships. Fig. 4 illustrates the correctly detected rubber-mallet in a cluttered tool scene (range image size: 31101 points). Trying to detect the rubber-mallet with one Superquadric in this scene with many shape ambiguities does not end in a robust result. To evaluate the performance we examined 100 experiments on the same scene describing the rubber-mallet first with the head model only, second with the shaft only and finally with both Superquadric models exploiting their relationship. The results are summarized in Tab. II. Note, that trying to detect the mallet with a single Superquadric causes a false detection in every second trial. Only when exploiting the geometric relationship the true positive detections of the rubber-mallet raise to 76%. The remainder are false detections due to the similarity of objects. In the scene in Fig. 3 without similar shaped object parts (like the shaft of the metal hammer and the spray tin in Fig. 4) the detection rate of the rubber-mallet is 90%. The reason for not obtaining 100% is that the number of hypotheses  $m$  is bounded from 10 to 100 due to computational efficiency. It still remains a random process and if  $m$  is increased the detection results converge towards 100%.

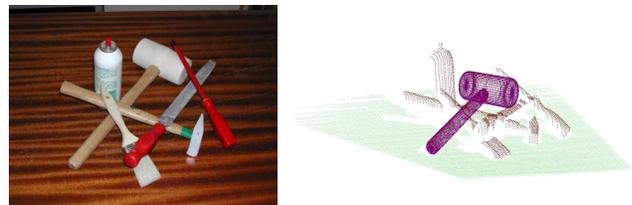


Fig. 4. Detection result of a rubber-mallet in a cluttered tool scene of similar objects exploiting the relation of the two modeled Superquadrics.

A comparison to previous detection methods showed that our algorithm achieves good results, even if the Superquadric model is far apart from the objects shape. The method from Krivic et al. [13] is the one with the most similar object detection performance and modeling approach. As in the introduction stated, the segmentation wastes much time, and in this case more than 10 minutes, while the object detection on the pre-segmented data sets needs approximately the same

TABLE I

SUMMARY OF THE ALGORITHM AND SUPERQUADRIC PARAMETER OF THE OBJECT DETECTION RESULTS IN FIG. 3.

part	$a_1$	$a_2$	$a_3$	$\varepsilon_1$	$\varepsilon_2$	$k_x$	$k_y$	$m$	$k$	$\alpha_k$	$d_k$
Bowl	60mm	60mm	20mm	0.4	1.0	0.3	0.3	10	-	-	-
Cup-Body	38mm	38mm	48mm	0.1	1.0	0.0	0.0	12	-	-	-
Cup-Handle	30mm	23mm	8mm	0.5	1.0	0.0	0.0	100	-	-	-
Cup	-	-	-	-	-	-	-	-	2	90°	50mm

TABLE II

PERFORMANCE EVALUATION OF THE OBJECT DETECTION IN FIG. 4 AND SUMMARY OF THE ALGORITHM AND SUPERQUADRIC PARAMETER.

part	$a_1$	$a_2$	$a_3$	$\varepsilon_1$	$\varepsilon_2$	$k_x$	$k_y$	$m$	$k$	detect	$\alpha_k$	$d_k$	time
Head	30mm	30mm	56mm	0.1	1.0	0.0	0.0	10	1	48%	-	-	2.3s
Shaft	12mm	10mm	120mm	0.1	1.0	0.0	0.0	26	1	55%	-	-	7.1s
Mallet	-	-	-	-	-	-	-	36	2	76%	90°	170mm	10.2s

amount of time as our detection approach without a previous segmentation procedure.

## V. CONCLUSION AND FURTHER STEPS

To sum up, we introduced a 3D object detection approach with a geometric model description using approximated Superquadrics. Fast and robust object detections are achieved combining the RANSAC algorithm with a hierarchical sub-scale search. A key for robust detection is the introduction of two new criteria of fit and a ranked voting in the hypothesis selection step. For compound objects two more criteria have been proposed and evaluated using the same voting procedure. Experiments confirmed the rapid and reliable detection of every day objects.

Future work will deal with learning the model parameters in a more natural way applied to convex shaped objects. Showing a cognitive system the object which it should detect, the system should be able to extract the relevant Superquadric parameters. This will enable the system to learn new tasks and to improve the human robot interface. First steps are being taken to achieve this goal to realize an autonomous robot system. Acquiring a color range image of the object enables a geometric decomposition [19] and a final Superquadric fit to describe the object. First tests were successful but more work is needed to achieve robust extraction of the model parameters.

