



# Volumetric Representation of Semantically Segmented Human Body Parts Using Superquadrics

Ryo Hachiuma<sup>(✉)</sup>  and Hideo Saito 

Keio University, Yokohama, Kanagawa, Japan  
{ryo-hachiuma,hs}@keio.jp

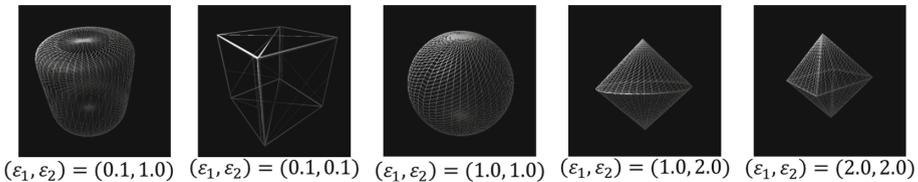
**Abstract.** Superquadrics are one of the ideal shape representations for adapting various kinds of primitive shapes with a single equation. This paper revisits the task of representing a 3D human body with multiple superquadrics. As a single superquadric surface can only represent symmetric primitive shapes, we present a method that segments the human body into body parts to estimate their superquadric parameters. Moreover, we propose a novel initial parameter estimation method by using 3D skeleton joints. The results show that superquadric parameters are estimated, which represent human body parts volumetrically.

**Keywords:** Superquadrics · Volumetric representation · Semantic segmentation

## 1 Introduction

The idea of automatically sensing and discovering information about the 3D human body has been of interest in several areas of computer vision for many years. For example, human activity recognition [17] and pose estimation [5, 11] are at the base of many augmented/virtual reality and robotic applications. The extracted information about the human body can be represented in many ways, such as cylinders [13], skeletons [10], and joint skeletons [5].

Recently, superquadric [4] has been revisited to represent objects efficiently and comprehensively [15]. Superquadrics are ideal shape representations for adapting various primitive shapes with a single equation. Applying the



**Fig. 1.** The various superquadric shapes according to  $\varepsilon_1$  and  $\varepsilon_2$ .

superquadric to an object enables the object to be expressed by various primitive shapes, such as cuboids, cylinders, and spheres with several parameters in the equation. Figure 1 shows examples of various superquadric surfaces with different shape parameters ( $\varepsilon_1, \varepsilon_2$ ). The superquadric parameters of real-world objects are estimated from 3D point cloud of them [18]. An equation obtained by substituting the 3D point cloud of an object into superquadric representation is regarded as a non-linear least squares problem, and the parameters are estimated using the Levenberg-Marquardt (LM) algorithm.

In the 1990s, superquadrics were employed to represent humans volumetrically [9]. Although a single superquadric can only represent the symmetric primitive shape, a human shape is represented with multiple superquadrics by approximating each body part as a symmetric primitive shape. Previous work [9] handled this task as a toy problem, and depth information was obtained from a structured light range scanner, which required a difficult setup to capture. Afanasyev *et al.* [2] proposed a method to estimate the body pose from a depth image, representing body parts with superquadrics. However, as they aimed to only estimate the pose and they fixed the superquadric shape and scale parameters. Paschalidou *et al.* [15] estimated multiple superquadric parameters by a convolutional neural network from the 3D mesh. Furthermore, Sundaresan *et al.* [19] proposed a method to estimate scale and pose parameters of superquadrics from 3D voxel data. Unlike the previous methods [15, 19], we present a method which estimates superquadric parameters from a single RGB and 2.5 D depth image.

