Hierarchical Superquadric Decomposition with Implicit Space Separation

Jaka Šircelj\textsuperscript{1,2}, Peter Peer\textsuperscript{1}, Franc Solina\textsuperscript{1}, Vitomir Štruc\textsuperscript{2}

\textsuperscript{1}Faculty of Computer and Information Science, UNI-LJ, Večna pot 113, Ljubljana, Slovenia
\textsuperscript{2}Faculty of Electrical Engineering, UNI-LJ, Tržaška cesta 25, Ljubljana, Slovenia

Abstract

We introduce a new method to reconstruct 3D objects using a set of volumetric primitives, i.e., superquadrics. The method hierarchically decomposes a target 3D object into pairs of superquadrics recovering finer and finer details. While such hierarchical methods have been studied before, we introduce a new way of splitting the object space using only properties of the predicted superquadrics. The method is trained and evaluated on the ShapeNet dataset. The results of our experiments suggest that reasonable reconstructions can be obtained with the proposed approach for a diverse set of objects with complex geometry.

1 Introduction

A vital field in 3D vision and robotics is 3D object reconstruction. By compressing input data into simpler representations such as meshgrids or sets of simpler objects, we can understand and interact with objects and their surroundings more easily, solving problems like collision avoidance [7] or grasp planning [2], [9].

A common type of reconstruction is to represent a given (3D) object with a set of simpler shape primitives, often referred to as \textit{geons}. A popular approach in geon reconstruction is to estimate a fixed amount of geons and, concurrently, predict which of those should be kept in the final geon set, thus retaining the variability in geon numbers [8], [6]. In our previous work, for example, we used a MaskRCNN model to first segment the input depth image and then reconstruct the parts using a specific type of geons, called superquadrics [10].

Following the overall idea of our prior work, we introduce a novel method to reconstruct 3D objects with a set of superquadrics (SQ) in this paper. However, in this novel approach we base the procedure on a hierarchical tree decomposition algorithm. Instead of using a model that returns a fixed amount of geons at once, we introduce a hierarchical decomposition model, that incrementally splits the input object into more and more SQ representations, as illustrated in Figure 1. While a similar method was introduced by Paschalidou \textit{et al.} in [5], we propose a splitting method based on the superquadrics characteristic, alleviating the model from determining how the object should be hierarchically split.

We evaluate the proposed approach on the ShapeNet dataset. Two models with different reconstruction capabilities are trained for the experiments. The first, facilitating reconstructions with a maximum of 4 superquadrics, is trained and tested on the full ShapeNet dataset. The second, enabling reconstructions with up to 8 superquadrics, is assessed on the pistol ShapeNet subset. Results are presented in terms of IoU scores and with visual examples, and point to the feasibility of the proposed solution.

2 Method

2.1 Superquadrics

To reconstruct the 3D shapes we use superquadrics, which are geometric shapes that can describe objects such as spheres, ellipsoids, cylinders and rectangular cuboids. A common way to describe the surface of a superquadric is using the implicit function $F(x, y, z) = 1$. Here $F$ is called the inside-outside function and is defined as

\begin{equation}
F(x, y, z) = \left(\frac{x}{a_1}\right)^{\frac{2}{\varepsilon_1}} + \left(\frac{y}{a_2}\right)^{\frac{2}{\varepsilon_2}} + \left(\frac{z}{a_3}\right)^{\frac{2}{\varepsilon_1}},
\end{equation}

where $a_1, a_2, a_3$ define the size of the superquadric and $\varepsilon_1, \varepsilon_2$ its shape. We can also move the superquadric in space by offsetting the $x, y, z$ coordinates by $t_1, t_2, t_3$ respectively and by rotating the coordinates using quaternion notation $q_0, q_1, q_2, q_3$. To abbreviate the notation we write the spatial coordinates using vector notation $\mathbf{x} = [x, y, z]$.
and all superquadric parameter as

$$\lambda = [a_1, a_2, a_3, \varepsilon_1, \varepsilon_2, t_1, t_2, t_3, q_1, q_2, q_3, q_4].$$

A useful property of the inside-outside function, that will be used further on, is that it contains information whether the space is inside or outside the superquadric (hence the name). Points with $F(x) < 1$ lie inside the object, $F(x) > 1$ lie outside, and, as already suggested, points with $F(x) = 1$ lie on its surface. For numerical stability we rather compare $F(x)\varepsilon_1$ values [3], the spatial inequalities remains the same.

