

Superquadric Segmentation in Range Images via Fusion of Region and Boundary Information

Dimitrios Katsoulas, Christian Cea, and Dimitrios Kosmopoulos

D.Katsoulas and D. Kosmopoulos are with the Institute of Informatics and Telecommunications, Computational Intelligence Laboratory, National Center for Scientific Research Demokritos, GR-15310 Aghia Paraskevi, Athens, Greece

C. Cea is with the Computer Science Department, Institute for Pattern Recognition and Image Processing, University of Freiburg, Georges-Koehler-Allee, Gebaude 052, D-79110 Freiburg i.Br., Germany

Abstract

The high potential of Superquadrics as modeling elements for image segmentation tasks has been pointed out since years in the computer vision community. In this work we employ superquadrics as modeling elements for multiple object segmentation in range images. Segmentation is executed in two stages. Firstly, a hypothesis about the values of the segmentation parameters is generated. Secondly, the hypothesis is refined locally. In both stages, object boundary and region information are considered. Boundary information is derived via model-based edge detection in the input range image. Hypothesis generation uses boundary information to isolate image regions which can be accurately described by superquadrics. Within hypothesis refinement, a game-theoretic framework is used to fuse the two information sources by associating an objective function to each information source. Iterative optimization of the two objective functions in succession, outputs a precise description of all image objects. We demonstrate experimentally, that this approach substantially improves the most established method in superquadric segmentation, in terms of accuracy and computational efficiency. We demonstrate the applicability of our segmentation framework in real world applications by constructing a novel robotic system for automatic unloading of jumbled box-like objects from platforms.

Index Terms

I.4.8.g range data, I.4.8.j shape, I.4.9 applications, I.4.7.e size and shape, I.4.6.e region growing, partitioning, I.4.6.a edge and feature detection, I.4.8.l surface fitting.

I. INTRODUCTION

Image segmentation is admittedly a difficult problem in machine vision, and despite the promising techniques already proposed, e.g., in [1], [2], [3], [4], we are still far from solving the problem in its generality. Considerable improvement on the output of the segmentation task regarding all robustness, accuracy, and computational efficiency can be achieved when *prior* knowledge related to various properties of the objects in the image is utilized. Such knowledge is incorporated in the segmentation task using modeling entities, or simply *models*, able to accurately express the object properties. This work contributes to segmentation of three- dimensional objects. A variety of three- dimensional entities have already been proposed for their modeling. Among them, generalized cylinders [5], implicit (fourth degree) polynomials [6], geons [7], [8], spherical harmonic and Fourier surfaces [9], [10], symmetry seeking models [11], blob models [12], hyperquadrics [13], and others. In this work we employ *Superquadrics* [14], [15], [16], [17] as modeling entities. The reasons for selecting

the particular models are the following: Firstly, superquadrics comprise a small number of parameters with intuitive meaning, which makes their handling straightforward. Secondly, they have a large expressive power, so that they can model a big number of artificial and natural objects. Thirdly, their expressiveness can be easily further enhanced by the addition of global and local shape deformations. Fourthly, closed form expressions exist for the models, as well as for special curves on their surface, which makes both rendering and usage fast and direct. Finally, fast and robust methods for superquadric fitting to three dimensional data have been so far developed.

This work describes a new framework for segmenting images that contain multiple piled objects, which are modeled by superquadric entities. Input is a range image of the object configuration. Usage of range imagery is advantageous in our case, where three- dimensional objects are dealt with, because object depth information is incorporated in the image. Using intensity images from multiple cameras is an alternative, but in this case three dimensional object information must be reconstructed from intensity images, which is known to be strenuous and of questionable robustness, mainly due to illumination variability and camera calibration errors. Our strategy improves the state-of-the-art in this domain by an order of magnitude in speed and reduces the model fitting error to a factor of about 3. In the remainder of this section we describe how superquadric models have been employed up to now for performing segmentation, we give a rough description of our approach, and discuss the contributions of this work.

A. Related work on superquadric segmentation

Existing approaches attempting superquadric segmentation can be roughly divided in those that *decouple* model parameter estimation from image segmentation, and those that interweave them.

In the context of the former category (e.g., [18], [19], [20], [21], [22], [23]), an initial segmentation of the image is provided. Then, superquadrics are used for representing the already segmented image regions. In other words, given the segmentation the superquadric fitting problem is primarily addressed. The models are elaborated with global and local deformations. Hence, these methods demonstrate superior results as far as object representation is concerned: They are able to represent complicated single- and multi- part objects with remarkable accuracy. These approaches are referred to as *segment- then- fit* methods.

Approaches belonging to the latter category, e.g. [24], [25], [26], [27], [28], [29], [16],

[17], [30] address both the segmentation and model fitting problem. Due to the fact that the model is as well used as a means for segmentation, the ability of these methods to represent complicated objects is reduced: They employ undeformed superquadrics, or at the most superquadrics enhanced with simple global deformations for object modeling. Their advantage however, is robustness [17] p.105. These approaches are referred to as *segment-and-fit* methods. They belong to the generic category of model-based image segmentation methods using *probabilistic* techniques: Assuming that the parameters of all models in the input image, as well as their number comprise the parameter set of the segmentation task, the probabilistic techniques estimate the segmentation parameters by maximizing their *posterior* probability in the image. Note, that this is the *optimal* way to perform segmentation, since it guarantees minimization of the *average risk* of failing to select the actual segmentation parameters [31].

Due to the difficulties in directly optimizing the posterior [32], the segmentation task is usually broken down into a parameter initialization stage, and a local optimization stage. The parameter initialization stage is regarded as generation of a *hypothesis* about the values of the segmentation parameters. The optimization stage *refines* the hypothesis, by locally optimizing the posterior distribution. This is the reason why this two-stage segmentation framework will be hereinafter referred to as the *hypothesis generation and refinement* strategy.

The state-of-the-art for superquadric segmentation, the *Recover-And-Select* (**RAS**) paradigm, [33], [16], [17], belongs to the latter category, and adheres to the hypothesis generation and refinement strategy: In the hypothesis generation stage the segmentation parameters are initialized. The hypothesis refinement stage updates the parameters by means of local optimization. More specifically, the hypothesis generation stage involves a grid-like placement of small models in the image. The hypothesis-refinement stage updates the parameters of the initialized models, so that they more accurately express the corresponding data on one hand, and rejects models judged to inaccurately express the data despite parameter updating on the other. RAS is more robust and efficient than other approaches. Its main drawback however, is the misclassification of image points in the neighborhood of object boundaries, because the segmentation process uses the depth values of sets of image pixels only, that is, *region* information, for performing segmentation. Our framework outperforms RAS by fusing *boundary* and region information within segmentation.

B. Our approach- contributions

Our approach comprises two processing stages: hypothesis generation and refinement. In both stages, in addition to region information, *object boundary* information is incorporated, which is acquired via edge detection in the input range image. In the hypothesis generation stage, boundary information is used as a guide to the initialization of the segmentation parameters. In hypothesis refinement a *game theoretic* framework is employed to fuse the information sources in a balanced way.

Our approach considerably outperforms the state-of-the-art in segmentation of superquadrics in terms of speed and accuracy. In addition, we show how a game theoretic framework can be used for segmenting multiple objects in range images. The key to success of our approach lies in three aspects: (i) The consideration of model information from the very early steps of all algorithm stages. (ii) The adjustment of the data provided by each source in order to explicitly represent domain specific information. (iii) The adjustment of the model to represent more accurately the data provided by each information source.

Superquadric segmentation could be applied in a variety of occasions. Representative examples are object categorization and recognition, 3D image representation, image registration and merging, next viewpoint planning guidance, image compression, and others [17]. In this work, our segmentation framework is applied for robotic grasping. We developed a novel robotic system, which automatically unloads piled box-like objects from platforms in real-time. This system combines advantages like flexibility, computational efficiency, accuracy and robustness, the combination of which cannot be found in existing applications.

C. Paper structure

The remainder of this article is organized as follows: In section II, we present some background information related to our modeling elements, which is used by the subsequent sections. In section III we focus on RAS, and point out its problems. In section IV we present our segmentation approach in detail: We present the components of our approach: object boundary information acquisition, the way in which a hypothesis about the values of the segmentation parameters is obtained, and how the hypothesis is refined by fusing information sources. In section V we evaluate our framework experimentally, and we demonstrate its superiority through its quantitative comparison to the Recover-And-Select framework. In section VI we prove the applicability of our framework in an application of high commercial

interest, the robotic unloading of piled objects from platforms. Finally, section VII concludes this work.

II. SUPERQUADRICS AND SOME GEOMETRIC PROPERTIES

The term Superquadrics, was firstly used in [14] to define a family of shapes that includes *superellipsoids*, *superhyperboloids* of one sheet and *superhyperboloids* of two sheets, as well as *supertoroids*. In the computer vision literature however, the term superquadrics is used to refer to superellipsoid objects. In this work, we will also use the term superquadrics as a synonym for superellipsoids.

The explicit form of a superquadric model delivers three- dimensional points on the model surface, and is given by the following equation:

$$\mathbf{S}_{\mathbf{p}}(\eta, \omega) = \begin{bmatrix} x(\mathbf{p}; \eta, \omega) \\ y(\mathbf{p}; \eta, \omega) \\ z(\mathbf{p}; \eta, \omega) \end{bmatrix} = \begin{bmatrix} a_1 \cos(\eta)^{\epsilon_1} \cos(\omega)^{\epsilon_2} \\ a_2 \cos(\eta)^{\epsilon_1} \sin(\omega)^{\epsilon_2} \\ a_3 \sin(\eta)^{\epsilon_1} \end{bmatrix} \quad (1)$$

In the preceding equation, η and ω lie in the range $[-\pi/2, \pi/2]$ and $[-\pi, \pi]$ respectively. The vector $\mathbf{p} = (a_1, a_2, a_3, \epsilon_1, \epsilon_2)$, incorporates the superquadric parameters. More specifically, ϵ_1, ϵ_2 control the shape of the model, and a_1, a_2, a_3 express the size of the model along its $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ axes of the model coordinate system respectively. Fig. 1(a) shows various superquadric models which are rendered using (1). Each of these models corresponds to a different value of the shape parameters, within the range $[0.1, 2]$.

The implicit superquadric equation can be derived by the explicit form of (1), by simple manipulations of the latter:

$$\left(\left(\frac{x}{a_1} \right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{a_2} \right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{a_3} \right)^{\frac{2}{\epsilon_1}} = 1 \quad (2)$$

In order to be able to express the model translation and rotation, six additional pose parameters are incorporated in the parameter vector, which now looks like:

$$\mathbf{p} = (a_1, a_2, a_3, \epsilon_1, \epsilon_2, p_x, p_y, p_z, \phi, \theta, \psi) \quad (3)$$

We adopt the standard approach of [15] to fit a model to the set of three- dimensional points $\mathbf{X}_i, i = 1 \dots N$, according to which the function L_r , shown in (4), which is proportional to the likelihood of the data given the parameter vector and expresses the sum of squares of algebraic distances of the points to the model, is maximized:

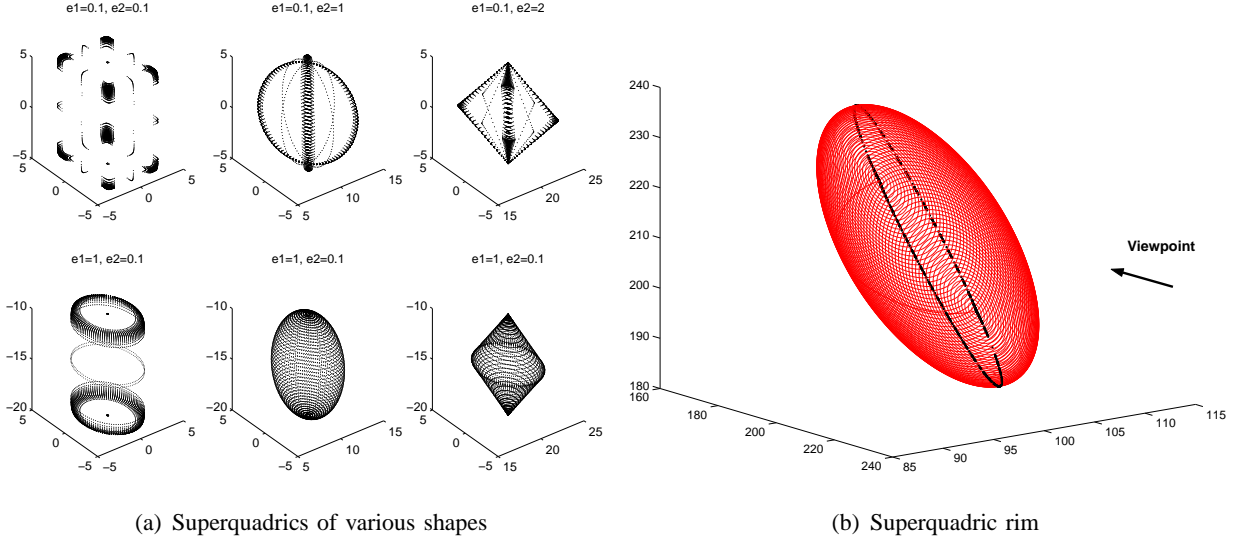


Fig. 1. Properties of Superquadric models

$$L_r(\mathbf{p}) = - \sum_{i=1}^n (\sqrt{a_1 a_2 a_3} (F^{\epsilon_1}(\mathbf{p}; \mathbf{X}_i) - 1))^2, \quad (4)$$

where the function F , equals the left part of (2).

