

Robust and Accurate Superquadric Recovery: a Probabilistic Approach

Weixiao Liu^{1,2} Yuwei Wu¹ Sipu Ruan¹ Gregory S. Chirikjian^{1*}
¹National University of Singapore ²Johns Hopkins University
 {mpewxl, yw.wu, ruansp, mpegre}@nus.edu.sg

Abstract

Interpreting objects with basic geometric primitives has long been studied in computer vision. Among geometric primitives, superquadrics are well known for their ability to represent a wide range of shapes with few parameters. However, as the first and foremost step, recovering superquadrics accurately and robustly from 3D data still remains challenging. The existing methods are subject to local optima and sensitive to noise and outliers in real-world scenarios, resulting in frequent failure in capturing geometric shapes. In this paper, we propose the first probabilistic method to recover superquadrics from point clouds. Our method builds a Gaussian-uniform mixture model (GUM) on the parametric surface of a superquadric, which explicitly models the generation of outliers and noise. The superquadric recovery is formulated as a Maximum Likelihood Estimation (MLE) problem. We propose an algorithm, Expectation, Maximization, and Switching (EMS), to solve this problem, where: (1) outliers are predicted from the posterior perspective; (2) the superquadric parameter is optimized by the trust-region reflective algorithm; and (3) local optima are avoided by globally searching and switching among parameters encoding similar superquadrics. We show that our method can be extended to the multi-superquadrics recovery for complex objects. The proposed method outperforms the state-of-the-art in terms of accuracy, efficiency, and robustness on both synthetic and real-world datasets. The code is at http://github.com/bmlklwx/EMS-superquadric_fitting.git.

1. Introduction

Understanding 3D environments (and the objects therein) has been a crucial task for computer vision and intelligent systems [24]. With the recent breakthroughs in 3D vision and the advancement of computing power, modern computer vision systems are able to reconstruct and reason about scenes with low-level representations, such as point

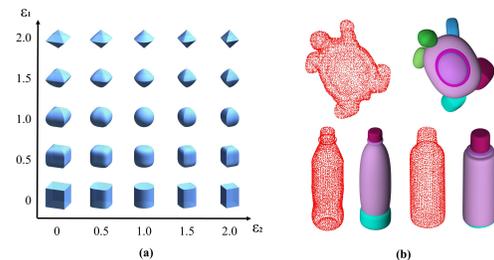


Figure 1. (a) Shape vocabulary of convex superquadrics. We focus on the interpretation of point clouds with convex superquadrics. (b) Superquadric representations inferred by our method.

clouds [11, 25, 26], meshes [18] and voxels [9, 38]. In contrast, the human visual system favors abstracting scenes into canonical parts for a better perceptual understanding [3, 15]. Human vision seems to work well with basic geometric structures of the scene, and not rely on detailed point-wise models [24]. Therefore, understanding the geometric structure of visual inputs appears to be a promising approach for intelligent systems to achieve high-level goals such as physical reasoning, decision making, planning and interaction with the environment. Inspired by this idea, researchers turned to exploring the possibility of using basic volumetric primitives for object description, e.g., cuboids [20, 33, 39]. However, due to the limited expressiveness of cuboids, objects can only be describe in a highly abstracted way.

Superquadrics are a family of geometric primitives with a rich shape vocabulary, including cuboids, cylinders, ellipsoids, octahedra and their intermediates (Fig. 1), but only encoded by 5 parameters. In the recent few years, superquadrics have raised considerable attention in the community and are widely applied in robotics and computer vision tasks, e.g., object modeling [5, 22, 23], collision detection and motion planning [28, 29], pose estimation [4], and grasping [27, 36, 37]. A single superquadric is already expressive enough to reasonably model many everyday objects [14, 30]. The pioneering work on single superquadric recovery includes [1, 12, 32], where the recovery is formulated as a *Least-Square* (LSQ) problem. In [6, 16], the methods are extended to model complex objects with multiple superquadrics. However, the existing methods are vulnera-

*Corresponding author

ble to noise and outliers, and the inherent deficiency of the local optimizer makes them sensitive to initialization.

In this paper, we focus on the single superquadric recovery problem and propose the first probabilistic method to recover superquadrics from noisy point clouds. We build a probabilistic model to simulate how observations (point cloud) are possibly generated from a given parameterized superquadric surface. GUM is adopted to accommodate the generation of noise and outliers. The superquadric recovery is thus formulated as a MLE problem (Sec. 3.1). *Expectation and Maximization* (EM) algorithm is most commonly used to solve MLE problems with latent variables [10]. However, the EM algorithm typically converge to a local optimum, resulting in an inaccurate shape approximation. To solve this problem, we propose a novel algorithm, *Expectation, Maximization and Switch* (EMS), which takes advantage of the geometric features of superquadrics to avoid local optima (Sec. 3.2). In the E-step, the probability of a point being an outlier is inferred from a posterior perspective (Sec. 3.3). In the M-step, the parameter of the superquadric is updated given the current estimation of the latent variables in the E-step (Sec. 3.4). The E and M-step alternate until converged to a local optimum, where we introduce the S-step: globally search for candidate parameters encoding similar superquadrics in terms of shape and pose, and then switch to the one from which we can further increase the likelihood (Sec. 3.5). This strategy is made possible by investigating the geometric symmetry and algebraic ambiguity of superquadrics. Furthermore, we show our probabilistic formulation can be extended to multi-superquadrics recovery tasks (Sec. 3.6).

The motivations and key features of this paper come from the two following questions. (1) Can we design an algorithm robust, accurate and efficient enough to abstract an optimal superquadric primitive from a noisy point cloud whose underlying shape is indeed within the vocabulary of superquadrics? This question is answered in Sec. 4.1 and Sec. 4.2. (2) If the point cloud is too complex to be approximate by a single superquadric, can the algorithm figure out a major part which can be approximated by a superquadric? This is a challenging goal which cannot be achieved by any existing single superquadric recovery algorithm. We demonstrate qualitatively that the proposed method is capable of doing this in Sec. 4.3.