## REFERENCES

- [1] A.H. Barr. Superquadrics and angle preserving transformations. *IEEE Computer Graphics and Applications*, 1(1):11–23, 1981.
- [2] A.H. Barr. Global and local deformations of solid primitives. In *ACM SIGGRAPH Computer Graphics, Proceedings of the 11th Annual Conference on Computer graphics and interactive techniques*, volume 18(3), pages 21–30, 1984.
- [3] M. Bertero, T.A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988.
- [4] V. Cantoni and L. Lombardi. Hierarchical architectures for computer vision. In *Proceedings of the IEEE Euromicro Workshop on Parallel and Distributed Processing*, pages 392–398, 1995.
- [5] L. Chevalier, F. Jaillet, and A. Baskurt. Segmentation and superquadric modeling of 3D objects. *J. of WSCG03*, 11(1), 2003. ISSN 1213-6972.
- [6] V. Ferrari, T. Tuytelaars, and L. van Gool. Simultaneous object recognition and segmentation by image exploration. In *Proceedings of the European Conference on Computer Vision*, pages 40–54, 2004.
- [7] F.P. Ferrie, J. Lagarde, and P. Whaitte. Recovery of volumetric object descriptions from laser rangefinder images. In *Proceedings of the European Conference on Computer Vision*, pages 387–396, 1990.
- [8] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–385, 1981.
- [9] A. Gupta, L. Bogoni, and R. Bajcsy. Quantitative and qualitative measures for the evaluation of the superquadric models. In *Proceedings of the IEEE Workshop on Interpretation of 3D Scenes*, volume 12(2), pages 162–169, 1989.
- [10] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K.W. Bowyer, D.W. Eggert, A.W. Fitzgibbon, and R.B. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.
- [11] D. Katsoulas. Reliable recovery of piled box-like objects via parabolically deformable superquadrics. In *Proceedings of the IEEE 9th Int. Conference on Computer Vision*, pages 931–938, 2003.
- [12] D. Kragic, M. Bjorkman, H.I. Christensen, and J.-O. Eklundh. Vision for robotic object manipulation in domestic settings. *Robotics and Autonomous Systems, Elsevier*, 23(11):943–955, 2005.
- [13] J. Krivic and F. Solina. Part-level object recognition using superquadrics. *Computer Vision and Image Understanding, Elsevier*, 95(11):105–126, 2004.
- [14] S.H. Lee, H.K. Hong, and J.S. Choi. Assembly part recognition using part-based superquadric model. In *Proceedings of the IEEE TENCON*, pages 479–482, 1999.
- [15] A. Leonardis and A. Jaklic. Superquadrics for segmenting and modeling range data. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(11):1289–1295, 1997.
- [16] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision, Kluwer*, 60(2):91–110, 2004.
- [17] J.J. Moré. The levenberg-marquardt algorithm: Implementation and theory. *Numerical Analysis - Lecture Notes in Mathematics, Springer Verlag*, 630:105–116, 1977.
- [18] B. Parhami. Voting algorithms. *IEEE Transactions on Reliability*, 43(4):617–629, 1994.
- [19] A. Pichler, R.B. Fisher, and M. Vincze. Decomposition of range images using markov random fields. In *Proceedings of the IEEE Int. Conference on Image Processing*, volume 2, pages 24–27, 2004.
- [20] M. Salganicoff, L.H. Lyle, and R. Bajcsy. Active learning for vision-based robot grasping. *Machine Learning*, 23(2):251–278, 1996.
- [21] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–147, 1990.
- [22] L. Tao, U. Castellani, and V. Murino. Robust 3D segmentation for underwater acoustic images. In *IEEE 2nd Int. Symposium on 3D Data Processing, Visualization, and Transmission*, pages 813–819, 2004.
- [23] G. Taylor and L. Kleeman. Robust range data segmentation using geometric primitives for robotic applications. In *Int. Conference on Signal and Image Processing*, pages 467–472, 2003.
- [24] G. Taylor and L. Kleeman. Integration of robust visual perception and control for a domestic humanoid robot. In *IEEE Int. Conference on Intelligent Robots and Systems*, pages 1010–1015, 2004.
- [25] M. Wheeler and K. Ikeuchi. Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):252–265, 1995.