In this work we need to model the boundary of a Superquadric model. The ideal modeling element for this purpose is the superquadric RIM, which is defined as the three- dimensional curve on the model surface, that separates its visible from its invisible part. Figure 1(b) shows a model and its RIM for a particular viewpoint. A detailed discussion about the definition of the RIM and its properties is presented in [34]. In general, given that the \mathbf{Z} axis (or depth axis) of the sensor coordinate system is the viewing point, three- dimensional points on the RIM of the model with parameters \mathbf{p} are derived as follows:

$$\mathbf{R}_p(\omega) = \mathbf{S}_p(\eta(\omega), \omega), \quad (5)$$

where \mathbf{S}_p are points on the superquadric surface obtained via (2), ω ranges in $[-\pi, \pi]$, and $\eta(\omega)$ amounts to:

$$\eta(\omega) = \arctan \left\{ -\frac{a_3}{a_y} \left(\frac{n_y}{a_1} \cos(\omega)^{2-\epsilon_2} + \frac{o_z}{a_2} \sin(\omega)^{2-\epsilon_2} \right)^{\frac{1}{2-\epsilon_1}} \right\} \quad (6)$$

Note, that a uniform (*equal-distance*) sampling of ω in the range $[-\pi, \pi]$ does not produce uniformly sampled points along the RIM of a model. As we will see in section IV-C.1, uniform point sampling along the RIM is required for accurate boundary recovery. Our method for

RIM sampling is based on [35]: equal- distance sampling of points on the RIM is produced using spherical products of proper equal-distance sampled superellipses.

III. RECOVER-AND-SELECT FOR SUPERQUADRIC SEGMENTATION

The Recover-And-Select (RAS) paradigm performs the segmentation task in two stages: Firstly a rough approximation, a hypothesis, about the number and the parameters of the models in the input range image is generated. Secondly, the hypothesis is locally refined using image information.

A. Hypothesis generation

This involves an initial placement of the models to the image, by generating a grid-like subdivision of the input range image in numerous uniform rectangular windows each containing a few range points. The set of points contained in each window is assumed to be a *region* of points corresponding to a superquadric model. Subsequently, the model parameters are determined by fitting a model to each point region. Models with large fitting errors probably cross object boundaries, and thus are rejected from further consideration. The term *seeds* is used to refer to the remaining models, to express both their limited expressive ability, as well as their small size. Hence, the term *seed generation* or *seed placement* alternatively stands for the hypothesis generation stage. The fig. 2(c) depicts seed placement using RAS given the range image of fig.2(b), which was acquired by a laser sensor from the object configuration of fig. 2(a).

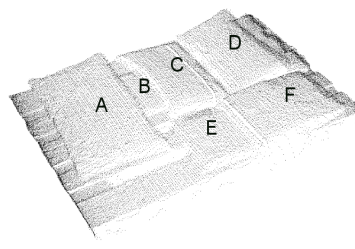
B. Hypothesis refinement

The hypothesis refinement stage comprises 1) seed region growing and 2) model selection.

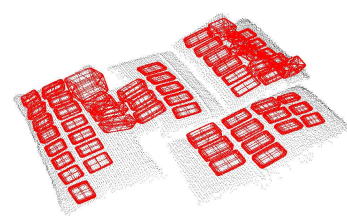
1) *Region growing*: Due to model rejection by the filtering step of the seed generation procedure, image points exist which are not assigned to models. The iterative region growing [36] aims to classify these points to existing regions and to refine the model parameters so that the newly classified points are taken into consideration. Within the region growing process, model parameter estimation through fitting is interwoven with image point classification to models, and therefore in [16], [17] it is referred to as the *classify-and-fit* process.



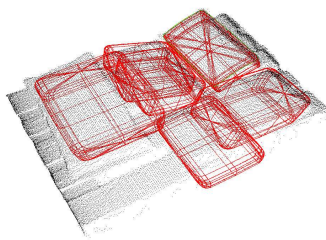
(a) Intensity image



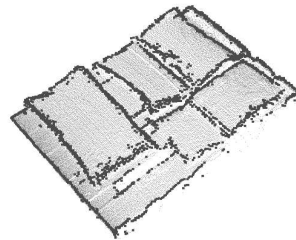
(b) Range image with object numbering



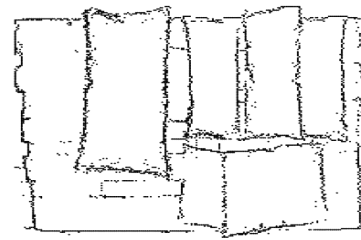
(c) Grid-like seed placement



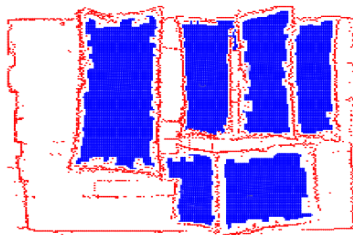
(d) Output of RAS



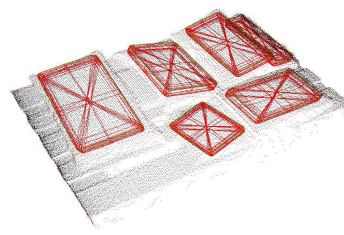
(e) Edge image



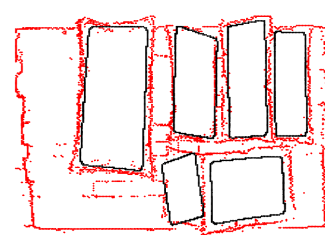
(f) Edge map



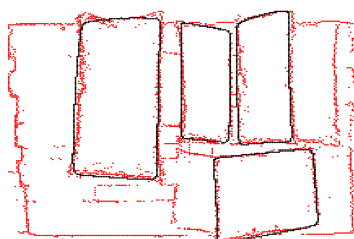
(g) Image regions corresponding to the generated seeds, superimposed on the edge map



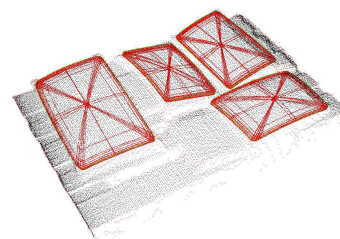
(h) Generated seeds, superimposed to the range image



(i) Seed boundaries on the edge map



(j) Recovered boundaries on the edge map



(k) Recovered models on the range image

Fig. 2. A configuration of piled bags used for qualitatively comparing our approach with RAS

2) *Model selection*: This process selects the models which represent the underlying points accurately. The initialization of many seeds and their independent growing, may cause representation of points by multiple models, which is manifested as model overlapping in the image. Model selection rejects redundant models by applying the *minimum description length* (MDL) criterion: The model configuration in which the fewest possible models describe the largest number of image points with low deviation is preferred. The criterion is realized via a greedy optimization algorithm, which as such is fast but may deliver sub-optimal solutions.

The hypothesis refinement is performed by iteratively invoking 1) followed by 2), until a satisfactory data representation is achieved. Invoking model selection after all seeds are fully grown (*recover-then-select* approach [17]), is inefficient, owing to the big number of initialized models. Computational costs are reduced by *interweaving* 1) and 2). Early invocation of model selection may lead to unreliable data representations and therefore in [17], computational efficiency and reliability are balanced by interrupting the region growing procedure when at least one growing model reaches twice its original size or when there are no models left to grow.

C. Discussion

The RAS paradigm is the state-of-the-art in multiple superquadric segmentation. The iterative region growing process is robust against gross measurement errors, but points belonging to neighboring objects are often erroneously included in the model's region: Due to its inherent locality the region growing process has no mechanism to detect that the model has reached the boundaries of the object it represents (*over-growing*). This problem is *implicitly* addressed within RAS, by placing numerous seeds in the image and by invoking the model selection procedure after region growing, to get rid of the seeds with large deviation from the data, which probably cross object boundaries.

We have extensively evaluated the recover-and-select paradigm and we present here some qualitative results by applying the framework to the range image of fig. 2(b). Fig. 2(d) shows the final result. The region over-growing problem persists, despite the invocations of the model selection process: The recovered models corresponding to the objects *A*, *E* in fig. 2(b) are overgrown: The greedy model selection process does not possess capability to *optimally* arrange the models in a piecewise description. Hence, in many cases either overgrown models are favored, or an overgrown model describing a superset of a data set is selected in addition to the model optimally describing the particular data set, as for the object *C*.

The execution time required for running the paradigm, was about 15 minutes in a Pentium 4 PC, with 1.6 Ghz: Initialization of numerous seeds as a means for addressing the region over-growing problem, in combination with the frequent invocations of the model selection process, renders the framework inappropriate for real time implementation.

To conclude, model overgrowing is the primary cause of problems of the RAS strategy. Our approach *explicitly* addresses region over-growing by incorporating object boundary information in the segmentation process.

IV. INFORMATION FUSION FOR RECOVERING MULTIPLE SUPERQUADRICS

The mechanism guiding the segmentation task in RAS is the region-growing process, which is perhaps the most famous representative of the so called *region-based* approaches for image segmentation. Fusion of region- and *boundary based approaches* are known to deliver superior segmentation results ([37] ch.5 section 2).

While region-based approaches use the input range image I directly for deriving region information, boundary- based approaches use a boundary image I_b for deriving object boundary information. This is usually obtained by applying an edge detector. The problem that arises now is how to integrate the information residing in both I , and I_b in a segmentation framework. In general, there is a different set of segmentation parameters related to each of the information sources. Given \mathbf{p}_r , and \mathbf{p}_b the segmentation parameters related to I and I_b respectively, segmentation of each source separately within a probabilistic framework for image segmentation, involves the computation of the likelihoods $P(I|\mathbf{p}_r)$, and $P(I_b|\mathbf{p}_b)$ [37]. Equal consideration of both information sources requires the computation of the *joint* likelihood $P(I, I_b|\mathbf{p}_r, \mathbf{p}_b)$. Since the parameter spaces related to each source are not related in general, this is not an easy task. For this reason the majority of methods attempting to integrate region- and boundary- based information, ([2], [4], [38]), usually form a giant objective function by constructing a *weighted sum* of the two likelihood functions. Subsequently, they simultaneously estimate both region- and boundary- related parameters by maximizing this function. In this way computation of the joint likelihood is avoided. Such techniques, named *single objective* techniques, have the problem of inadequate analytic and computational tractability. Furthermore, they are of questionable robustness as shown in [37], [39], [40], [41].