2. Related work

2.1. Preliminary: Superquadrics

Superquadrics are a family of geometric primitives (Fig. 1), which can be defined by the implicit function [2]:

$$F(\mathbf{x}) \doteq \left(\left(\frac{x}{a_x} \right)^{\frac{2}{\epsilon_2}} + \left(\frac{y}{a_y} \right)^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + \left(\frac{z}{a_z} \right)^{\frac{2}{\epsilon_1}} \quad (1)$$

where $\mathbf{x} \doteq [x, y, z]^T \in \mathbb{R}^3$ is a point defined in the superquadric frame; a_x, a_y and $a_z \in \mathbb{R}_{>0}$ are the scale parameters corresponding to the x, y, z-axis; ϵ_1 and $\epsilon_2 \in \mathbb{R}_{\geq 0}$ are the shape parameters. ϵ_1 governs the shape along the z-axis and ϵ_2 controls the shape orthogonal to the z-axis. A superquadric is convex, if both ϵ_1 and ϵ_2 are within $(0, 2]$. Eq. (1) is also called the ‘inside-outside’ function, since \mathbf{x} locates on the surface if $F(\mathbf{x}) = 1$, inside if $F(\mathbf{x}) < 1$ and outside otherwise. We are able to fully parameterize a superquadric of a general pose with $\theta \doteq \{\epsilon_1, \epsilon_2, a_x, a_y, a_z, g\}$, where $g = [\mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3] \in SE(3)$ is the Euclidean transformation.

2.2. Superquadric Recovery

The single superquadric recovery plays an important role in all superquadric recovery problems. In the previous works [1, 12, 13, 32, 35], it is formulated as a LSQ problem and solved with the Levenberg-Marquardt (LM) algorithm [17]. Recently, researchers have also explored the possibility of deep learning approaches [21, 31]. Solina *et al.* [1, 32] propose a cost function based on the implicit function of superquadrics. Gross *et al.* [12] modifies the cost function with the radial distance between points and the superquadric surface. In [13, 34], the robustness of the algorithms is investigated, and outliers are randomly rejected by a customized heuristic function. All the above methods suffer from numerical instability when either ϵ_1 or ϵ_2 approaches 0. As a consequence, they compromise by constraining the lower bounds of ϵ_1 and ϵ_2 to 0.1, resulting in less accuracy when modeling shapes with sharp edges, *e.g.*, cuboids and cylinders. Recently, this problem is revisited and solved by approximating the implicit function with an auxiliary function in the unstable region [35]. Also in [35], the authors point out that the methods are sensitive to the initialization, and suggest to try 3 different initial guesses obtained by performing the Principal Component Analysis (PCA) to the point cloud. However, all the methods are still subject to local optima, especially when the point cloud is partially sampled on the object surface.

For the multi-superquadrics recovery, Leonardis [16] Borges [5], and Chevalier *et al.* [6] try to interpret complex objects with multiple superquadrics by first segmenting the point cloud into parts and fitting a superquadric to each of them. In contrast, our multi-superquadrics recovery extension works in a hierarchical way. Other than a multi-superquadrics representation, we can also obtain the hierarchical relationships among the superquadrics (Fig. 3).

Another line of the multi-superquadrics recovery is based on deep learning [22, 23], where superquadrics can be obtained from meshes or even an RGB image. However, their approaches are data-driven with limitations in generalization, while we focus on the recovery of *unseen* real-world objects based on a case by case probabilistic reasoning.

3. Our Method

3.1. Probabilistic Formulation

In this section, we demonstrate how to formulate the superquadric recovery as a MLE problem. Our model is built as follows. First, on the superquadric surface $\mathcal{S}_\theta \subset \mathbb{R}^3$ parameterized by θ , a Gaussian centroid $\mu \in \mathcal{S}_\theta$ is sampled randomly according to the uniform density function

$$p(\mu) = \frac{1}{A_\theta}, \quad A_\theta = \int_{\mathcal{S}_\theta} 1 dS \quad (2)$$

where A_θ is the area of the superquadric surface. Subsequently, an observation $\mathbf{x} \in \mathbb{R}^3$ is generated from a Gaussian-uniform model (GUM), the probability density function of which is given as

$$p(\mathbf{x}|\mu) = w_o p_o(\mathbf{x}) + (1 - w_o) \mathcal{N}(\mathbf{x}|\mu, \Sigma) \quad (3)$$

where $\mathcal{N}(\cdot|\mu, \Sigma)$ represents the density function of a Gaussian distribution with mean μ and covariance Σ . We assume the noise to be isotropic, *i.e.*, $\Sigma = \sigma^2 \mathbf{I}$, where $\mathbf{I} \in \mathbb{R}_{3 \times 3}$ is the identity matrix. A uniform outlier component is introduced to model the generation of outliers. $w_o \in [0, 1]$ is the probability of a point being sampled from the outlier component. To make Eq. (3) a proper probability density function (integrates to 1 over \mathbb{R}^3), we introduce a working space \mathbb{V} encapsulating the point set, whose volume equals V . The probability density of the outlier component $p_o(\mathbf{x}) = \frac{1}{V}$ if $\mathbf{x} \in \mathbb{V}$ and $p_o(\mathbf{x}) = 0$ otherwise.

For the convenience of the upcoming derivation, we reformulate Eq. (3) into an equivalent form by introducing a latent random variable $z \in \{0, 1\}$ encoding the membership of \mathbf{x} . When $z = 0$, \mathbf{x} is sampled from the uniform outlier component. On the contrary, $z = 1$ indicates that \mathbf{x} is generated from the Gaussian inlier component. We assume the membership z is independent of the mean μ of the Gaussian component. Therefore, Eq. (3) is equivalent to

$$p(\mathbf{x}|\mu, z) = p_o(\mathbf{x})^{1-z} \cdot \mathcal{N}(\mathbf{x}|\mu, \Sigma)^z \quad (4)$$

$$z \sim p(z) = \text{Bernoulli}(1 - w_o)$$

Given a set of points $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^3 | i = 1, 2, \dots, N\}$, the parameters of the superquadric surface can be estimated by maximizing the following likelihood function

$$L(\theta, \sigma^2) = \prod_{i=1}^N p(\mathbf{x}_i|\mu_i, z_i) p(\mu_i) p(z_i) \quad (5)$$

or equivalently, minimizing the negative log-likelihood function

$$l(\theta, \sigma^2) = \sum_{i=1}^N z_i \left(\frac{\|\mathbf{x}_i - \mu_i\|_2^2}{2\sigma^2} - \log c \right) + N \log(A_\theta) \quad (6)$$

where c is the normalizing constant of the Gaussian distribution defined in Eq. (3). Note that the terms independent of θ and σ^2 have been omitted for simplicity.