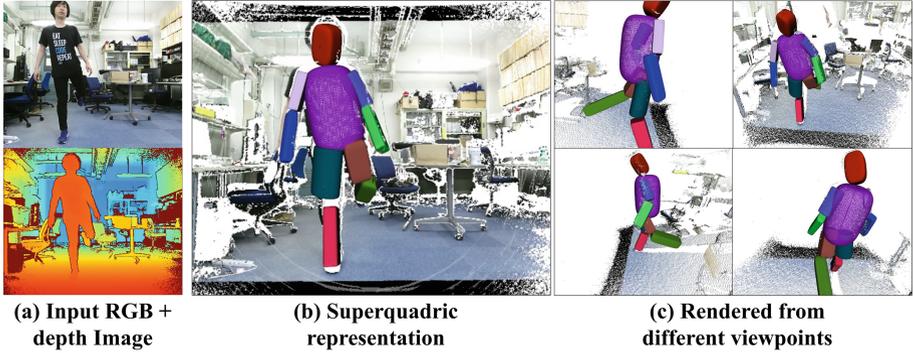
This paper revisits the task of representing the 3D human body with multiple superquadrics. We present a method to estimate multiple superquadrics that represent the 3D human body from a single RGB-D image (Fig. 2). Our method consists of two steps: segmenting the RGB-D image into 3D body parts and estimating each superquadric parameters from each segmented point cloud. We propose the initial parameter estimation that uses 3D human skeleton joints. In the experiment, we recorded three sequence with Kinect v.2 to verify the effectiveness of our proposed method. We employ the Chamfer distance as the evaluation metric. We confirmed that effectiveness of the proposed superquadric initial parameter estimation method.

## 2 Superquadrics

The way to define a superquadric in an superquadric-centered coordinate system is the inside-outside function with a scale parameter ( $s_x, s_y, s_z$ ) and a shape parameter ( $\varepsilon_1, \varepsilon_2$ ):

$$F(x, y, z, \mathbf{A}) = \left\{ \left( \frac{x}{s_x} \right)^{\frac{2}{\varepsilon_2}} + \left( \frac{y}{s_y} \right)^{\frac{2}{\varepsilon_2}} \right\}^{\frac{\varepsilon_2}{\varepsilon_1}} + \left( \frac{z}{s_z} \right)^{\frac{2}{\varepsilon_1}}, \quad (1)$$

where  $\mathbf{A}$  is a tuple as  $(s_x, s_y, s_z, \varepsilon_1, \varepsilon_2)$ . Parameters  $s_x$ ,  $s_y$ , and  $s_z$  are scale parameters that define the superquadric size at the  $x$ ,  $y$ , and  $z$  coordinates,



**Fig. 2.** Volumetric representation of human body parts using superquadrics. We take the input of a single RGB and depth image (a). The output is multiple superquadrics that represent the segmented human body parts (b). As each segmented part is represented by superquadrics, the hidden area’s shape can also be recovered as volumetric representation (c).

respectively. Parameters  $\varepsilon_1, \varepsilon_2$  are shape representation parameters that express squareness along the  $z$  axis and the  $x$ - $y$  plane. Also, a point which lies on a superquadric surface can be defined below:

$$\mathbf{x}(\eta, \omega) = \begin{bmatrix} s_x \cos^{\varepsilon_1}(\eta) \cos^{\varepsilon_2}(\omega) \\ s_y \cos^{\varepsilon_1}(\eta) \sin^{\varepsilon_2}(\omega) \\ s_z \sin^{\varepsilon_1}(\eta) \end{bmatrix}. \quad (2)$$

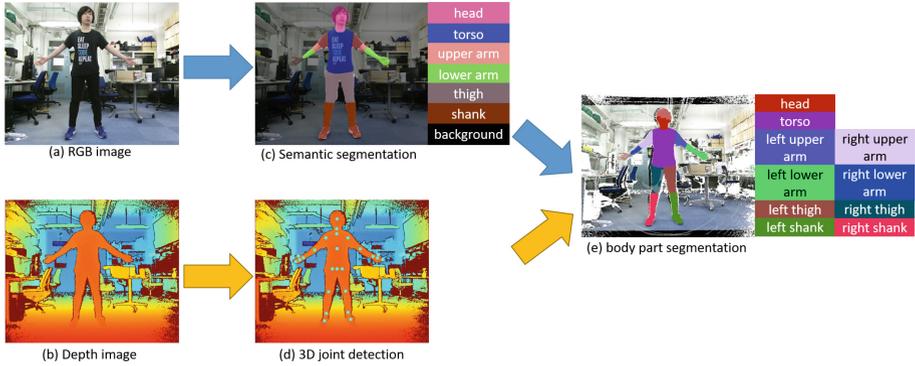
$$-\pi/2 \leq \eta \leq \pi/2, -\pi \leq \omega \leq \pi.$$