The flow diagram of our approach is illustrated in Fig. 3, and comprises two stages: Firstly, edge detection is performed in the input range image I , by means of the scan line

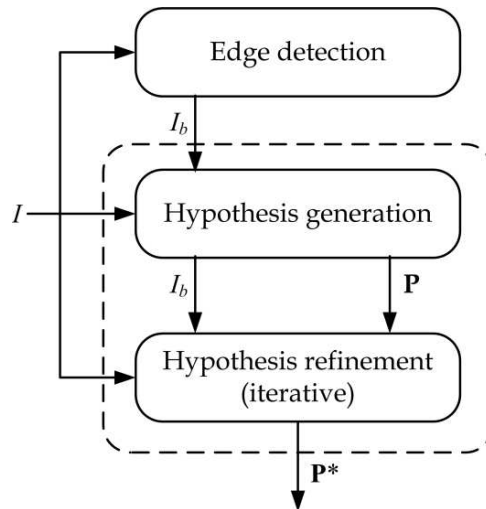


Fig. 3. Our approach for multiple superquadric recovery

approximation technique [42]. Output of the edge detection is the edge map I_b . Secondly, the main part of the segmentation is initiated, bounded by the dashed box in fig. 3. Both I and I_b are employed by this stage. Output of the hypothesis generation process is a set of superquadric models in the range image. If \mathbf{P} denotes a vector containing the parameters of all these models, the hypothesis refinement stage updates \mathbf{P} to obtain the set of models \mathbf{P}^* , the final output of our segmentation process.

In hypothesis refinement, boundary information is integrated in a way inspired by the game theoretic framework, of [40], which is here adapted to processing range images: Parameter refinement is realized by means of iterative invocations of two independent parametric modules in succession. Each module is related to an information source. A benefit function is associated to each module: The region module, fits superquadric models to image regions assumed they belong to unique objects, and associates range points with objects. The boundary module fits the RIM of the superquadric models to the edge points in the boundary image. Hence, each module acts on its own information domain, or in other words, dealing with boundary- related information is *decoupled* from region information and vice versa. Information integration is achieved by simply letting each module pass its corresponding parameters to the other, after its execution. The hypothesis refinement process ends when no significant change in the model parameters can be performed.

Performing integration of region- and boundary-based information within a game-theoretic framework, shows a variety of advantages: Firstly, both information sources are equally

taken into consideration by the segmentation process, without computing the joint likelihood function. Secondly, the benefit of adopting a decoupled scheme for information integration results to computational efficiency, since each module deals with the subset of the entire parameter set, which is related to the corresponding information domain. Finally, robustness is achieved even when the quality of one of the sources is not high, since each module tends to pull the other away from noise and local maxima on one hand, while pushing itself toward the global optimum on the other.

Our approach is described in detail in the paragraphs that follow: We firstly show how boundary detection is performed. Then, we discuss the hypothesis generation and refinement stages. Finally, we point out the innovations of our strategy.

A. Edge detection

We adopt a model-based approach for edge detection in the input range images, which is inspired by [42]. Edge detection is performed via approximation of the rows and the columns of the image with 2D geometric parametric models. In its original form, the technique involves the following steps: Firstly, an input scan line is split into smaller segments, which can be approximated by the modeling elements with a low deviation. A top-down recursive procedure is employed for this purpose [43]. Candidate edge points are the end points of the neighboring segments produced by the splitting process. The parameters of two neighboring segments, are used to compute the *edge strength* of each candidate edge point, which is a proportional to the probability of a candidate edge point being a real edge point. Since two kinds of edges can be found in range images, *jump* edges (discontinuities in depth) and *crease* edges (discontinuities in normals), two strength values are associated to each candidate edge point. The real edge points are the candidates for which one of the two strength values exceeds a user- defined threshold.

We have selected this edge detection approach because it exhibits many advantages: Computational efficiency, which is the outcome of the fast scan line splitting process. Robustness and accuracy in the localization of the edge points, which is the outcome of incorporating global shape information in the process: The classification of a range point as edge point, does not depend on local information, as is the case for the mainstream edge detection approaches for both intensity and range images, but on the parameters of neighboring approximating models, the estimation of which is influenced by a big number of range points. We introduced some additional operations, to further increase the robustness of [42]: a model- based segment-

merging step after the splitting operation, as well as a procedure for fine localization of the edge points. In [44], we show how these operation considerably improve the output of [42] in our data set. Our detector employs the two-dimensional analog of the superquadrics, the superellipses, for approximating the scan-lines of the image for our purposes.

The output of the edge detection operation for the range image of fig. 2(b) is the *edge image*, which comprises three- dimensional edge points. This image, superimposed on the range image fig.2(b), is illustrated in fig. 2(e). Fig. 2(f) illustrates the *edge map* I_b . The edge map is a two- dimensional binary image, a pixel's value $\mathbf{x} = (x, y)$ of which is 1, that is $I_b(\mathbf{x}) = 1$, only if the corresponding range point $I(\mathbf{x}) = (x, y, z)$ is an edge point. In short, the edge map incorporates the orthogonal projections of the boundaries of the objects' visible surfaces on the image plane. Thus, the difference in the information contained in the edge image and the edge map is that the former incorporates the *depth* of the edge points in addition. Since depth information can be derived by the range image, depth information in the edge image is redundant. Therefore, we employ the edge map instead of the edge image for deriving boundary information. In this way we end up dealing with two- instead of three- dimensional boundary information, which, as we will see, considerably increases the computational efficiency of the subsequent boundary related operations, without hampering their robustness.

B. Hypothesis generation

Target of the hypothesis generation (or seed placement) stage is an initial estimation of the segmentation parameters. The most widespread strategies for seed placement [45], [46], [47], [48] determine the sought parameters by model- fitting to image regions corresponding to the objects' interior. For determining these regions the edge map is employed. More specifically, a *distance transform* is applied to the edge map. The contours in the transformed edge map, along which the distance to the objects' edges has a constant value (*iso-distance contours*), enclose image areas in the interior of the objects by definition. The main problem that these approaches face is how to robustly determine the appropriate distance value. This is not easy to do since the objects have different dimensions, and the edge map is usually noisy.

Our approach for seed placement is inspired by [49], where the regions inside the objects are determined by *adaptive closure* of the edge map. The strategy of [49], is based on the observation that any contour in the edge map can be closed by *dilating* the edge map. Contour

closure is performed iteratively, and involves connected component labeling, model fitting to the connected regions, and morphological *dilation* of the regions' boundaries, for which a large fitting error occurred. In [37], we show that using this approach the boundary of each seed region corresponds to the iso-distance contour of the minimum distance from the boundaries in the edge map, whose interior can be satisfactorily modeled by a superquadric entity. Instead of using an iso-distance contour threshold for determining seed regions, we use a threshold for the maximum acceptable fitting error, which is easier to define since it depends on sensor noise and not on the object dimensions in the image.

The seeds created by the application of the adaptive contour closure approach in the edge map of fig. 2(f) are presented in fig. 2(g), 2(h). More specifically, fig. 2(g), shows the region map of the seeds. Fig. 2(h) shows the superquadrics fitted to the range points corresponding to each region in the map. A number of conclusions can be drawn by inspecting these images: Each region corresponds to a different object, and the parameters of the initialized models are sufficiently close to the actual object parameters. Furthermore, the boundaries of the regions are very close to the actual object contours and their shape is similar to the latter. Our seed placement procedure manages to produce an accurate initialization of the segmentation parameters for the particular image.

C. Hypothesis refinement

The hypothesis refinement stage consists of a classify-and-fit process which performs local refinement of model parameters, and of a post- processing step, which rejects models that describe the data inaccurately.

Inputs of the classify-and-fit process are the seed parameter vector \mathbf{p} related to a seed, as well as the region of points associated to the seed. The seed corresponds to an object in the image, which will be hereinafter referred to as the *object of interest* of the process. The goal of the process is to accurately segment the object of interest. This is done iteratively as shown in the flow diagram of fig.4(a).

The process, firstly initializes a 2D image, which partly incorporates region information by identifying pixels corresponding to the object of interest. This image, will be hereinafter referred to as the *region image* and will be denoted as I_r . Its dimensions are equal to those of the range image. A pixel of this image has the value 1, if the corresponding pixel of the range image belongs to the object of interest. Initially, the pixels of I_r set to 1, are the pixels comprising the seed's region. Subsequently, the core of the process starts. Each

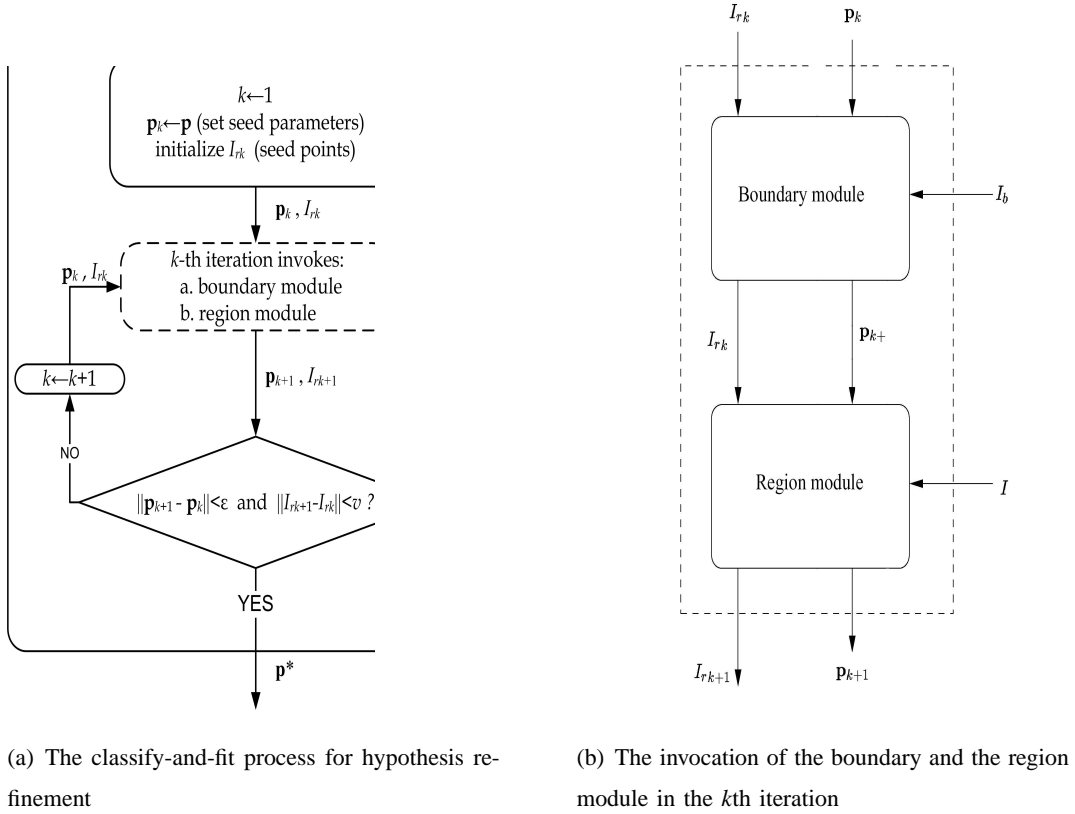


Fig. 4. The iterative hypothesis refinement process, where (b) analyses the dashed process in (a)

iteration, involves a sequential invocation of two modules: **(a)** The *boundary module* employs the superquadric RIM to model object boundary information. This is allowable, since object boundaries separate the visible from the invisible part of the object, just like the RIM does to the model describing the object, given a particular viewpoint. The module updates the seed parameter vector by fitting the RIM of the model to the edge map I_b . **(b)** The *region module* updates the superquadric parameters by fitting the model to the range image I . In addition, it updates the region image I_r . Each module recovers the model parameters by *interweaving* point classification with model parameter estimation. In other words, each of the modules acts as the standard classify-and-fit process on its own information domain.

The process stops after its k th iteration, when two conditions hold simultaneously: The difference between the newly obtained value of the parameter vector \mathbf{p}_{k+1} to the vector obtained by the previous iteration \mathbf{p}_k is smaller than a user-defined threshold ϵ , and the new pixels that the current iteration assigns to the object of interest, or the difference in number of pixels set to 1 between the new region image I_{rk+1} and the one obtained by the previous iteration I_{rk} is smaller than a user defined threshold ν . In this case, the current model vector

is delivered as the output \mathbf{p}^* of the process.

After parameter refinement of all the models in the image has been performed, a post-processing step is invoked to reject objects with large fitting errors to the models, that is objects which can not be satisfactory modeled by superquadric entities. In the following paragraphs we come to a detailed description of the two modules of the classify-and-fit process.