3.2. Expectation, Maximization and Switching

Eq. (6) is intractable to solve directly because of the existence of a set of continuous latent variables $\mu_i \in \mathcal{S}_\theta, i = 1, 2, \dots, N$, and discrete latent variables $z_i \in \{0, 1\}, i = 1, 2, \dots, N$. The EM algorithm [10] is most commonly used to solve MLE with latent variables. However, in our scenario, it suffers from multiple local optima and has difficulties in calculating the expectation of the latent variables in an exact form. Therefore, we propose a novel algorithm called EMS, which is adapted from the EM algorithm by utilizing the geometric properties of superquadrics. The algorithm first estimates the latent variables from a posterior perspective given the current estimation of the superquadric parameter (E-step). Next, the parameters are inferred and updated by minimizing the negative log-likelihood function (M-step). These two steps alternate until converged to a local minimum. Then, we globally search for candidate parameters which encode superquadrics geometrically similar to the current estimation but have the potential of further decreasing the negative log-likelihood function (S-step). If a valid candidate is found, we reinitialize the EM-phase with this candidate; else, we declare a termination and return the final result. Detailed motivations and derivations of each step are discussed in the following sections.

3.3. E-step: Latent Variables Estimation

Given the current superquadric parameter θ , we estimate μ_i via *maximum a posteriori* (MAP)

$$\hat{\mu}_i = \arg \min_{\mu_i \in \mathcal{S}_\theta} p(\mu_i|\mathbf{x}_i) \quad (7)$$

which guarantees to minimize Eq. (6). By the Bayes' rule, we show that it is equivalent to find the nearest point on the superquadric surface to \mathbf{x}_i

$$\hat{\mu}_i = \arg \min_{\mu_i \in \mathcal{S}_\theta} \|\mathbf{x}_i - \mu_i\|_2 \quad (8)$$

As far as we know, there is no analytical solution to Eq. (8), and numerical optimization is prohibitively time consuming for a large number of observations. Therefore, we approximate $\hat{\mu}_i$ with a sub-optimal solution

$$\hat{\mu}_i^s = \mathbf{x}_i - \left(1 - F^{-\frac{\epsilon_1}{2}}(g^{-1} \circ \mathbf{x}_i) \right) (g^{-1} \circ \mathbf{x}_i) \quad (9)$$

which is the intersection of the vector \mathbf{x}_i (defined in the superquadric frame) and the superquadric surface \mathcal{S}_θ . $F(\cdot)$ is the implicit function of the superquadric. This idea was first introduced by [12] to establish the radial distance measure for superquadrics fitting.

Given the current estimation of μ_i , the expectation of the posterior probability of \mathbf{x}_i being an inlier is inferred via the Bayes' rule

$$E(z_i = 1 | \mathbf{x}_i, \hat{\mu}_i^s) = \frac{\mathcal{N}(\mathbf{x}_i | \hat{\mu}_i^s, \sigma^2 \mathbf{I})}{\mathcal{N}(\mathbf{x}_i | \hat{\mu}_i^s, \sigma^2 \mathbf{I}) + \frac{w_o p_o}{1-w_o}} \quad (10)$$

For simplicity, we denote the value of the expectation as \hat{z}_i .

3.4. M-step: Parameter Optimization

Substituting μ_i and z_i in Eq. (6) with their posterior estimations $\hat{\mu}_i^s$ and \hat{z}_i , we obtain an updated negative log-likelihood function

$$\hat{l}(\theta, \sigma^2) = \sum_{i=1}^N \hat{z}_i \left(\frac{\|\mathbf{x}_i - \hat{\mu}_i^s\|_2^2}{2\sigma^2} - \log c \right) + N \log(A_\theta) \quad (11)$$

To the best of our knowledge, A_θ cannot be expressed in closed-form with respect to θ . For efficiency, A_θ is approximated by a bi-linear interpolation along ϵ_1 and ϵ_2 . Details are presented in the Supplementary Material. $N \log(A_\theta)$ indicates a preference for a small superquadric. To avoid local optima, it is preferable to delay the introduction of this term until σ^2 falls below a threshold, *i.e.*, the shape is roughly captured. We apply the trust-region-reflective [8] to optimize θ . The main reason we do not use the LM solver lies in its numerical instability when approaching the lower bound of the shape parameters. In contrast, the reflective transformation [7], empirically, always keeps the optimization numerically stable. More discussion can be found in the Supplementary Material. After θ is solved, σ^2 is updated by setting the corresponding partial derivative of Eq. (11) to zero.

3.5. S-step: Geometric Local Optimum Avoidance

Superquadrics hold many geometric and algebraic features, *e.g.*, symmetry and parametric ambiguity. Trivial cases can be raised where two superquadrics possess a similar or even identical shape and pose but are distant in the parameter space. The likelihood function Eq. (11) quantifies how well an underlying geometric primitive fits the observations, and thus two similar (or identical) superquadrics must possess similar (or identical) likelihoods as well. Inspired by this idea, we propose a geometry-guided local optima avoidance strategy: globally search for candidate parameters encoding similar shapes, and switch to the one which has the potential of further increasing the likelihood. The switch action is discontinuous in the parameter space, while the change in terms of the underlying geometric shape is almost smooth. We investigate the conditions when two parameters distant to each other represent similar superquadrics, and summarize the cases into two categories. The idea of switch and examples of similarities are visualized in Fig. 2.