As The surface of superquadrics is located in the original coordinate system, the superquadrics can be expressed in a generic coordinate system by adding six further variables, representing the six superquadric pose, with a total of eleven independent variables, i.e.  $\mathbf{q} \in \mathbb{R}^{11}$ .

### 3 Method

In this paper, we follow the representation of the human body part defined by Krivic *et al.*’s [9], which is separated into 10 parts: *head* (1), *torso* (2), *right/left upper/lower arm* (3–6), *right/left thigh/shank* (7–10).

Our method consists of two steps to estimate superquadric parameters of each human body part. Each superquadric parameter is estimated from point cloud of each point cloud of human body part. Therefore, first, we segment the 3D point cloud which is obtained from RGB-D sensor into each body part. Second, we estimate superquadric parameters from each segmented point cloud.



**Fig. 3.** The flow of human body part segmentation. The color code is visualized at the right side of the image for Figure (c) and (e). (Color figure online)

### 3.1 Body Part Segmentation

To extract the 3D point cloud of human body parts, we combine the human body semantic segmentation and 3D human joint detection. The flow of human body part segmentation is shown in Fig. 3. From an RGB image, we apply semantic segmentation to label human body part to each pixel. We employ Light-Weight RefineNet [14] which shows high accuracy on the PASCAL Person-Part dataset [6] while keeping the computational efficiency. In the PASCAL Person-Part dataset, there are seven categories of labels: *head*, *torso*, *upper arm*, *lower arm*, *thigh*, and *shank*. Figure 3(c) shows the result of semantic segmentation. Note that left/right limbs (lower arm, upper arm, thigh and shank) are not segmented each other.

Therefore, we use 3D skeleton joints to segment left/right limbs. The skeleton’s 3D positions are estimated using the method proposed by Shotton *et al.* [16]. Figure 3(d) shows the result of 3D skeleton joints estimation. The 3D skeleton joints are projected onto the depth image and colored in cyan. As the RGB and depth images’ coordinate systems are not aligned, we transform the semantically labeled image to the depth coordinate using the intrinsic and extrinsic parameter of the RGB-D sensor.

For each 3D point which is labeled as a limb, we calculate the Euclidean distance between the point and each 3D skeleton joints. If the nearest joint belongs to a left limb, the 3D point is labeled a left body part. For example, if the nearest joint is *left shoulder*, the point is labeled as a *left upper arm*. Figure 3(e) shows the result of human body part segmentation. Compared to Fig. 3(c), left limbs and right limbs are segmented each other.

### 3.2 Superquadric Parameter Estimation

The superquadric parameter is estimated from the extracted point cloud of each body part. The minimization of the algebraic distance from points to the

superquadric surface can be solved by defining a non-linear least-squares minimization problem:

$$\min_{\mathbf{q}} \sum_{k=0}^K (\sqrt{s_x s_y s_z} (F^{\varepsilon_1}(\mathbf{p}_k; \mathbf{q}) - 1))^2, \quad (3)$$

where  $K$  denotes number of 3D points in the point cloud,  $\mathbf{p}_k$  denotes each 3D points in the point cloud, and  $\mathbf{q}$  is superquadric parameters which fits the input 3D point cloud.  $(F(\text{Tr}_{\Phi}(\mathbf{p}_i); \Lambda) - 1)^2$  imposes the point to superquadric surface distance minimization, where the term  $\sqrt{s_x s_y s_z}$  is proportional to superquadric volume, compensates for the fact that the previous equation is biased toward larger superquadric surfaces. We employ Levenberg-Marquardt [12] algorithm to minimize the above equation.