2.2 Model architecture

Similarly to Paschalidou et al. [5] we use a hierarchical decomposition procedure to reconstruct a complex 3D object using a set of superquadrics. This allows the decomposition to predict different numbers of superquadrics, increasing their number for fine detail, and decreasing for larger parts of the object. The procedure works by first extracting a feature vector out of the input depth image, which we refer to as a depth feature. This is concatenated with a hierarchical feature and passed into a recursive neural network which predicts two new hierarchical features each encoding data to predict a new superquadric. This prediction is done with an SQ predictor network which predicts the parameters $\lambda$ previously described. Each of the two hierarchical features can again be concatenated with the depth feature vector and passed into the recursive neural network, thus splitting the initial superquadric feature into two. A hierarchical feature in the first step can be a simple vector filled with zeroes. The model architecture corresponding to the outlined idea is shown in Figure 2.

Such a model produces a superquadric-pair binary tree, where each node contains two sets of parameters and two links to two children, again containing two sets of parameters. We use the following notation to reference the superquadric parameters in the tree:

$$\text{SQ}_{\text{pair}}_{d,i} = [\lambda_a, \lambda_b],$$

where $d$ is the depth of the node and $i$ is the index of the node in the tree. This notation is also shown in the tree example in Figure 1.

2.3 Training

For training, we use the natural property of the inside-outside function, whose sign describes the inside and the outside space of the superquadric. Similarly to [5], we define the occupancy function $g(x; \lambda)$, which translates this property to act as a binary classifier of the inside space

$$g(x; \lambda) = \sigma(s * (1 - F\varepsilon_1(x; \lambda))),$$

where $\sigma$ is the sigmoid function, and $s$ is a scaling parameter that defines the slant of the values around the surface of the superquadric – for more information see [5].

Since we want to use this occupancy property of the superquadrics, we generate the training dataset by sampling points in space and annotating them with 1 if they lie inside the 3D object and 0 if they lie outside. The root node of the superquadric pair tree, Eq. (2), is trained on these sets of initial ground-truth points and labels, denoted as $P$ and $L_{1,1}$ respectively.

The child SQ pairs should only cover the space in proximity to the space occupied by their parent superquadric, and should not be fitted to the full inside of the 3D model i.e. $L_{1,1}$. Paschalidou et al. [5] solve this problem by also predicting the centroid of the part covered by a predicted superquadric. The space is then split by matching every point to its closest centroid (see Figure 4 in [5]). In this paper, we use the mathematical properties of superquadrics to split the space. More specifically, we found that it is best to split the space inside and outside of the superquadric pairs differently. The space inside of the superquadrics is split using the inside-outside function by finding the maximum of the inside-outside function, i.e.:

$$\arg\max_{i=a,b} F\varepsilon_1(x; \lambda_i).$$

Conversely, the space outside of the superquadric pair is split using the radial Euclidean distance

$$d(x) = |x||1 - F^{-\frac{1}{2}}|,$$

taking the superquadric with the minimum value

$$\arg\min_{i=a,b} d_i(x).$$

This distance function is used outside because it behaves more properly as a distance function further away from a superquadric. For more information on the radial distance function, look at Jaklič et al. [3]. The space-splittin
properties of the inside-outside function and the radial Euclidean distance function are illustrated in Figure 3. The proposed splitting method is simpler than the one used in [5], since we don’t need to predict any additional values, and uses the space splitting capabilities of the superquadrics, thus, splitting the space more naturally and in accordance with the geometry of the superquadrics.