1) *Boundary module: Boundary finding influenced by region information:* The boundary module refines the model parameter vector of the object of interest, using two- dimensional boundary information. Assume that the boundary module is about to be invoked for the k th time. Inputs of the module are the edge map I_b , and the parameter vector \mathbf{p}_k , which is output of the $(k-1)$ th invocation of the module, as well as the region image I_{rk} . Output of the k th invocation of the module is the model parameter vector \mathbf{p}_{k+} (see fig. 4(b)).

The refinement of \mathbf{p} is implemented via maximization of its posterior probability, given both I_b , and I_r . The way this is implemented is inspired by [50], [10]. This is described as:

$$\begin{aligned}\mathbf{p}_{k+} &= \arg \max_{\mathbf{p}_k} L(\mathbf{p}_k, I_b, I_{rk}) \\ &= \arg \max_{\mathbf{p}_k} [\mu_1 L_b(\mathbf{p}_k, I_b) + \mu_2 L_r(\mathbf{p}_k, I_{rk})],\end{aligned}\quad (7)$$

where the user defined constants μ_1 , and μ_2 express the relative significance of the terms L_b , and L_r respectively.

To better illustrate the following, we use the term *eRIM* to refer to the orthogonal projection of the RIM of the evolving model, to the image plane. To clarify, assuming (x, y, z) a three- dimensional point on the RIM obtained by (5), (6), (x, y) will be the corresponding point of the eRIM on the image plane. Now the term L_b of the benefit function L , is proportional to the log-likelihood of the edge map: Maximization of L_b results into fitting the model's eRIM to the edge pixels of the object of interest in the edge map. L_b amounts to:

$$L_b(\mathbf{p}_k, I_b) = \log P(I_b | \mathbf{p}_k) \propto - \sum_{i=1}^M I_{bd}(x(\mathbf{p}_k; \omega_i), y(\mathbf{p}_k; \omega_i))^2 \quad (8)$$

In the preceding expression, I_{bd} is the Euclidean distance-transformed image of the edge map, x, y the image plane coordinates of M three- dimensional points (x, y, z) on the RIM of the model being refined, and $\omega_i = \frac{2i\pi}{M}, i = 0 \dots M-1$. Note that the need for a uniform RIM sampling (discussed in section II) emerges here: The more uniform the sampling is, the

more accurately the likelihood function will express the similarity of the model's eRIM to the object's eRIM in the edge map, and the more accurate the fitting operation will be. The reader is referred to supplementary material 2 for details on the fitting operation.

We now denote as $rRIM$ the boundary of the region in the region image I_{rk} . If so, we define the second term L_r of the benefit function L to be proportional to the log-likelihood of the region image. This is maximized when fitting the model's eRIM to rRIM. Its incorporation in the benefit function integrates region information within the boundary module. In short:

$$L_r(\mathbf{p}_k, I_{rk}) = \log P(I_{rk}|\mathbf{p}_k) \propto - \sum_{i=1}^M I_{rkbd}(x(\mathbf{p}_k; \omega_i), y(\mathbf{p}_k; \omega_i))^2 \quad (9)$$

In the preceding equation, I_{rkbd} is the Euclidean distance-transformed image of the binary image containing the rRIM.

2) *Region module: Region growing influenced by boundary information:* The region module refines the model parameter vector corresponding to the object of interest, using three-dimensional range information. The k th invocation of the region module, succeeds the k th invocation of the boundary module. At this time point, inputs of the module are the range image I , the region image I_{rk} obtained in the previous iteration, and the output of the k th invocation of the boundary module, that is the vector \mathbf{p}_{k+} . Outputs of the module are the \mathbf{p}_{k+1} , and I_{rk+1} (see fig. 4(b)).

The region module is realized by means of the standard region growing process [36]. It performs parameter refinement by the iterative succession of a model fitting step, followed by a pixel classification step. In the model-fitting step it is assumed that the image coordinates of range points corresponding to the object of interest, will be enclosed by the eRIM of the evolving model. In this way boundary information is integrated in the region module. If \mathbf{x}_b the set of the enclosed pixels, then this step fits the model to all 3D points with image coordinates in \mathbf{x}_b . This is equivalent to an image likelihood maximization and is expressed by the following expression:

$$\mathbf{p}_{k+1} = \arg \max_{\mathbf{p}_{k+}} [\log P(I|\mathbf{p}_{k+})] \quad (10)$$

$$= \arg \min_{\mathbf{p}_{k+}} \sum_{\mathbf{x} \in \mathbf{x}_b} \sqrt{a_1 a_2 a_3} (F^{\epsilon_1}(\mathbf{p}_{k+}; I(\mathbf{x})) - 1)^2 \quad (11)$$

The superquadric model with parameters \mathbf{p}_{k+1} generated by the fitting step of the region module, and the region image I_{rk} are used to generate the updated region image I_{rk+1} within the context of the classification step of the region module. More specifically, the pixel classification step examines the set of pixels in the neighborhood of the rRIM. The pixels of

this set which have a small radial euclidean distance from the model \mathbf{p}_{k+1} , are added in the region of I_{rk} to obtain I_{rk+1} .

D. Discussion

The game- theoretic framework for information integration was proposed in [51], and applied to segmentation of intensity images in [40], containing a couple of non occluded objects. The contribution of the approach presented here lies in the application of the framework for *range* images containing multiple piled objects, using *three-dimensional* entities as models. This advances significantly the state-of-the-art, as we will show in section V.

More specifically, the novelty of our system lies in the exploitation of decoupling in model recovery, offered by the game-theoretic framework: **(i)** Although target of both domains, boundary and region, involved in the recovery is the estimation of the model's parameter vector, the model is used only in the region domain in data representation. Using the same model in boundary domain, would complicate the fitting process since the constraint that data correspond to model boundaries should be encoded. Using a curve on the model's surface (the model's RIM) for fitting in boundary domain is advantageous since it corresponds exactly to the boundary data, and depends exactly on the same parameters as the superquadric model. **(ii)** Since depth information is employed in the region domain for model recovery, it does not have again to be taken into consideration in the boundary domain calculations. For this reason fitting in the boundary domain is performed in two dimensions, which considerably accelerates the computational efficiency of all operations without accuracy losses. **(iii)** In cases where additional knowledge about the shape of the target objects is available, (e.g., when the objects are box-like), decoupled handling of boundary and region domains leads into a decomposition in the parameter recovery. This increases the efficiency of our strategy even more. The principle was applied in our robotic system for object unloading with excellent results (see section VI).

V. EXPERIMENTS

We present three sets of experiments to demonstrate the efficiency of our approach. We used images acquired from three different range sensors for this purpose. The goal of the first set is to point out the problems of using region and boundary information sources for object segmentation separately, and to show the superiority of the result when fusing the two information sources. We used a commercial range finder, the SICK LMS-200 [52], which is

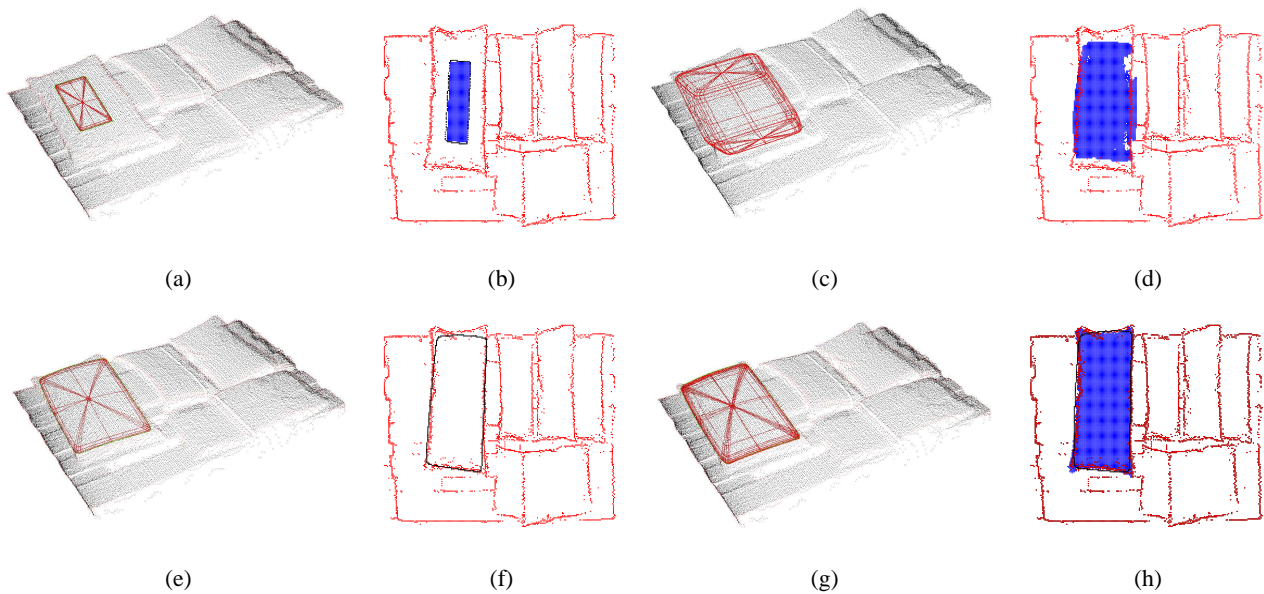


Fig. 5. Recovery results under various fusion modes, as illustrated in section V-A. 5(a), 5(b): seed on 3D and image plane. 5(c), 5(d): region- based recovery result. 5(e), 5(f): boundary- based recovery result. 5(g), 5(h): Result with fusion of region-boundary information sources.

a laser radar based on the time-of-flight measurement technique for acquiring images for this purpose. The acquired images are relatively noisy: The advantages gained by fusing multiple information sources are more clearly illustrated when noisy images are dealt with. The second set compares our approach with the recover-and-select framework, and demonstrates the improvements it introduces with respect to computational efficiency and accuracy. We experimented with images acquired using the prototype structured light range scanner [53], [17]. The same images were used in [17] for assessing the RAS framework. Finally, we performed a third set of experiments in order to illustrate the behavior of our approach when segmenting multi- part objects. The images used here were taken from the popular OSU range image database, and the MSU's Technical Arts 100X range scanner was used for their acquisition. A Pentium 4, 1.6 GHz PC was employed in all experiments.

A. Robustness of information fusion

We demonstrate the advantages gained when fusing multiple sources of information in image segmentation, using the object configuration shown in fig. 2(a). In particular, we dealt with recovering one object in the configuration, starting from the small seed shown in fig. 5(a), 5(b). More specifically, fig. 5(a) depicts the seed superimposed on the range image. Fig. 5(b) shows the edge map along with the eRIM of the seed, as well as the region associated to

the seed.

We firstly tried to recover the object using a region-based only approach, namely the standard region- growing process. The result is presented in fig. 5(c), 5(d). The former figure depicts the recovered model in three dimensions, while the latter shows the region of points associated to the model, superimposed on the edge map. Note, that the seed has been placed inside the object of interest and in a relatively big distance from its boundaries. Seed placement in this position increases the robustness of the region growing approach [17] p.157. Despite this, the result of region segmentation is discouraging: Pixels belonging to neighboring points have been erroneously included in the region of the evolving model. The model is over- grown. The process needed 30 iterations to produce the result shown.

We then attempted object recovery using our approach. In the first case, model recovery is driven by boundary information only. Hence, we used $\mu_1 = 1$, and $\mu_2 = 0$ in the benefit function of the boundary module (see (7)). The result of boundary- guided recovery is illustrated in fig. 5(e), 5(f). The problem of using boundary information only for object segmentation, is that noisy edge points on the exposed surface of the object of interest attract the eRIM curve, so that the model does not grow as it should. Hence, in comparison to the region growing- based segmentation, the boundary- based segmentation suffers from the inverse problem: model under-growing. Note, that in this case the system executed 5 iterations in order to converge.