It is known that the optimization function (Eq. 3) will be numerically unstable [20] when  $\varepsilon_1, \varepsilon_2$  are less than 0.1. Moreover, the superquadric will have concavities when  $\varepsilon_1, \varepsilon_2 > 2.0$ . Therefore, we employ the constraints when minimizing the function in Eq. 3 for the shape parameters:  $0.1 < \varepsilon_1, \varepsilon_2 < 2.0$  and for the scale parameters:  $s_1, s_2, s_3 > 0.0$ .

As the minimization function is not a convex function, the initial parameters determine which local minimum the minimization converges to. Most of the works [7, 8] which estimate superquadric parameters employ the initial parameter estimation method proposed by Solina *et al.* [18]. Unlike the estimation method in the previous work, we propose a novel approach to estimate initial parameters using 3D skeleton joints. We denote the previous initial parameter estimation method as the baseline method.

**Initial Translation Parameter.** The baseline method estimated initial translation parameters by calculating the centroid from all 3D points, and the centroid is set to the initial translation parameter. However, as the point cloud is captured from a single viewpoint, the centroid point is drifted to the direction of the origin of the coordinate system. On the other hand, the model proposed by Shotton *et al.* [16] is trained to estimate the 3D skeleton joints from human meshes so that the estimated joints do not drift even if from a single viewpoint. Therefore, we take the average of each body part’s 3D joint skeletons in order to set the initial translation parameter. For example, we take the average of the 3D coordinate of joints belonging to the *torso* to estimate its translation parameter.

**Initial Rotation Parameter.** The baseline method calculated the covariance matrix from all 3D points and the eigenvectors of the matrix are set to the initial rotation parameter. We set the initial rotation parameter which aligns the z-axis of superquadric surface to be parallel to the vector of two connected 3D joints in each body part.

**Initial Scale and Shape Parameter.** We compute the covariance matrix of each body part’s 3D point cloud, and the three eigenvalues of the matrix are set to initial scale parameters. For initial shape parameters, we set  $\varepsilon_1 = 1.0, \varepsilon_2 = 1.0$  for the *head* and  $\varepsilon_1 = 0.1, \varepsilon_2 = 1.0$  for other body parts that approximates the each body part.

**Table 1.** The average Chamfer distance [cm] for each sequence (*seq1*, *seq2*, *seq3*). Lower is better.

	<i>seq1</i>	<i>seq2</i>	<i>seq3</i>	Average
Baseline	1.735	1.671	1.535	1.647
Proposed	<b>1.105</b>	<b>1.533</b>	<b>1.349</b>	<b>1.329</b>

## 4 Experiment

We recorded three sequences using Kinect v.2. We denote three sequences as *seq1*, *seq2* and *seq3*. There are at total 125, 95, and 53 frames in each sequence, respectively.

### 4.1 Evaluation Metric

We employed a Chamfer distance metric to evaluate if the estimated superquadric surfaces represented the original point cloud. Chamfer distance calculates the distances between given two set of point clouds  $S_1, S_2$ :

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2. \quad (4)$$

The algorithm of the Chamfer distance finds the nearest neighbor points in one set and sums the squared distances up. In this experiment, one point set is 3D points that are labeled with body parts, while the other is 3D points sampled from the estimated superquadric surface.

By sampling 3D points from a superquadric surface according to Eq. 5, the regions which exhibit high curvature are emphasized, shown in Fig. 4 (a). For an unbiased sample distribution we need to apply equidistant sampling using spherical angles as introduced by Bardinnet *et al.* [3],

$$\hat{\mathbf{x}}(\eta, \omega) = \begin{bmatrix} s_x \rho \cos(\eta) \cos(\omega) \\ s_y \rho \cos(\eta) \sin(\omega) \\ s_z \rho \sin(\eta) \end{bmatrix}, \quad (5)$$