Since our model obtains a hierarchy of superquadrics, we must also construct a hierarchy of labels \( L_{d,i} \) on which we compute the losses. These labels are computed recursively according to how the space is split using the superquadric pairs. The inside of the object is considered inside for the SQ-pair child only if its parent superquadrics’ split space also contains it (see Figure 4). This translates to a logical AND operation between the parents ground truth occupancy \( L_{d,i} \) and the space split of the parent

\[
I_{d,i} = \hat{l}_{\text{parent node}(d,i)} \land \hat{l}_{\text{split}(x; \lambda_{\text{parent sq}(d,i)}, \lambda_{\text{uncle sq}(d,i)})},
\]

where \( I_{d,i} = \sum_{d,i \subset \epsilon \in L_{d,i}} \), and \( \lambda_{\text{parent sq}(d,i)} \) denotes the parent SQ pair node, \( \lambda_{\text{parent node}(d,i)} \) is the parent superquadric (a or b) and \( \lambda_{\text{uncle sq}(d,i)} \) is the other superquadric from the parent pair node. For example, for node \( (d, i) = (2, 2) \):

\[
\begin{align*}
\text{parent node}(2, 2) & = (1, 1) \\
\text{parent sq}(2, 2) & = (1, 1, a) \\
\text{uncle sq}(2, 2) & = (1, 1, b).
\end{align*}
\]

Finally the occupancy loss is calculated using the occupancy values for each pair tree node and the ground truth occupancy values \( L_{d,i} \)

\[
L = \sum_{d=1}^{2^{l-1}} \sum_{i=1}^{d} L_{\text{BCE}}(\max_{\lambda_{sq}=a,b} g(x; \lambda_{sq}), I_{d,i}),
\]

where \( L_{\text{BCE}} \) is the binary cross entropy loss. This loss is similar to the part-reconstruction loss term from Paschalidou et al. [5], but differs in that we look at how accurately an SQ pair matches the designated space of its parent superquadric.

### 3 Results

In this section, we present the results for two trained hierarchic decomposition models. We use the ShapeNet dataset [1] for training along with the NVIDIA Kaolin v0.1 library [4], which is used to compute signed distance function values of the 3D objects and to infer the ground truth inside-outside space. The first model (abbreviated ModelP) is trained on the Pistol subset of ShapeNet with the maximum depth of the superquadric pair tree set to 3 levels. The second model (ModelS) is trained on the full ShapeNet dataset with the maximum depth set to 2 levels. Training with more levels proved challenging, and resulted in degenerated SQs, often small and out of bounds.

**Performance Indicator.** After a train-validation-test split, both models are evaluated on the test subset using the intersection-over-union metric (IoU)

\[
IoU(\Lambda_d, L_{0,0}) = \frac{\sum_{l \in \Lambda_d} \hat{l}_d(x) \land l}{\sum_{l \in L_{0,0}} \hat{l}_d(x) \lor l},
\]

where \( \Lambda_d \) denotes all predicted superquadric parameters in the tree at layer \( d \), \( \Lambda_d = \{ \lambda_{d,i} | i \in [1, 2^{d-1}] \} \) and \( \hat{l}_d(x) \) denotes the predicted inside-outside label for point \( x \). The latter is inferred as inside or outside \( l_d(x) = 1 \) if \( g(x; \lambda_{d,i}) > 0.5 \), for any superquadric in layer \( d \), otherwise the point is outside of the reconstruction.

**Model Comparison.** The IoU results of the ModelS on the full ShapeNet dataset are shown in Table 1 and for separate subsets of Shapenet in Table 2. The IoU results of ModelP for the pistol subset are reported in Table 3. Since the full ShapeNet dataset contains much more variability, ModelS scores a lower IoU over the full ShapeNet dataset.
than ModelIP over the pistol subset. An interesting observation is that ModelS still performs better on the pistol subset than ModelIP, even though the latter was trained exclusively on pistols. This suggests that training on more complex and diverse sets can improve results on specific subsets. However, training on the full ShapeNet did prove more difficult. For instance, we managed to train ModelS for only up to 2 tree levels, while we managed to train ModelIP to predict trees up to 3 levels deep.