Finally, we considered that both boundary and region based information guide the recovery of the object of interest, that is we set $\mu_1 = \mu_2 = 0.5$ in (7). The result is illustrated in fig. 5(g), 5(h), of superior accuracy compared to the previous ones, and this is due to the combination of a region *force* which is responsible for continued model growing along noisy edge points, with the boundary force which prohibits object growing beyond real boundaries. The recovery results for all objects in fig. 2(a) is shown in figures 2(j), 2(k). In this case the system converged after 12 iterations. There is a considerable improvement in terms of robustness and accuracy with respect to the results delivered by recover-and-select (fig. 2(d)). In general, the values of the weights μ_1 , μ_2 , depend on a-priori knowledge on the relative quality of the boundary and region information respectively. High quality boundary information suggests a big value for μ_1 . Similarly, high quality region information suggests a big value for μ_2 . In the experiments with all our data sets, $\mu_1 = \mu_2 = 0.5$ produced substantially better results.

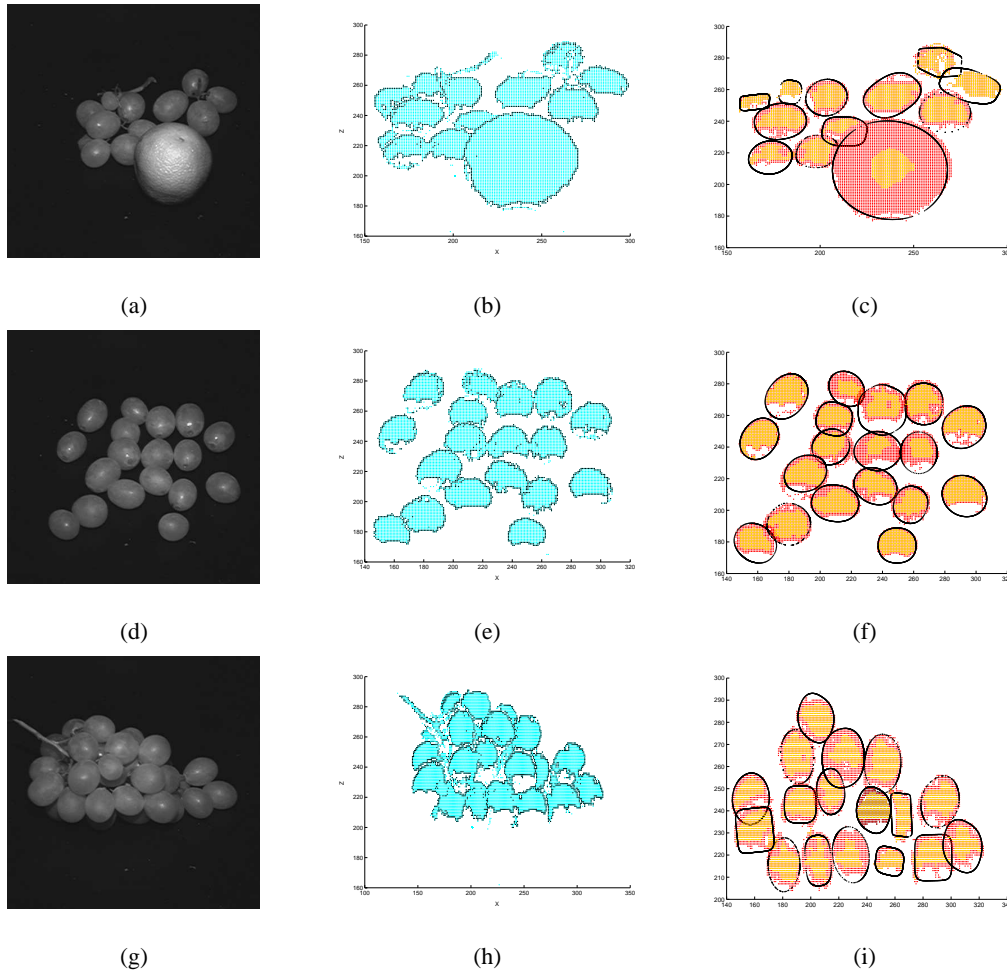


Fig. 6. Configurations of rounded superquadrics. In a, d, g are displayed the intensity images; in b, e, h the edge map (dark points) superimposed to the range image; in c, f, i show results on the image plane: the recovered eRIMs, the regions corresponding to the seeds (light areas) and the regions added to the seed regions after object recovery (dark areas).

B. Comparison to RAS

We tested our approach on the same images on which RAS was evaluated in [17]. The results are shown in figures 6(a)-6(i), 7(a)- 7(o).

We performed a comparison of our approach with the RAS paradigm in terms of accuracy. This comparison is based on a *ground truth* segmentation of the images: For each range image we *manually* segmented the range points corresponding to each object. We denote the set of points belonging to an object i in the figure as \hat{R}_i . Now, given a recovered model with parameters \mathbf{p}_i which corresponds to the object i , the quantity d_{3D} amounts to the average radial Euclidean distance of the points in \hat{R}_i to the model \mathbf{p}_i , and expresses a measure of recovery accuracy in three dimensions. That is:

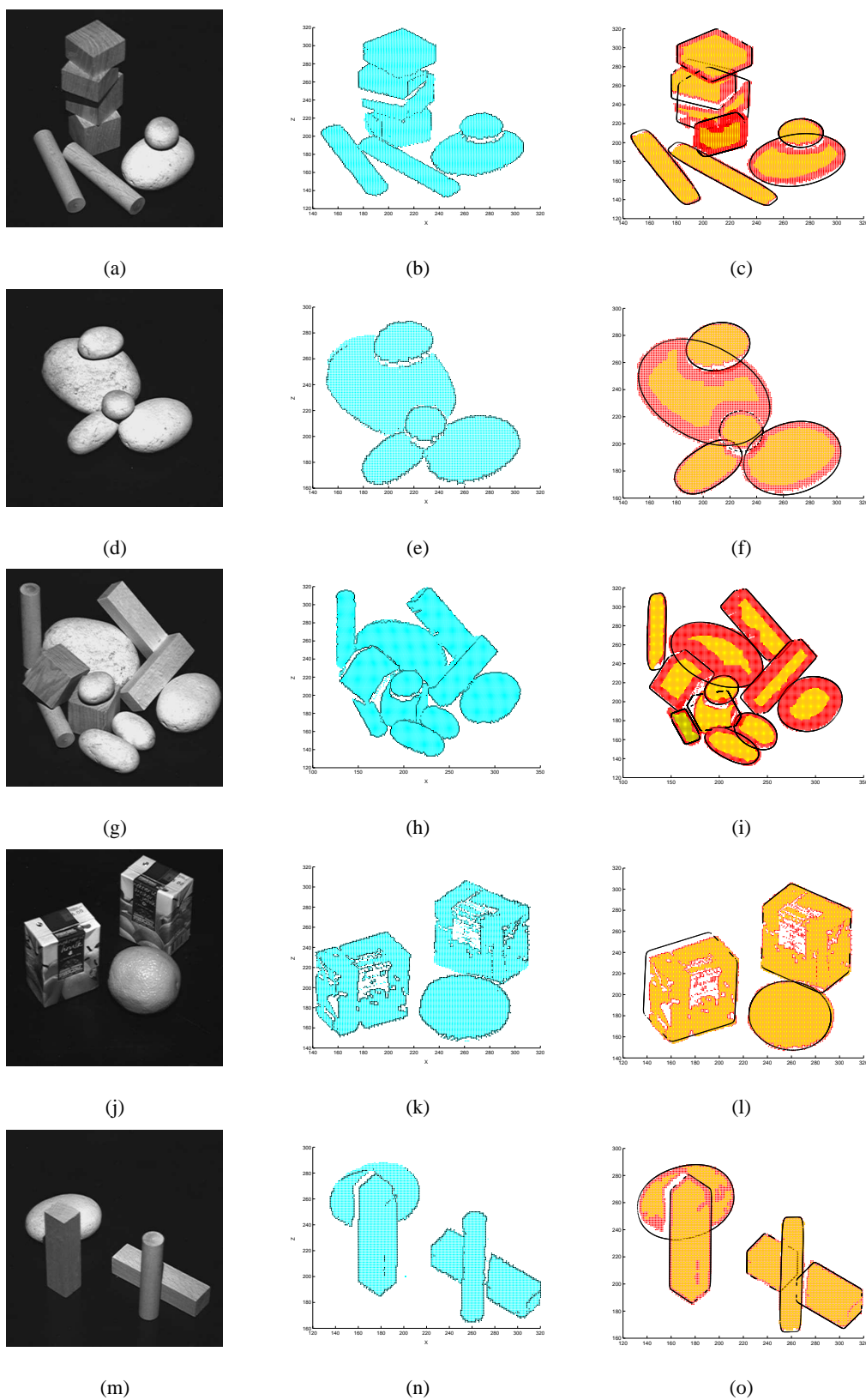


Fig. 7. Configurations of various superquadrics. In a, d, g, j, m are displayed the intensity images; in b, e, h, k, n, the edge map (dark points) superimposed to the range image; in c, f, i, l, o show results on the image plane: the recovered eRIMs, the regions corresponding to the seeds (light areas) and the regions added to the seed regions after object recovery (dark areas).

$$d_{3Di}(\hat{\mathbf{R}}_i, \mathbf{p}_i) = \frac{1}{|\hat{\mathbf{R}}_i|} \sum_{\mathbf{x} \in \hat{\mathbf{R}}_i} d(\mathbf{x}, \mathbf{p}_i) \quad (12)$$

In addition, we assessed the reconstruction of the objects' boundary. We regarded the image plane coordinates of all points in $\hat{\mathbf{R}}_i$ and then extracted the boundary $\hat{\mathbf{S}}_i$ of this region. This two-dimensional curve corresponds to the eRIM of the object i . The metric d_{2D} measures the average distance between the eRIM of the recovered object \mathbf{p}_i (denoted as $\mathbf{C}_{\mathbf{p}_i}$), to the points on $\hat{\mathbf{S}}_i$ and expresses the accuracy in the recovery of the objects boundary that is:

$$d_{2Di}(\hat{\mathbf{S}}_i, \mathbf{C}_{\mathbf{p}_i}) = \frac{1}{|\hat{\mathbf{S}}_i|} \sum_{\mathbf{x} \in \hat{\mathbf{S}}_i} d(\mathbf{x}, \mathbf{C}_{\mathbf{p}_i}) \quad (13)$$

The accuracy results of both our approach and RAS, are illustrated in table V-B. For each examined image, the average values of d_{2D} and d_{3D} for all the objects in the image are denoted as \bar{d}_{2D} , and \bar{d}_{3D} respectively. Given this, the second and third column of table V-B, show the \bar{d}_{3D} , \bar{d}_{2D} , when our approach is used for segmentation. The fourth and fifth column illustrate the corresponding values when RAS is used. For every image, our approach performs better than RAS. The last row of the table shows the averaged results along all the rows. The average \bar{d}_{3D} for all images examined is 0.612 using our approach and 1.196 using RAS. Given this metric our approach delivered about two times more accurate results. The corresponding results for \bar{d}_{2D} is 0.747 and 2.422 respectively. As far as this metric is concerned, our approach introduces an improvement of a factor around 3. Worth-noticing is, that from 79 objects examined, RAS delivered better reconstruction accuracy results in only 6 cases. The primary reason for the failure of our approach in these cases, is the low quality of both region and boundary based information.

Moreover, we compared our approach to RAS, with respect to computational efficiency. As discussed, our approach recovers each object independently from the others. In this way it offers the potential of parallelization, if within a parallel processing framework, each object is recovered by one processor. Given 8 images containing 79 objects in total (see figures 6(a)-7(o)), we measured the average time needed for the recovery of a unique object. This was found to be 25.29 seconds per object. RAS cannot be parallelized. The corresponding time RAS needed for object recovery was about 300 seconds. In this respect our approach introduces an improvement of a factor of about 12 regarding efficiency. Note, that our approach was implemented in MATLAB, while RAS in C. An improvement on the efficiency

Image id	\bar{d}_{3D}	\bar{d}_{2D}	$\bar{d}_{3D}(RAS)$	$\bar{d}_{2D}(RAS)$
fig. 7(a)	0.867	0.431	1.424	2.739
fig. 7(d)	0.363	1.009	0.881	1.705
fig. 7(g)	0.822	1.153	1.318	3.822
fig. 7(j)	0.752	0.199	0.930	0.644
fig. 7(m)	0.449	0.093	0.777	0.364
fig. 6(a)	0.672	1.090	1.038	2.602
fig. 6(d)	0.378	0.363	1.251	2.726
fig. 6(g)	0.593	1.642	1.947	4.773
average	0.612	0.747	1.196	2.422

TABLE I

ACCURACY OF RECONSTRUCTION IN 3D AND 2D BETWEEN OUR APPROACH AND RAS

of our method is expected to arise if our approach is ported to C. The reader is referred to supplementary material 3 for a more detailed analysis of these experiments.