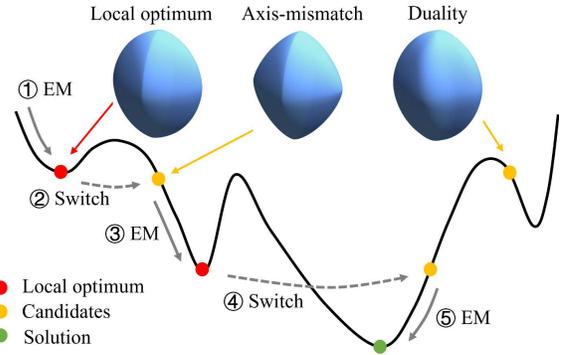


Figure 2. Visualization of the S-step and similarities. The EM steps change the shape and pose of the superquadric towards a local optimum, while the S-steps tries to preserve the current shape but jump between distant parameters to avoid the local optimum.

Axis-mismatch similarity: According to the definition in Sec. 2.1, the x and y-axis are of an equivalent class in terms for interchangeability, *i.e.*, if we interchange the value of a_x and a_y , we can obtain an identical superquadric by rotating 90 degrees about the z-axis. However, the z-axis distinguishes itself because it is generally not interchangeable with them. We thus name the z-axis as the principal axis. Now let us consider a special case when the shape parameters $\epsilon_1 = \epsilon_2$. In this scenario, the implicit function (Eq. (1)) degenerates and makes all the three axes equivalent. From another perspective, the ‘degree’ of interchangeability decrease as the difference between ϵ_1 and ϵ_2 increases. Therefore, when ϵ_1 and ϵ_2 are close to each other, a similar superquadric can be obtained by reassigning the principal axis to either the x or y-axis and then applying a corresponding rotation. We call this axis-mismatch, because it helps to avoid the local optimum caused by the incorrect assignment of the principal axis, which prohibit the shape from further varying to the global optimum.

Duality similarity: This case is derived from the shape ambiguity regarding ϵ_2 . When $a_x = a_y$ and $\epsilon_2 = 0$, we can always represent the same superquadric (in terms of shape and pose) by scaling a_x and a_y with $1/\sqrt{2}$, setting $\epsilon_2 = 2$, and rotating 45 degrees about the principal axis. This is because $\epsilon_2 = 0$ and $\epsilon_2 = 2$ define a pair of squares dual to each other in the subspace orthogonal to the principal axis. Utilizing this property, when $a_x \approx a_y$, a similar superquadric can be constructed by setting $\epsilon_2 = 2 - \epsilon_2$, re-scaling and rotating. The duality similarity is of great importance to local optimum avoidance, since it connects two extreme ends of the parameter space. Local optimizer tends to converge monotonically towards one of the ends, while the duality similarities glue the ends together, allowing the algorithm to check and explore the counterpart.

Similar superquadrics can also be resulted by a combination of the two similarities. Detailed mathematical formulation about the candidates generation and switching strategy

can be found in the Supplementary Material.

3.6. Extension: Multi-superquadrics Recovery

Although our method is designed for single superquadric recovery, it can be easily extended to understand the shape and structure of a complex object with multiple superquadrics. One of the key features of our single superquadric recovery method is that it can capture the major superquadric-like part of a point cloud and identify the unfitted points as ‘outliers’. Therefore, from a point cloud \mathbf{X} of a complex object, we can first recover a superquadric capturing the major superquadric structure along with a set of unfitted ‘outlier’ clusters \mathcal{C} . Then, we feed \mathcal{C} back to the algorithm as the input to recover a new set of superquadrics along with an updated \mathcal{C} . The process (Algorithm 1) is repeated hierarchically until \mathcal{C} is empty or a maximum depth of layer is reached. In this way, not only the point cloud is represented with the union of the acquired superquadrics, but also the hierarchical relationship between the superquadrics is obtained, as illustrated in Fig. 3.

Algorithm 1 A layer of hierarchical superquadric recovery

- 1: **Input:** $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$
 - 2: **Output:** $\{\theta_1, \theta_2, \dots, \theta_n\}, \mathcal{C} = \{\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \dots, \bar{\mathbf{X}}_m\}$
 - 3: $\mathcal{C} \leftarrow \{\}$
 - 4: **for** $i = 1, \dots, n$ **do**
 - 5: $(\theta_i, outliers) \leftarrow EMS(\mathbf{X}_i)$
 - 6: $\{\bar{\mathbf{X}}_{1_i}, \bar{\mathbf{X}}_{2_i}, \dots, \bar{\mathbf{X}}_{\gamma_i}\} \leftarrow Clustering(outliers)$
 - 7: ▷ clustering the outliers by Euclidean distance [19]
 - 8: $\mathcal{C} \leftarrow \mathcal{C} \cup \{\bar{\mathbf{X}}_{1_i}, \bar{\mathbf{X}}_{2_i}, \dots, \bar{\mathbf{X}}_{\gamma_i}\}$
 - 9: ▷ clusters with points less than a threshold are pruned
 - 10: **end for**
-

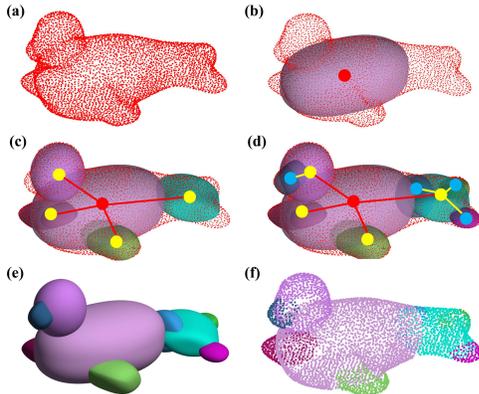


Figure 3. Hierarchical process of the multi-superquadrics recovery. We demonstrate with a *Seal* from the KIT ObjectModels. (a) The original point cloud. (b-d) Steps of the hierarchical process and the generation of the structure graph. (e) The recovered superquadric model. (f) Segmentation of the original point cloud based on the superquadric model.