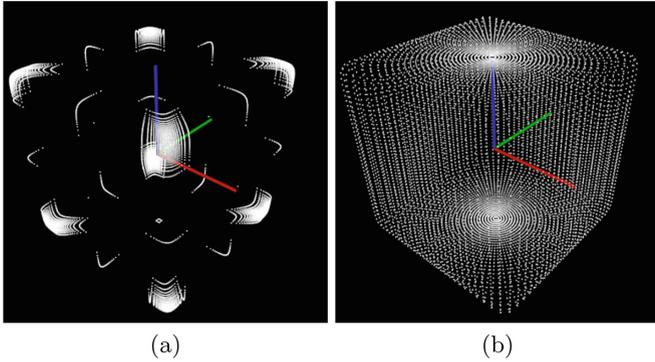
where

$$\rho = \left( \left( |\cos \eta \cos \omega|^{\frac{2}{\varepsilon_2}} + |\sin \eta \cos \omega|^{\frac{2}{\varepsilon_2}} \right)^{\frac{\varepsilon_2}{\varepsilon_1}} + |\sin \eta|^{\frac{2}{\varepsilon_1}} \right)^{\frac{-\varepsilon_1}{2}}.$$

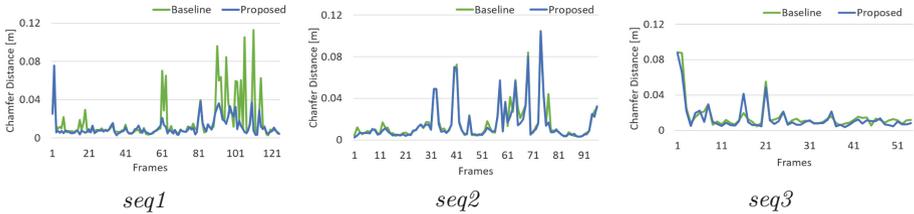
Figure 4(b) shows the sampling result by Eq. 5. The 3D points are uniformly sampled from superquadric surface.

### 4.2 Qualitative Results

To evaluate the effectiveness of our initial parameter estimation method, we compared the estimation results with the previous work [18] (baseline). This



**Fig. 4.** Sampled points from superquadric surface by Eqs. 1 and 5. The left figure is sampled point cloud with the Eq. 1, and the right figure is the sampled point cloud with Eq. 5. The Superquadric parameter is set to  $(\varepsilon_1 = 0.1, \varepsilon_2 = 0.1, s_1 = 1.0, s_2 = 1.0, s_3 = 1.0)$ .