C. Experiments with multi- part objects

We performed experiments with multi- part objects. Two range images of multi- part objects from the OSU (MSU/WSU) range image database were used for this purpose. The images along with the recovery results are shown in figures 8(a)- 8(f). Not all objects in those images can be well modeled by superquadrics. Despite this, in the majority of the cases our model-based edge detector is able to capture the intersections between the parts, and to provide relatively accurate boundary information. The adaptive boundary closure operation delivers a satisfactory seed placement. Finally the hypothesis refinement framework delivers in the majority of these images encouraging segmentation results. Note, that the triangular part of the planar object in the first figure set cannot be expressed via the model. The same is the case for the left part of the planar object of the second figure set. Results on the recovery of the objects for these figures have been as well published in [26]. Despite the problems of our approach in some cases, the overall results are considerably better than in [26].

Furthermore, we performed an experiment with a range image of a wooden doll. The image appears in [17], p. 171. The results appear in figures 9(a)- 9(d). All parts of the doll have been recovered. Again, the result shown here is better than the result delivered by RAS: The upper parts of the left and right hands of the doll, that is, the parts connected to the torso

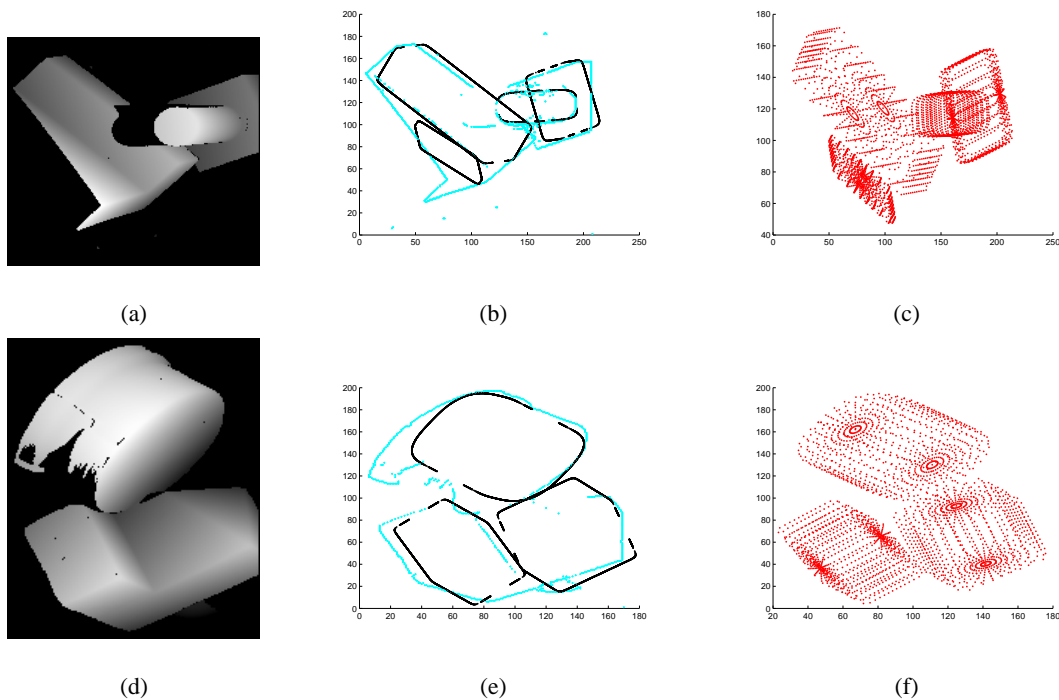


Fig. 8. Recovery results on multi- part objects. In a, d, are displayed the intensity images; in b, e, the recovered RIMs (dark points) superimposed to the edge image (light points); Finally, in c, f we show the recovered three dimensional models. The recovery results are not satisfactory in some cases. The reason is that some object parts cannot be modeled with superquadric models. The reader is referred to section V-C for details.

are represented in [17] with two superquadric models. In our case one model expresses each part, as it should. Note, that in this case some object parts are not perfectly modeled with superquadrics. The difference of this parts to superquadric models is not big. The majority of the boundary points are detected by the edge detector, and the obtained segmentation is reasonable.

We performed a final experiment in order to quantify the behavior of our algorithm against occlusion. In order to perform viewpoint invariant experiments, we considered a spherical superquadric with radius 30mm, that is: $\epsilon_1 = \epsilon_2 = 1$ and $\alpha_1 = \alpha_2 = \alpha_3 = 30$. We then created range points on the surface of the model, and embedded the points on its hemisphere in a synthetic range image, with inter pixel distance of 1 mm. The RIM of this model is a circle, on which we considered $M = 100$ points for our experiments. In this experiment we varied independently the amount of bounding contour occlusion and the amount of region occlusion.

When the missing percentage of bounding contour information is lower than 55 per-cent,

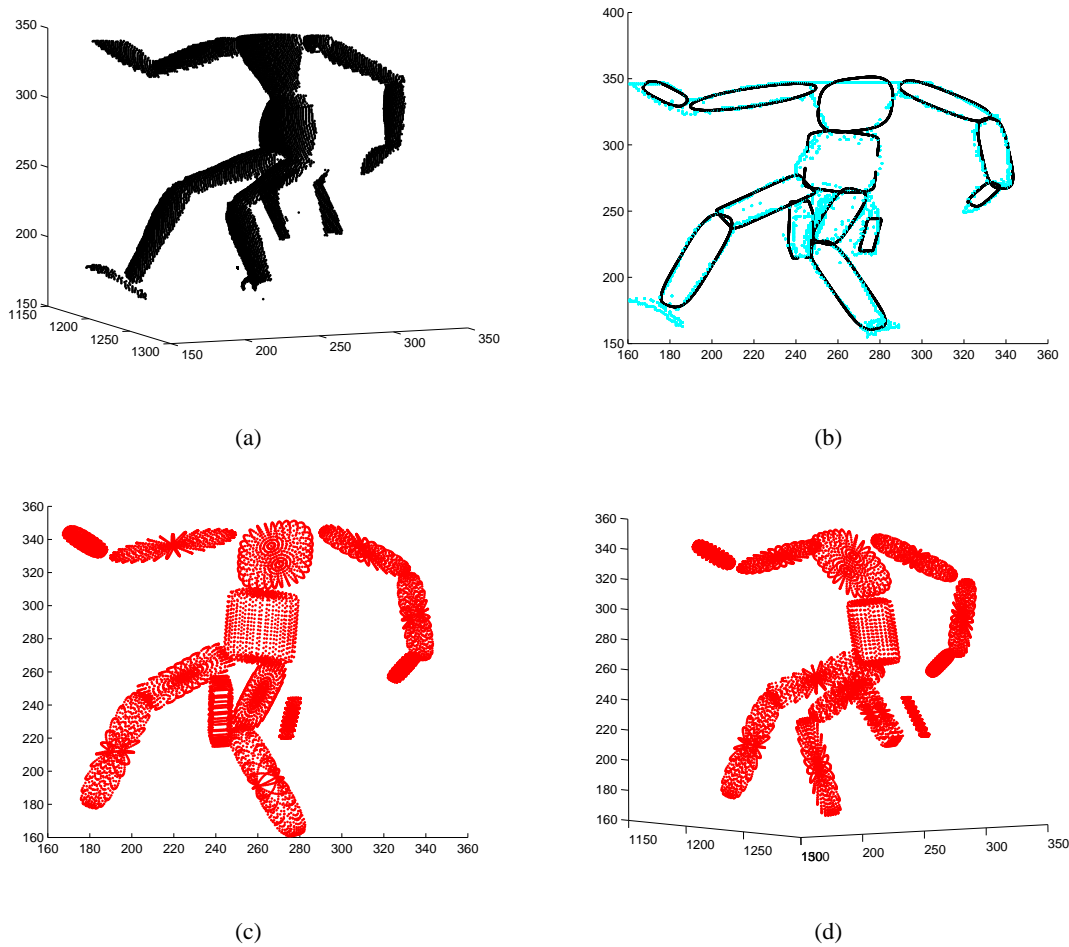


Fig. 9. We performed experiments with the image of a wooden mannequin, originally presented in [17], p. 171. a, shows the range image. b the recovered RIMS (dark points) superimposed to the edge image (light points). c shows the recovered models in three dimensions. Finally, e shows an additional three- dimensional view of the recovered superquadrics.

our approach manages to recover the object of interest with an accuracy of 0.5 pixels for d_{2D} (see (12)) and 1 mm for d_{3D} (13). For missing boundary information in the range 40-55 per-cent the size of the seed generated by the hypothesis generation stage is small. Despite this, the region module gradually increases the size of the seed due to the incorporated region growing operation. Due to the fact that no noisy boundary points inhibit the growth of the model, the model manages to reach its actual size. For more than 55 per-cent missing boundary information, the generated seed is very small, due to the multiple invocations of the boundary closure operation. In this case less than 10 range points are assigned to the initialized model. Hence, the superquadric fitting function delivers very unreliable results and the system breaks down.

In addition we performed experiments by varying the amount of region information. For

missing region information up to 60 per-cent our approach delivers a satisfactory result. The accuracy in recovering boundary information expressed by d_{2D} is up to 3.5 pixels. The accuracy in recovering region information expressed by d_{3D} is up to 5 mm. For missing region information in the range 70 – 90 per-cent, the size of the initialized seed is small. The object's boundary points are far away from the seed, so their attraction forces they exert to the seed are low. In addition, there is no available region information which will be used by the region module to grow the seed. As a result the result of our system deteriorates considerably. Note, that for more than 90 per-cent missing region information, our system breaks down, since the number of range points are not enough to produce a reliable model via the model fitting operation.

Finally, we measured the reconstruction accuracy of the visible surface of the model as well. In this case for missing boundary information up to 70 per cent, the recovery accuracy values were less than 5 pixels for d_{2D} and 7 mm for d_{3D} . For missing region information up to 78 per-cent, the recovery accuracy values were less than 4 pixels for d_{2D} and 9 mm for d_{3D} . The overall results shown in this section are encouraging and they demonstrate that our system is robust against occlusion.

VI. APPLICATION: ROBOTIC UNLOADING OF PILED BOX-LIKE OBJECTS

We applied our segmentation framework to a novel robotic system (a description of which appears also in [54], [37], [55]), for automatic unloading of piled box-like objects, as those in figures fig. 2(a), as well as in figures VI-14 . The driving force for the implementation of the particular system, is mainly the industrial requirement for the reduction of costs in the logistic processes.

In the context of our system a six degrees of freedom industrial robotic manipulator is employed for object unloading. A laser sensor (*SICK* LMS-200), is mounted on the hand of the robot for acquiring images of the top side of the pallet. Besides, a vacuum gripper is as well mounted on the hand of the robot for object grasping and placement. Our system operates as follows: Firstly, a range image of the object configuration is acquired. This is done by a linear movement of the robotic hand across the two ends of the pallet. Secondly, the image is analyzed and a number of upper objects on the pile are recovered. Thirdly, the recovered objects are grasped by the robotic hand and placed at positions defined by the user. This process continues iteratively, until no objects lie on the pallet. Fig. 10 illustrates instances of the unloading process.

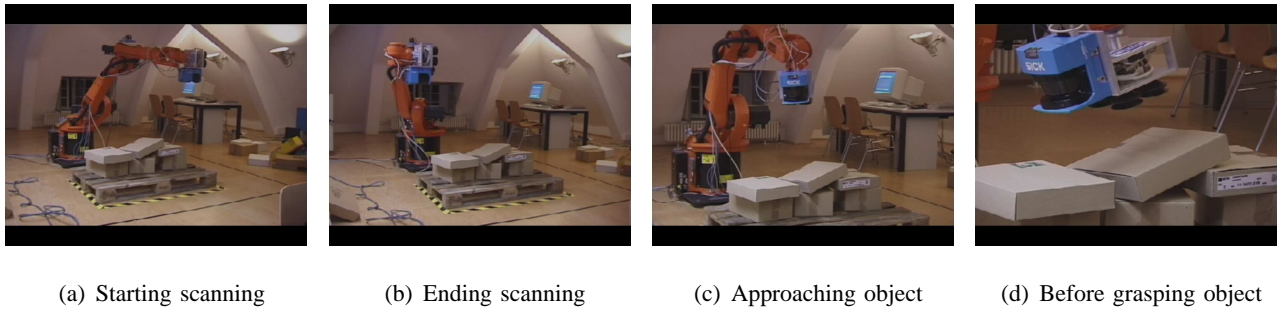


Fig. 10. Our robotic system in operation: The start and end positions of the scanning process are shown in 10(a) and 10(b), respectively. 10(c), shows the robotic hand while approaching an object. Finally, 10(d) illustrates the robot, just before grasping the object.