4. Experiments

We conduct experiments on both synthetic and real-world point cloud datasets to verify the accuracy, efficiency and robustness of our method. All the algorithms and experiments are implemented in MATLAB, on a computer running Intel Core i9-9900K(3.6GHz). In all the single superquadric recovery experiments, we compare with 4 baseline methods: Implicit-LSQ [32], Radial-LSQ [12], Robust-fitting [13] and Numerical Stable method [35].

Initialization: Following [35], if not specified, the baseline methods are initialized with 3 different superquadric parameters obtained via PCA to alleviate the affect of local optima. To show the effectiveness of the proposed local optimum avoidance strategy, our method is only initialized once with the first one among the three. Implementation details can be found in the Supplementary Material.

4.1. Recovery on Synthetic Datasets

In this section, we evaluate the performance of the algorithms on recovering the underlying geometric primitives given points sampled on a superquadric. Three experiments are designed to evaluate the robustness to partial data, outliers, and noise, respectively. To cover the whole parameter space of superquadrics in convex region, we randomly generate 500 superquadrics under the following strategy: the shape parameter $[\epsilon_1, \epsilon_2]$ is uniformly sampled on $(0, 2]^2 \subset \mathbb{R}^2$; the scale parameter $[a_x, a_y, a_z]$ is uniformly sampled on $[0.5, 3]^3 \subset \mathbb{R}^3$; the rotation $\mathbf{R} \in SO(3)$ is generated by first sampling a rotation axis on the unit sphere and then applying a random rotation in $[0, 2\pi]$ about the axis; the translation \mathbf{t} is uniformly sampled on $[-1, 1]^3 \subset \mathbb{R}^3$. Point clouds are generated by sampling evenly on the surface of the random superquadrics with an equal[†] interval of 0.2. In [35], a recursive algorithm is proposed to realize this task. However, the points sampled by the algorithm get denser when approaching the poles along the z-axis. Therefore, we further improve their algorithm by adaptively re-scaling the sampling intervals at different latitudes to achieve an overall equal-distance sampling (see Supplementary Material).

The most natural metric to evaluate the goodness of fitting is the average point-to-surface orthogonal distance. However, given an arbitrary point, there is no analytical solution of the corresponding closest point on a superquadric surface. Therefore, we approximate the point-to-surface metric with

$$error(\mathbf{X}, \mathbf{S}) \doteq \frac{1}{N} \sum_{i=1}^N \min_{\mathbf{s}_j \in \mathbf{S}} \|\mathbf{x}_i - \mathbf{s}_j\|_2 \quad (12)$$

where $\mathbf{S} = \{\mathbf{s}_j \in S_\theta | j = 1, 2, \dots, M\}$ is a set of points

[†]No strict and theoretical way of equal-distance sampling exists. Here we indicate a practical approximation of almost equal-distance sampling.

densely and evenly sampled (with an interval of 0.02 in the synthetic experiments) on the superquadric surface S_θ .

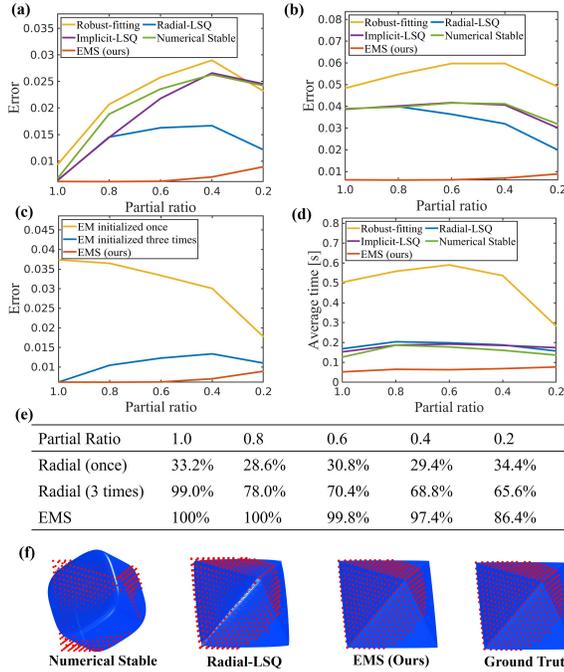


Figure 4. Results of recovery from partial data. (a) Average error (the baselines are initialized 3 times). (b) Average error (the baselines are initialized once). (c) Ablation study. (d) Average runtime. (e) Global Optimality rate (declared when the fitting error is less than 0.01) (f) Examples at partial ratio 0.4, when the baselines get stuck in local optima, while ours have managed to jump out.

Recovery on partial data: This experiment simulates the cases where points are sampled from a partial area of a superquadric in the sense that sometimes we can only obtain a partial view of an object. For each point cloud of the 500 superquadrics, a point is randomly selected and a certain percentage (partial ratio) of the points which are the closest to the random point are kept. We test on 5 different levels of partial ratio ranging from 1.0 (complete point cloud) to 0.2, resulting in totally 2500 point clouds. Results are shown in Fig. 4. Our method significantly outperforms the baseline methods at all levels of partial ratios. All the methods perform well when a complete point cloud of a superquadric is available. However, with the decreasing of the partial ratio, it becomes harder to make a reasonable initial guess of the parameter. Therefore, the baseline methods become more likely to get stuck at a local optimum, even with multiple trials. While, our method continuously avoids the local optima with the geometry-guided switching strategy, and thus maintains a high success rate in converging to the global optimum. We also perform ablation studies about the S-step. The results verify its effectiveness (Fig. 4(c)). Although our algorithm only guarantees local optimality, the

empirical results indicate that our method is very likely to achieve the global optimum in practice (Fig. 4(e)).

Robustness to outliers: We add different levels of Gaussian outliers to the 2500 point clouds. Outlier ratio (the number of outliers / the number of inliers) is used to quantify the level of outliers added. Results are shown in Fig. 5. Our method shows superior robustness to outliers and is able to recover the underlying superquadrics even from severely corrupted point clouds. The Implicit-LSQ, Radial-LSQ and Numerical Stable are vulnerable to outliers, since they are not designed to tolerate outliers. The Robust-fitting performs better, however, is limited by its heuristic outlier rejection policy, which also imposes an undesirable negative effect on clean point clouds. In contrast, our method infers outliers in a posterior probabilistic approach and thus maintains high robustness among all levels of outliers.