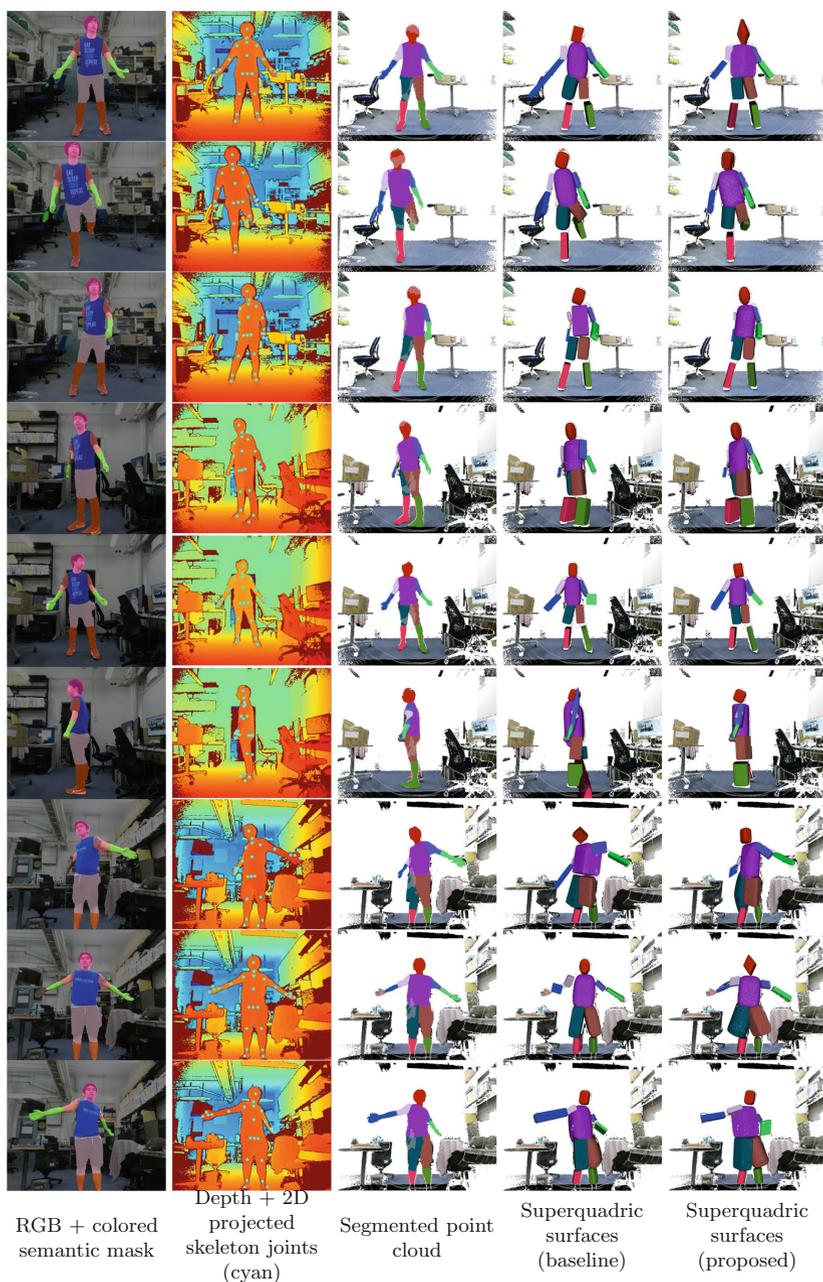


**Fig. 5.** Chamfer distance between estimated superquadric surface and the 3D points which labeled as the human of three sequences. Lower is better.

previous method has been widely used for estimating the initial parameter estimation [1, 8]. Figure 6 shows the superquadric estimation of three frames at three sequences. It demonstrates that our proposed method successfully estimated multiple superquadric parameters which approximate the point cloud of each body part. At the second row, superquadric parameters are estimated even if the person raised the left shank, and the proposed method estimated accurate superquadric pose parameters compared to the baseline method. Moreover, at the sixth row, we can verify the effectiveness of our proposed method with the superquadric parameters of the left upper arm and the right shank.

### 4.3 Quantitative Results

The averaged Chamfer distance across the entire frame in each sequence is summarized in Table 1. For example in *seq1*, 1.74cm for the baseline method and 1.11cm for our proposed method. Figure 5 shows the Chamfer distance of each frame in each sequence. From the figure, we can verify that our novel initial parameter estimation method found more optimal parameters than the previous method [18].



**Fig. 6.** The result of superquadric parameter estimation by the baseline method and the proposed method at three frames in each sequence. The first row to the third are the frames from *seq1*, the fourth row to the sixth are the frames from *seq2* and the seventh row to the ninth are the frames from *seq3*.

## 5 Conclusion

To revisit the task of representing the human shape volumetrically with multiple superquadrics, we presented a method to estimate superquadric parameters that represent the 3D human body. Moreover, we proposed a novel initial parameter estimation method that uses 3D skeleton joints. The results showed that our method successfully represented the 3D human body with multiple superquadrics. Additionally, we compared our initial estimation method with the previous method and verified its effectiveness by comparing the Chamfer distance between the estimated superquadric surfaces and the point cloud of humans.