Given the input range image, our superquadric recovery framework is applied to determine the parameters of the graspable objects in the pile. The framework is adapted to the application as follows: Firstly, box-like superquadrics are only used for modeling. This means that the shape ϵ_1, ϵ_2 parameters of the models are set in the value 0.2 and are kept constant within the recovery process. Secondly, four additional parameters expressing global parabolic deformations are introduced in the parameter vector of (3). These allow for modeling of the slight occurring on some non-rigid objects (e.g., bags or pillows). Furthermore, a decomposition in the recovery of the parameter vector of the model is realized: According to our framework, superquadric parameter recovery is performed by two modules each dealing with its own domain of information. However, given the box-like geometry of the objects, not all the superquadric parameters are related to the domains on which each of the two modules acts. Decomposing the recovery of the model parameter vector results into a search for the optimal parameters in lower dimensional parameter spaces. As a consequence both the robustness and computational efficiency of the recovery process increase. The reader is referred to supplementary material 1 for details on this issue.

Finally, a post-processing step is included in the framework. The process succeeds the hypothesis refinement stage, and determines if the refined model corresponds to a *graspable* object. A graspable object is by definition an object which is not occluded by another object. This implies that edge points along the entire RIM of the object are captured by the boundary detection process, and are thus present in the boundary image. Therefore, for the graspable objects in the image, the fitting error residual of the model's RIM is expected to be low. In addition, since such objects can be well modeled by superquadrics, the fitting error residual of a superquadric model to the range points of the image belonging to the object is as well

Avg. Distance measures	d_{3D} (mm)	d_{2D} (pixels)
Values	8.68	1.01

Measures	Values
Graspable objects	174
True positives	159
False positives	6
False negatives	15

TABLE II

ACCURACY AND ROBUSTNESS OF ROBOTIC OBJECT UNLOADING SYSTEM

Modules	Avg. Duration (sec)
Hypothesis generation per object	3.13
Hypothesis refinement per object	15.5
Hypothesis refinement per scene	91.75

TABLE III

EFFICIENCY OF ROBOTIC OBJECT UNLOADING SYSTEM

expected to be low. In short, the post processing method examines both model fitting error residuals to the range image and the boundary image, and decides that the object of interest is graspable if both are smaller than user-defined thresholds. The model parameters determined to correspond to graspable objects can now be forwarded to a robotic system for performing grasping operations.

We tested the performance of our system on about 40 box-like object configurations. Representative results are shown in fig.11-15. The first image of each row shows an object configuration. The second, the range image on which the edge image is superimposed. The edge map with the recovered object boundaries (in bold), follow. The last image shows the recovered models in 3D, superimposed on the range image. Note than in some occasions, for example in figures 12, 13, 14, there are objects to which no model corresponds. These objects are determined to be non-graspable by the post processing step, and the corresponding models are rejected. The same holds for the recovery output of fig. 2(k).

Accuracy results are shown in table II. The table shows averaged results along all images for the metrics d_{2D} , d_{3D} , extracted as in section V. Note, that the recovery accuracy in 3D is about 9mm, that is in the order of the sensor's resolution, which is about 10mm. As well, we achieved almost sub-pixel reconstruction accuracy in 2D.

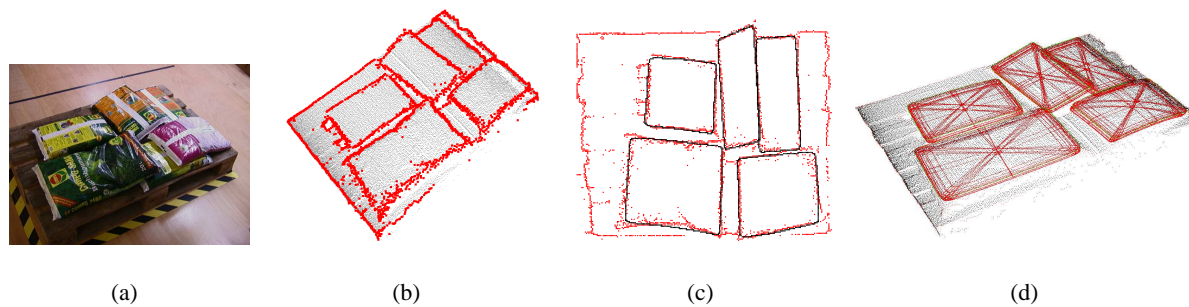


Fig. 11. Bags configuration

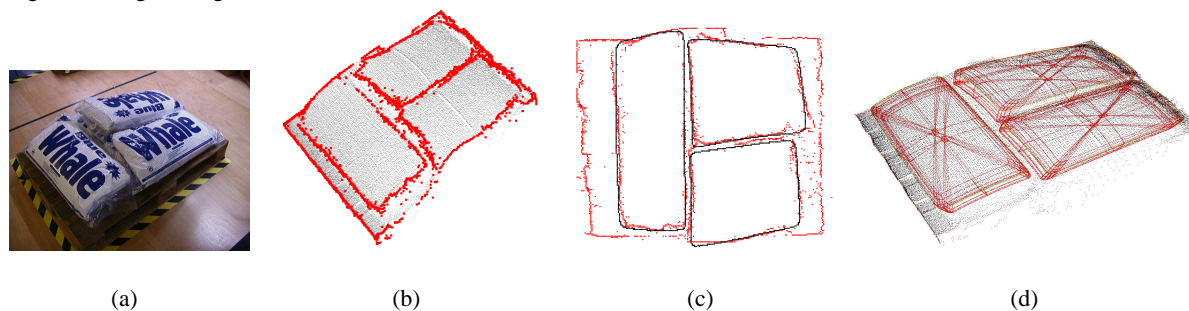


Fig. 12. Pillows configuration

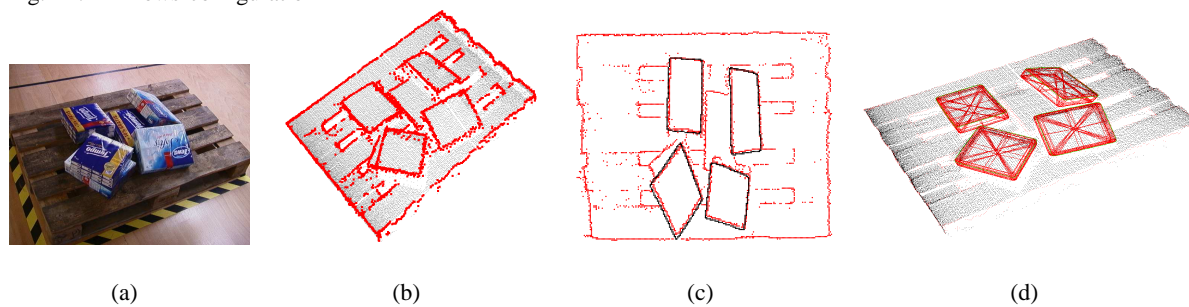


Fig. 13. Box-like object configuration

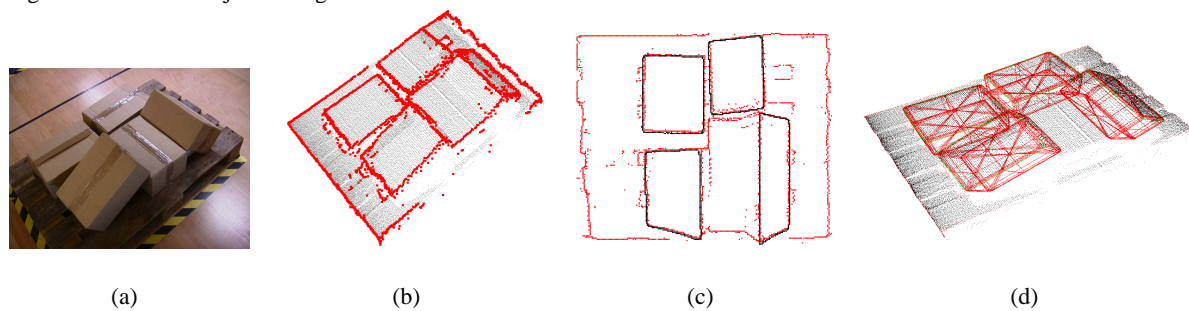


Fig. 14. Cardboard boxes configuration

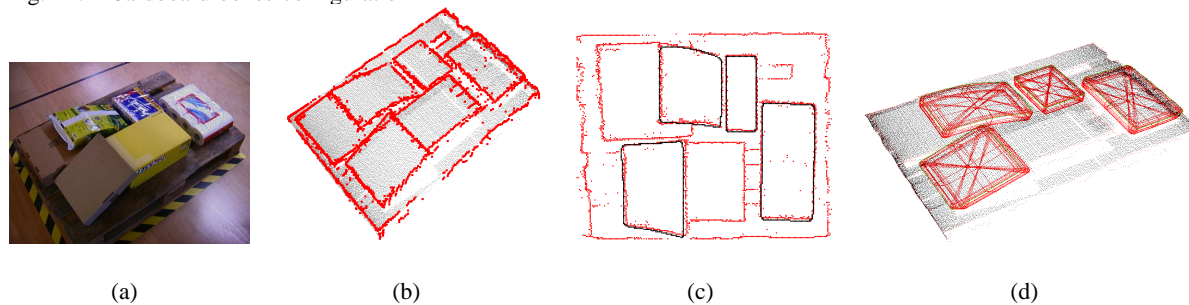


Fig. 15. Configuration of various types of objects

We assessed the robustness of our system by comparing the number of graspable objects in all box-like object configurations against the one determined by a human. Results are shown in table III: In the 40 examined images, 174 objects exist in total, 159 of which have been successfully recovered. That is, the percentage of true positive responses was more than 91 per-cent. Note, that the system gave false negative responses in only 6 cases. Note as well, that using the Recover-and-Select framework one is not in the position to decide on the graspability of objects. This feature is offered by our approach due to the consideration of boundary information.

Finally, our system is fast: The average object recovery time was 15.5sec (hypothesis refinement) plus 3.13 sec (hypothesis generation), that is about 18 seconds in a Pentium 4 PC, with 1.6Ghz. The combination of the advantages of our system, that is, flexibility, robustness, accuracy, computational efficiency, low-cost, cannot be found in any other existing work, to the best of our knowledge [37].

VII. CONCLUSIONS

We presented a strategy for multiple superquadric segmentation in range images. The segmentation task is performed within two stages. Firstly, a hypothesis about the number of objects in the image and the parameters of the models representing them is generated. Secondly this hypothesis is refined by means of iterative fitting of the models to the data. Within our approach two aspects of information a range image can provide are considered: region and object boundary information. The former is provided directly by the input range image. The latter by a two-dimensional edge map, obtained via edge detection on the range image. Hypothesis generation is realized via adaptive closure of the edge map, which indicates range point regions belonging to the object's interior, on which seed models are fitted. The hypothesis refinement stage improves the result delivered by the latter stage. A game-theoretic framework is applied, to perform refinement while simultaneously fusing the two information sources. This is done via iterative invocations of two modules in succession: The region module, refines the model parameters by model fitting in the range image. The boundary module fits the model's boundary to the edge map. This takes place until the model parameters do not change significantly.

We demonstrated experimentally, that our strategy outperforms the Recover-And-Select (RAS) framework [16], [17], the state-of-the-art in multiple superquadric segmentation, in terms of accuracy and computational efficiency. More specifically, our approach reduced

about 2 times the error metric for 3D model reconstruction and about 3 times the error metric for reconstructing the boundaries of the objects in the two-dimensional image plane. Additionally, it proved to be more than 12 times faster than RAS.