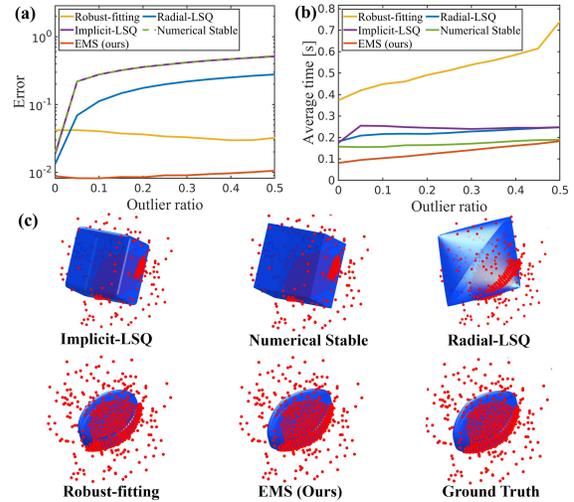


Figure 5. Outlier results. (a) Average error at different outlier ratios. (b) Average runtime. Runtime of Radial, Implicit and NS has no significant meaning since they cannot recover the superquadric correctly in most cases. (c) Examples at outlier ratio 0.4.

Robustness to noise: In this experiment, we corrupt each point in the point clouds with Gaussian noise of different variance. Results are shown in Fig. 6. It can be observed that the fitting error grows with the increase of noise level for all the methods, including ours. This is because, unlike outliers, the noise inherently corrupts the geometric information. But still, our method shows advantages in robustness and efficiency when dealing with noisy point clouds.

4.2. Recovery on Real-world Datasets

In this section, we evaluate the single superquadric recovery on two public datasets: the KIT ObjectModel [14] and BigBIRD [30]. Both of the datasets consist of various daily household objects, and are widely used as benchmarks

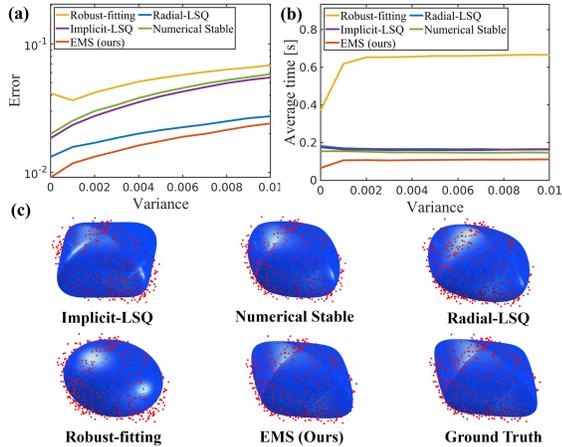


Figure 6. Noise results (a) Average errors at different noise levels (evaluated on uncorrupted point clouds). (b) Average runtime. (c) Examples at $\sigma^2 = 0.01$ Gaussian noise level.

in 3D-object reconstruction, recognition, and robot grasping. KIT is scanned by a high-accuracy laser scanner (Minolta VI-900), while BigBIRD is captured by a commercial-grade RGB-D camera (PrimeSense Carmine 1.08).

KIT ObjectModels: We follow the setting in [35]. The dataset includes 145 objects, among which [35] selects 105 items that can be reasonably approximated by a single superquadric. However, we decide to shrink our scope to 97 items, since some of the objects (*e.g.*, bottles) would be better recovered with multiple superquadrics. Therefore, we defer the discussion of those cases to Sec. 4.3. We down-sample the point clouds with voxelized grids to about 1500 points. Quantitative results are shown in Fig. 7(c). Our method outperforms all the baseline methods in accuracy and efficiency. Our method also demonstrates some interesting features which can not be shown quantitatively. As shown in Fig. 7(a), some objects have minor local deformations and/or parts deviating from the main structure, which affect the major shape capturing. Our probabilistic method is able to reduce the impact of those parts (inferred as outliers) and achieves an overall better fitting. Also in Fig. 7(b), the box-like objects are usually filled with content inside, resulting in a shape slightly extended in the middle. This shape is within the vocabulary of superquadrics, however, is so close to cuboids that the other algorithms fail in telling them apart, *i.e.*, get stuck at a local optimum. In contrast, our method is able to switch and check among similar shapes and recover the optimal superquadric.

BigBIRD: Compared with laser scanners, low-cost RGB-D cameras are inferior in measurement accuracy, resulting in noisy point clouds with outliers. Therefore, the BigBIRD is an ideal dataset for evaluating the robustness of an algorithm in daily operation. The dataset consists of 125 object instances, and we test on 91 of them which can be reasonably approximated with single superquadrics.

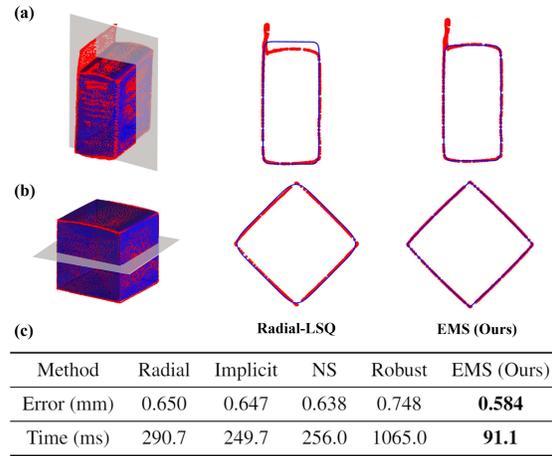


Figure 7. KIT results. (a) and (b) are the two cross-section views of fitting results between Radial-LSQ and the proposed method. (c) Quantitative results of average error (Eq. (12) with sampling interval as dense as 0.1 mm) and runtime.