In the future, we will develop applications that leverage the three big advantages of superquadrics. First, unlike the 3D skeleton joint representation, the superquadric representation contains not only the 3D position but also the volumetric information of the person. Mehta *et al.* [11] showed the virtual reality application of 3D pose estimation. In the application, the estimated 3D pose is used to provide the pose of a virtual avatar. By using superquadric representation, the avatar refers not only the pose of the user but also the shape of the user. Second, as superquadric scale parameters directly represent the size of each body part, the size of a human body can be easily measured from a single view and used to virtually fit or customize clothes. Finally, as the shape of the hidden area is recovered using superquadrics, the recovered information can be used to generate free viewpoint images.

**Acknowledgement.** This work was supported by AIP-PRISM, Japan Science and Technology Agency, Grant Number JPMJCR18Y2, Japan.

## References

1. Abelha, P., Guerin, F.: Learning how a tool affords by simulating 3D models from the web. In: IEEE/RSJ IROS, pp. 4923–4929, September 2017
2. Afanasyev, I., et al.: 3D human body pose estimation by superquadrics. In: VIS-APP, vol. 2, January 2012
3. Bardinet, E., Cohen, L.D., Ayache, N.: A parametric deformable model to fit unstructured 3D data. *Comput. Vis. Image Underst.* **71**(1), 39–54 (1998)
4. Barr, A.H.: Superquadrics and angle-preserving transformations. *IEEE Comput. Graph. Appl.* **1**(1), 11–23 (1981)
5. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields. In: CVPR (2017)
6. Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasun, R., Yuille, A.L.: Detect what you can: detecting and representing objects using holistic models and body parts. In: CVPR, pp. 1979–1986 (2014)
7. Drews Jr., P., Trujillo, P.N., Rocha, R.P., Campos, M.F.M., Dias, J.: Novelty detection and 3D shape retrieval using superquadrics and multi-scale sampling for autonomous mobile robots. In: ICRA, pp. 3635–3640 (2010)
8. Duncan, K., Sarkar, S., Alqasemi, R., Dubey, R.: Multi-scale superquadric fitting for efficient shape and pose recovery of unknown objects. In: ICRA, pp. 4238–4243 (2013)

9. Krivic, J., Solina, F.: Part-level object recognition using superquadrics. *CVIU* **95**(1), 105–126 (2004)
10. Shi, L., Cheng, I., Basu, A.: Anatomy preserving 3D model decomposition based on robust skeleton-surface node correspondence. In: International Conference on Multimedia and Expo, pp. 1–6, July 2011
11. Mehta, D., et al.: VNect: real-time 3D human pose estimation with a single RGB camera, vol. 36 (2017). <http://gvv.mpi-inf.mpg.de/projects/VNect/>
12. Moré, J.J.: The Levenberg-Marquardt algorithm: Implementation and theory. In: Watson, G.A. (ed.) *Numerical Analysis. LNM*, vol. 630, pp. 105–116. Springer, Heidelberg (1978). <https://doi.org/10.1007/BFb0067700>
13. Lee, M.W., Cohen, I.: A model-based approach for estimating human 3D poses in static images. *Trans. Pattern Anal. Mach. Intell.* **28**(6), 905–916 (2006)
14. Nekrasov, V., Shen, C., Reid, I.D.: Light-weight refinenet for real-time semantic segmentation. In: *BMVC* (2018)
15. Paschalidou, D., Ulusoy, A.O., Geiger, A.: Superquadrics revisited: learning 3D shape parsing beyond cuboids. In: *CVPR* (2019)
16. Shotton, J., et al.: Real-time human pose recognition in parts from single depth images. In: *CVPR*, pp. 1297–1304, June 2011
17. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS 2014*, vol. 1, pp. 568–576 (2014)
18. Solina, F., Bajcsy, R.: Range image interpretation of mail pieces with superquadrics. In: *AAAI*, pp. 733–737 (1987)
19. Sundaresan, A., Chellappa, R.: Model driven segmentation of articulating humans in laplacian eigenspace. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 1771–85 (2008)
20. Vaskevicius, N., Birk, A.: Revisiting superquadric fitting: a numerically stable formulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(1), 220–233 (2019)