The considerable improvements in accuracy introduced by our approach, are due to the consideration of boundary information within segmentation, in addition to region information. In detail, using boundary information in the hypothesis generation stage produces a set of segmentation parameters which are much closer to the actual parameters than those delivered by the corresponding stage of RAS. Furthermore, consideration of boundary information within hypothesis refinement, solves the region over-growing problem, which is the main source of low accuracy within RAS. Finally, high accuracy is the outcome of incorporation of model information from the early steps and during all stages of the process.

Regarding computational efficiency, consideration of boundary information in the hypothesis generation process, results into a smaller number of initialized seeds, and thus in lower overall processing time. Besides, using boundary information for parameter recovery accelerates the growing of the models: Except for the standard region growing, models now grow as a result of the attraction forces exerted by the boundary points. In the beginning of the growing operation, when the model is small, these forces are responsible for fast growing of the models toward the real object boundaries. Note finally that recovery in all domains is performed using parametric models of closed form which considerably facilitates the fitting processes.

The novelty of our approach lies in the adjustment of a game-theoretic framework of [51], [40] to the segmentation of range images. More specifically, we exploit the feature of decoupling model recovery between the two information domains so as to enhance the descriptive ability of the data, and ease the model fitting process. The former is due to letting boundary operations occur in two- instead of three- dimensions. The latter is due to using the model's boundary curve instead of the model itself for model fitting in the boundary domain.

Last but not least, we built a novel robotic system for automatic piled box-like object unloading. The system is based on our segmentation framework. A detailed experimental evaluation of the system showed that it is accurate, robust and fast. We showed that the average time for the recovery of an objects was 15.5 seconds on the target platform. Additional knowledge on the shape of the objects in the configuration, e.g., knowledge that the configuration consists of rigid boxes only for example, can further reduce the execution time of our system to 5 seconds per object. Given that the robotic grasping and placement

operations do not last more than 4 seconds for an industrial robot in its full speed, our system has the potential to perform unloading operations faster than human workers. The advantages exhibited by this system, cannot be found in any other application of this kind, up to our knowledge.

ACKNOWLEDGMENTS

We would like to deeply thank Ales Jaklic, and Ales Leonardis, university of Ljubljana, Slovenia for sharing their code and the images they used in RAS evaluation, as well as for very useful discussions.

REFERENCES

- [1] H. Tek and B. Kimia, "Image Segmentation by Reaction-Diffusion Bubbles," in *IEEE International Conference in Computer Vision*, 1995, pp. 156–162.
- [2] S. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.
- [3] S. Sclaroff and L. Liu, "Deformable Shape Detection and Description via Model-Based Region Grouping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 5, pp. 475–489, 2001.
- [4] N. Paragios and R. Deriche, "Coupled Geodesic Active Regions for Image Segmentation: A Level Set Approach," in *European Conference on Computer Vision*, Vol. 2, 2002, pp. 224–240.
- [5] R. Nevatia and T. Binford, "Description and Recognition of Curved Objects," *Artif. Intell.*, vol. 8, no. 1, pp. 77–98, 1977.
- [6] D. Keren, D. Cooper, and J. Subrahmonia, "Describing Complicated Objects by Implicit Polynomials," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 38–53, 1994.
- [7] I. Biederman, "Human Image Understanding: Recent Research and a Theory," *Computer Vision Graphics and Image Processing*, vol. 32, pp. 29–73, 1985.
- [8] K. Wu and M. Levine, "3-D Shape Approximation using Parametric Geons," *Image Vision Comput.*, vol. 15, no. 2, pp. 143–158, 1997.
- [9] L. Staib and J. Duncan, "Model Based Deformable Surface Finding for Medical Images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 720–731, 1996.
- [10] —, "Boundary Finding with Parametrically Deformable Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061–1075, Nov. 1992.
- [11] D. Terzopoulos, A. Witkin, and M. Kass, "Symmetry-Seeking Models and 3D Object Reconstruction," in *IEEE International Conference on Computer Vision*, 1987, pp. 269–276.
- [12] S. Muraki, "Volumetric Shape Description of Range Data using Blobby Model," *Computer Graphics(SIGGRAPH 91)*, vol. 25, no. 4, pp. 227–235, July 1991.
- [13] J. Hanson, "Hyperquadrics: Smoothly Deformable Shapes With Convex Polyhedral Bounds," *Computer Vision, Graphics, and Image Processing*, vol. 44, pp. 191–210, 1988.
- [14] A. Barr, "Superquadrics and angle preserving transformations," *IEEE Computer graphics and Applications*, vol. 1, no. 1, pp. 11–23, 1981.

- [15] F. Solina and R. Bajcsy, "Recovery of Parametric Models from Range Images: The case for Superquadrics with Global Deformations," *IEEE T-PAMI*, vol. 12, no. 2, pp. 131–147, 1990.
- [16] A. Leonardis, A. Jaklic, and F. Solina, "Superquadrics for Segmenting and Modeling Range data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1289–1295, 1997.
- [17] A. Jaklič, A. Leonardis, and F. Solina, *Segmentation and recovery of Superquadrics*, ser. Computational Imaging and Vision. Dordrecht: Kluwer Academic Publishers, 2000, vol. 20.
- [18] F. Ferrie, J. Lagarde, and P. Whaite, "Darboux Frames, Snakes, and Super-Quadrics: Geometry from the Bottom Up," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 8, pp. 771–784, 1993.
- [19] D. Metaxas and S. Dickinson, "Integration of Quantitative and Qualitative Techniques for Deformable Model Fitting from Orthographic, Perspective, and Stereo Projections," in *IEEE International Conference on Computer Vision*, 1993, pp. 641–649.
- [20] D. Terzopoulos and D. Metaxas, "Dynamic 3D Models with Local and Global Deformations: Deformable Superquadrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 703–714, 1991.
- [21] N. Raja and A. Jain, "Recognizing Geons from Superquadrics Fitted to Range Data," *Image and Vision Computing*, vol. 10, no. 3, pp. 179–190, April 1992.
- [22] A. Pentland and S. Sclaroff, "Closed-form solutions for physically based shape modeling and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 715–729, 1991.
- [23] A. Gupta, G. Lea, and K. Wohn, "Segmentation, Modeling and Classification of the Compact Objects in a Pile," in *SPIE Symposium on Intelligent Robot and Computer Vision VIII: Algorithms and Techniques*, Nov. 1989, pp. 98–108.
- [24] A. Gupta and R. Bajcsy, "Volumetric Segmentation of Range Images of 3D Objects Using Superquadric Models," *Computer Vision Graphics and Image Processing: Image Understanding*, vol. 58, no. 3, pp. 302–326, 1993.
- [25] L. Chevalier, F. Jaillet, and A. Baskurt, "Segmentation and Superquadric Modeling of 3D Objects," in *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, Plzen-Bory, Czech Republic, 2003.
- [26] S. J. Dickinson, D. N. Metaxas, and A. Pentland, "The role of model-based segmentation in the recovery of volumetric parts from range data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 3, pp. 259–267, 1997.
- [27] D. DeCarlo and D. Metaxas, "Shape Evolution with Structural and Topological Changes Using Blending," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1186–1205, 1998.
- [28] T. Horikoshi and S. Suzuki, "3D Parts Decomposition from Sparse Range Data Using Information Criterion," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 168–173, 1993.
- [29] H. Zha, T. Hoshida, and T. Hasegawa, "A Recursive Fitting-And-Splitting Algorithm For 3-D Object Modeling Using Superquadrics," in *IEEE International Conference on Pattern Recognition*, 1998, pp. 658–662.
- [30] A. Pentland, "Automatic Extraction of Deformable Part Models," *Int. J. Comput. Vision*, vol. 4, no. 2, pp. 107–126, 1990.
- [31] R. Duda, P. Hart, and D. Stork, *Pattern Classification, Second Edition*. New York, USA: Wiley Interscience, 2001.
- [32] J. Schoukens and R. Pintelon, *Identification of Linear Systems. A Practical Guideline to Accurate Modeling*. Oxford, UK: Pergamon Press, 1991.
- [33] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation of Range Images as the Search for Geometric Parametric Models," *International Journal of Computer Vision*, vol. 14, no. 3, pp. 253–277, Apr. 1995.
- [34] C. Cea, "Superquadric Recovery in Range Images via Region Growing influenced by Boundary Information," Master's thesis, University of Freiburg, Computer Science Department, Institute for Pattern Recognition and Image Processing, 2004.

- [35] M. Pilu and R. Fisher, "Equal-distance Sampling of Superellipse Models," in *Proceedings of the 1995 British conference on Machine vision (Vol. 1)*. BMVA Press, 1995, pp. 257–266.
- [36] P. Besl and R. Jain, "Segmentation Through Variable-Order Surface Fitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 167–192, 1988.
- [37] D. Katsoulas, "Robust Recovery of Piled Box-Like Objects in Range Images," Ph.D. dissertation, University of Freiburg, department of computer science, 2004.
- [38] J. Freixenet, X. Muñoz, J. Mart, and X. Llad, "Colour Texture Segmentation by Region-Boundary Cooperation," in *European Conference on Computer Vision, Volume 2*. Springer, Sept. 2004, pp. 250–261.
- [39] S. Li and T. Basar, "Distributed Algorithms for the Computation of Non-cooperative Equilibria," *Automatica*, vol. 23, no. 4, pp. 523–533, 1987.
- [40] A. Chakraborty and J. Duncan, "Game Theoretic Integration for Image Segmentation," *IEEE T-PAMI*, vol. 21, no. 1, pp. 12–30, 1999.
- [41] A. Chakraborty, "Feature and module integration for image segmentation," Ph.D. dissertation, Yale University, 1996.
- [42] X. Jiang and H. Bunke, "Edge Detection in Range Images Based on Scan Line Approximation," *Computer Vision and Image Understanding: CVIU*, vol. 73, no. 2, pp. 183–199, Feb. 1999.
- [43] R. Duda and P. Hart, *Pattern Recognition and Scene Analysis*. John Wiley and Sons, 1973.
- [44] D. Katsoulas and A. Werber, "Edge Detection in Range Images of Piled Box-like Objects," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR-2004)*, vol. 2, 2004, pp. 80–84.
- [45] X. Muñoz, "Image Segmentation Integrating Colour, Texture and Boundary Information," Ph.D. dissertation, University of Girona, 2003.
- [46] J. Benois and D. Barba, "Image Segmentation by Region-Contour Cooperation For Image Coding," in *International Conference on Pattern Recognition, Vol. C*, The Hague, Netherlands, 1992, pp. 331–334.
- [47] D. Sinclair, "Voronoi Seeded Colour Image Segmentation," AT&T Laboratories Cambridge, Tech. Rep. TR99-04, May 1999.
- [48] T. Shimbashi, Y. Kokubo, and N. Shirota, "Region Segmentation Using Edge Based Circle Growing," in *IEEE International Conference on Image Processing Vol.C*, Washington D.C., 1995, pp. 65–68.
- [49] X. Jiang, "An Adaptive Contour Closure Algorithm and its Experimental Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1252–1265, 2000.
- [50] A. Chakraborty, L. H. Staib, and J. S. Duncan, "Deformable Boundary Finding Influenced by Region Homogeneity," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 624–627.
- [51] H. Bozma and J. Duncan, "A Game-Theoretic Approach to Integration of Modules," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 1074–1086, 1994.
- [52] C. Ye and J. Borenstein, "Characterization of a 2-D Laser Scanner for Mobile Robot Obstacle Navigation," in *IEEE International Conference on Robotics and Automation*, Washington D.C., May 2002, pp. 2512–2518.
- [53] D. Skocaj and A. Leonardis, "Range image acquisition of objects with non-uniform albedo using structured light range sensor," in *ICPR*, 2000, pp. 1778–1781.
- [54] D. Katsoulas and D. Kosmopoulos, "Box-like Superquadric Recovery in Range Images by Fusing Region and Boundary based Information," in *ICPR*, 2006, Submitted.
- [55] D. Katsoulas, "Reliable recovery of piled box-like objects via parabolically deformable superquadrics," in *ICCV*, 2003, pp. 931–938.