Object-specific Results. Regarding the specific subsets of ShapeNet, the model performed best on simpler object classes, e.g., dishwasher or microwave, which have a box-like shape and are, thus, easier to reconstruct using superquadrics. The poorest results are on concave objects, such as bowl or bathtub, since only two levels of SQ-pairs don’t give enough capacity to cover the object.

Qualitative Evaluation. In Figure 5 we show a few reconstruction examples of ModelS. As we can see, the largest issue is the low number of predicted superquadrics, given that ModelS has been trained for only 2 levels, and being capable to reproduce an object with a maximum of 4 superquadrics. While simpler objects like the two cars are well reconstructed, more complex objects prove impossible to reconstruct with 4 superquadrics. For example, both tables have four leg stands, which get covered by two rectangular superquadrics in the left case, and a cylindrical one on the right. The same problem can be observed on both of the planes, where the wings and the engine prove much too complex for only 4 superquadrics.

In Figure 6 we show a few ModelP reconstructions, from worst to best, using the 2nd and 3rd tree level. Here, we observe that the least performing objects were those pistol models that have been rare in the pistol subset, like the 1st and 2nd pistol. The most common handgun is the 4th show the reconstruction from the second SQ tree layer.

The hierarchical decomposition has proven to successfully decompose the targeted input objects, even with single point-of-view depth images, which give a highly occluded representation of the 3D object. We showed that, unlike in Paschalidou et al. [5], we don’t need to predict centroids to split the space into two. It is sufficient to use the already predicted superquadrics for space splitting. These results have been supported by fairly good IoU scores and good visual representations. There is also much to improve in the model. The largest issue we faced was the maximum allowed depth of the SQ pair tree, where we achieved the max depth of 3 with 8 predicted superquadrics.

4 Conclusion

The hierarchical decomposition has proven to successfully decompose the targeted input objects, even with single point-of-view depth images, which give a highly occluded

Figures 5 and 6: Reconstruction examples of ModelS and ModelP on the pistol subset. First column - original object, second - reconstruction from the second SQ tree level, third - reconstruction from the third SQ tree level. The reconstructions are sorted from worst to best with the reconstruction IoU annotated above.

References

[1] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs,GR], 2015.
[2] C. Goldfeder, P. Allen, C. Lackner, and R. Pelossof. Grasp planning via decomposition trees. In ICRA, 2007.
[3] A. Jaklic, A. Leonardiš, and F. Solina. Segmentation and recovery of superquadrics. Kluwer, 2000.
[4] K. M. Jatavallabhula, E. Smith, J.-F. Laflche, C. F. Tsang, A. Rozantsev, W. Chen, T. Xiang, R. Lebaredian, and S. Fidler. Kaolin: A pytorch library for accelerating 3d deep learning research. https://github.com/NVIDIAGameWorks/kaolin, 2022.
[5] D. Paschalidou, L. V. Gool, and A. Geiger. Learning unsupervised hierarchical part decomposition of 3D objects from a single RGB image. In CVPR, pages 1060–1070, 2020.
[6] D. Paschalidou, A. O. Ulusoy, and A. Geiger. Superquadrics revisited: Learning 3D shape parsing beyond cuboids. In CVPR, pages 10344–10353, 2019.
[7] N. E. Smith, R. G. Cobb, and W. P. Baker. Incorporating stochastics into optimal collision avoidance problems using superquadrics. Journal of Air Transportation, 28:65–69, 2020.
[8] S. Tulsiani, H. Su, L. J. Guibas, A. A. Efros, and J. Malik. Learning shape abstractions by assembling volumetric primitives. In CVPR, pages 2635–2643, 2017.
[9] G. Vezzani, U. Pattacini, and L. Natale. A grasping approach based on superquadric models. In ICRA, pages 1579–1586, 2017.
[10] 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 CVWW, 2020.