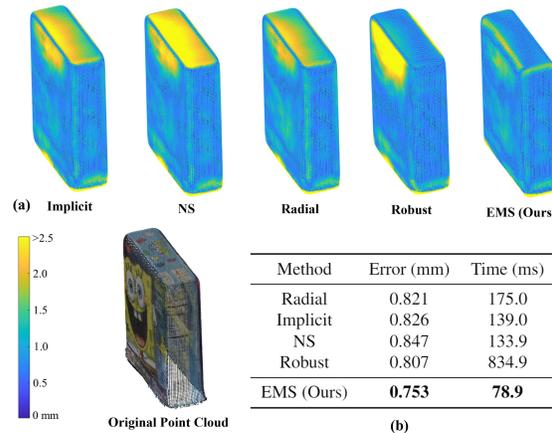


Figure 8. BigBIRD results. (a) A heatmap showing the local fitting error between the recovered superquadric and the original point cloud. (b) Quantitative results of the average error (Eq. (12) with sampling interval as dense as 0.1 mm) and runtime.

The point clouds are downsampled with voxelized grids to around 1000 points. Results are shown in Fig. 8. Among the baselines, the Robust-fitting shows better robustness but takes a much longer time to converge. Our method can achieve the best accuracy with the least time consumption.

4.3. Multi-superquadrics Recovery

In this section, we show the qualitative results of our extension to the multi-superquadrics recovery (Sec. 3.6). We use the complex objects which cannot be properly approximated by a single superquadric from the KIT ObjectModels. We downsample the point clouds to around 5000 points, so as to keep enough details of the complex shapes. We set the maximum hierarchical layers to 3 and the prun-

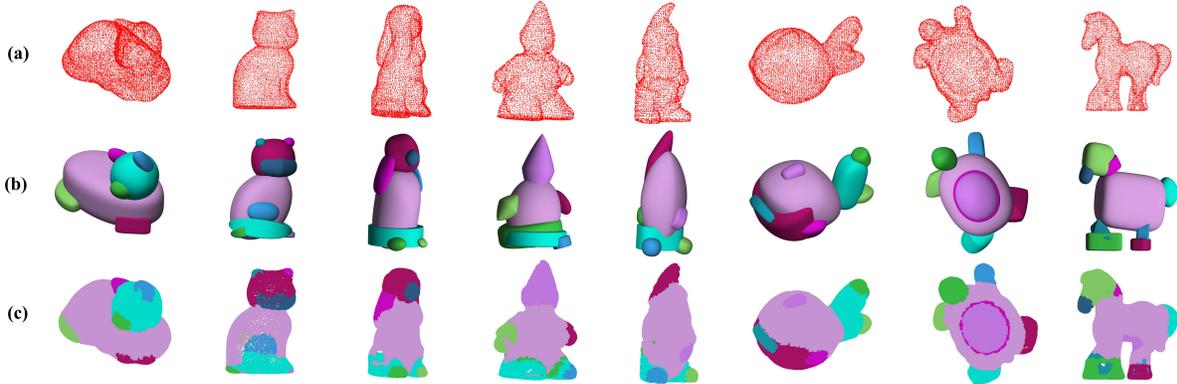


Figure 9. Examples of multi-superquadrics recovery. (a) The original point clouds. (b) The recovered multi-superquadrics representations with the proposed method. (c) Segmentation of the original point clouds based on the multi-superquadrics representations.

ing threshold to 60 points. As shown in Fig. 10, the baseline single superquadric method fails in capturing the major superquadric-like part of a bottle as a cylinder. Our method succeeds in doing this and identifies the points around the cap of the bottle as outliers, from which the bottle is recovered hierarchically. Other results on more complex objects are shown in Fig. 9. Our method is able to represent the objects vividly with superquadrics. Some detailed parts such as ears and mouths of animals can also be recovered.

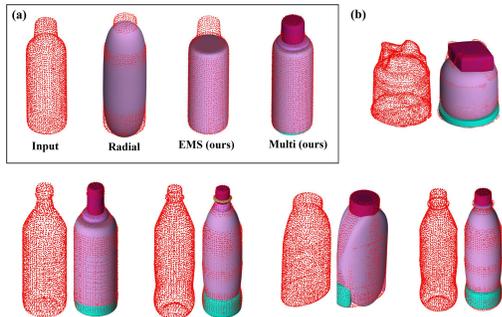


Figure 10. Qualitative multi-superquadrics recovery on bottles from KIT. (a) Comparison of single superquadric recovery between Radial-LSQ and EMS (ours). The proposed method is able to recover the major superquadric shape (a cylinder) from the point cloud. (b) Other results of the proposed method.

5. Discussion & Limitations

Various experiments show that our method is robust, accurate and efficient. As shown in Sec. 4.1 and Sec. 4.2, the baseline methods are sensitive to initialization, and are easily trapped into local optima. Therefore, to obtain a satisfying result, they usually require multiple trials starting from different initial guesses. In comparison, our method is able to avoid local optima on the fly and achieve a better recovery accuracy with one shot. Also, our method is robust to outliers and noise. This feature is brought in by our probabilistic formulation of the problem, where we explicitly model

the possible sources of outliers and noise with a uniform distribution and a Gaussian distribution, respectively. This feature also forms the basis for our multi-superquadrics extension. As shown in Sec. 4.3, our method is able to capture the major superquadric-like part from a point cloud and hierarchically generate subordinate superquadrics from the ‘outliers’ identified at the upper level.

However, there is a limitation in our multi-superquadrics extension: our approach requires the object being recovered possesses an inherent hierarchical geometric structure. For more general cases, a preliminary point cloud segmentation is preferred [25, 26]. Another limitation is due to the expressiveness of superquadrics. It is hard to well capture the shape of a thin-walled non-convex object such as a cup. In future works, we plan to adapt deformations and primitive subtractions to extend the vocabulary of superquadrics.

6. Conclusion

We present the first probabilistic method to recover superquadrics from point clouds. We formulate the recovering process as an MLE problem and propose a novel geometry-guided approach, EMS, to infer an optimal superquadric. The proposed method outperforms the state-of-the-art in terms of accuracy, robustness, and efficiency, which is substantiated by extensive experiments on both synthetic and real-world datasets. We also extend our method to represent a complex object with a hierarchy of superquadrics. Such representation can reveal the geometric and hierarchical structure of an object, which, we believe, has great potential in facilitating tasks such as object classification, segmentation and grasping.

Acknowledgements This research is supported by the National Research Foundation, Singapore, under its Medium Sized Centre Programme - Centre for Advanced Robotics Technology Innovation (CARTIN) R-261-521-002-592, and JHU internal funds.

References

- [1] R. Bajcsy and F. Solina. Three dimensional object representation revisited. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 231–240, 1987. 1, 2
- [2] A. H. Barr. Superquadrics and angle-preserving transformations. *IEEE Computer graphics and Applications*, 1(1):11–23, 1981. 2
- [3] I. Biederman. Recognition-by-components: a theory of human image understanding. *Psychological review*, 94(2):115, 1987. 1
- [4] G. Biegelbauer and M. Vincze. Efficient 3D object detection by fitting superquadrics to range image data for robot’s object manipulation. In *Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1086–1091, 2007. 1
- [5] D. L. Borges and R. B. Fisher. Class-based recognition of 3d objects represented by volumetric primitives. *Image and Vision Computing*, 15(8):655–664, 1997. British Machine Vision Conference. 1, 2
- [6] L. Chevalier, J. Jaillet, and A. Baskurt. Segmentation and superquadric modeling of 3D objects. In *The 11-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision’2003, WSCG 2003*, 2003. 1, 2
- [7] T. F. Coleman and Y. Li. On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds. *Math. Program.*, 67:189–224, 10 1994. 4
- [8] T. F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6(2):418–445, 1996. 4
- [9] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’96*, page 303–312, New York, NY, USA, 1996. Association for Computing Machinery. 1
- [10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. 2, 3
- [11] H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3D object reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1
- [12] A. D. Gross and T. E. Boult. Error of fit measures for recovering parametric solids. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 690–694, 1988. 1, 2, 3, 5
- [13] Y. Hu and W. G. Wee. Robust 3D part extraction from range images with deformable superquadric models. In *Signal Processing, Sensor Fusion, and Target Recognition IV*, volume 2484, pages 524 – 535. International Society for Optics and Photonics, SPIE, 1995. 2, 5
- [14] A. Kasper, Z. Xue, and R. Dillmann. The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research*, 31(8):927–934, 2012. 1, 6
- [15] D. K. Katherine and S. S. Elizabeth. Core systems in human cognition. In *From Action to Cognition*, volume 164 of *Progress in Brain Research*, pages 257–264. Elsevier, 2007. 1
- [16] A. Leonardis, A. Jaklic, and F. Solina. Superquadrics for segmenting and modeling range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1289–1295, 1997. 1, 2
- [17] K. Levenberg. A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. 2
- [18] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, Aug. 1987. 1
- [19] Matlab optimization toolbox, 2021a. The MathWorks, Natick, MA, USA. 5
- [20] C. Niu, J. Li, and K. Xu. Im2Struct: Recovering 3D shape structure from a single RGB image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1
- [21] T. Oblak, J. Šircelj, V. Štruc, P. Peer, F. Solina, and A. Jaklič. Learning to predict superquadric parameters from depth images with explicit and implicit supervision. *IEEE Access*, 9:1087–1102, 2021. 2
- [22] D. Paschalidou, L. V. Gool, and A. Geiger. Learning unsupervised hierarchical part decomposition of 3D objects from a single RGB image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2
- [23] D. Paschalidou, A. O. Ulusoy, and A. Geiger. Superquadrics revisited: Learning 3D shape parsing beyond cuboids. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2
- [24] A. P. Pentland. Perceptual organization and the representation of natural form. In M. A. Fischler and O. Firschein, editors, *Readings in Computer Vision*, pages 680–699. Morgan Kaufmann, San Francisco (CA), 1987. 1
- [25] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 1, 8
- [26] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. 1, 8
- [27] A. H. Quispe, B. Milville, M. A. Gutiérrez, C. Erdogan, M. Stilman, H. Christensen, and H. B. Amor. Exploiting symmetries and extrusions for grasping household objects. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3702–3708, 2015. 1
- [28] S. Ruan and G. S. Chirikjian. Closed-form Minkowski sums of convex bodies with smooth positively curved boundaries. *Computer-Aided Design*, 143:103133, 2022. 1

- [29] S. Ruan, K. L. Pobleto, Y. Li, Q. Lin, Q. Ma, and G. S. Chirikjian. Efficient exact collision detection between ellipsoids and superquadrics via closed-form Minkowski sums. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 1765–1771, 2019. [1](#)
- [30] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel. BigBIRD: A large-scale 3D database of object instances. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 509–516, 2014. [1](#), [6](#)
- [31] J. Šircelj, T. Oblak, K. Grm, U. Petković, A. Jaklič, P. Peer, V. Štruc, and F. Solina. Segmentation and recovery of superquadric models using convolutional neural networks. In *25th Computer Vision Winter Workshop*, 2020. [2](#)
- [32] 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. [1](#), [2](#), [5](#)
- [33] S. Tulsiani, H. Su, L. J. Guibas, A. A. Efros, and J. Malik. Learning shape abstractions by assembling volumetric primitives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. [1](#)
- [34] E. R. Van Dop and P. P. L. Regtien. Fitting undeformed superquadrics to range data: improving model recovery and classification. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 396–401, 1998. [2](#)
- [35] N. Vaskevicius and A. Birk. Revisiting superquadric fitting: A numerically stable formulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):220–233, 2019. [2](#), [5](#), [7](#)
- [36] G. Vezzani, U. Pattacini, and L. Natale. A grasping approach based on superquadric models. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1579–1586, 2017. [1](#)
- [37] G. Vezzani, U. Pattacini, G. Pasquale, and L. Natale. Improving superquadric modeling and grasping with prior on object shapes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6875–6882, 2018. [1](#)
- [38] A. Zeng, S. Song, M. Niessner, M. Fisher, J. Xiao, and T. Funkhouser. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. [1](#)
- [39] C. Zou, E. Yumer, J. Yang, D. Ceylan, and D. Hoiem. 3D-PRNN: Generating shape primitives with recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. [1](#)