Scale-Invariant, Unsupervised Part Decomposition of 3D Objects

Safoura Rezapour Lakani, Mirela Popa, Antonio J. Rodríguez-Sánchez, Justus Piater

Institute of Computer Science, University of Innsbruck, Austria

Abstract. This paper presents a novel 3D object decomposition method based on *supervoxels* and *low-curvature regions*. We consider representing an object with a combination of primitive shapes in the recognized regions. We propose a scale-invariant, non-parametric shape estimation method. The experiments show promising results on unseen objects.

1 Introduction

Objects can be represented by a combination of parts. Certain parts in an object can be used for object classification or recognition. For instance, in Figure 1, in order to classify *objects with handle*, only *handles* must be detected. This should lead to more efficient and generalizable recognition procedures than global representations of all the objects with a handle. Decomposing objects into meaningful parts relevant to a specific task has been studied in computer vision [2, 3, 1]. In these approaches, learning starts either from labeled patches or from patches of specific sizes. In this work, we propose an unsupervised method for decomposing 3D point clouds of objects into parts. We are motivated by the application of grasping. Assuming we know the grasp types for a set of *primitive shapes*, we aim to transfer these to previously-unseen objects. The primitive shapes are cone, cylinder, cube, sphere. We aim to have a general method for representing shapes rather than parametric methods. Therefore, we represent them by their surface-normal distributions. We made the assumption that the primitive shapes are smooth. Therefore, we enforce the smoothness property by decomposing the object into *low-curvature regions* and then estimate shapes in those regions.

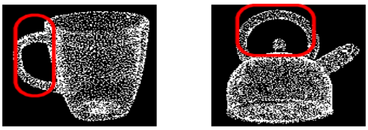


Fig. 1. Recognizing objects with a *handle*.

2 Model-based Object Decomposition Into Parts

The main goal is to recognize multiple instances of given primitive shapes in an object. As mentioned in Section 1, we estimate shapes in the low-curvature regions in the object. Hence, our algorithm has two steps. In the first step we find these regions in an object. Then, we estimate primitive shapes in each region.

2.1 Low-Curvature Regions

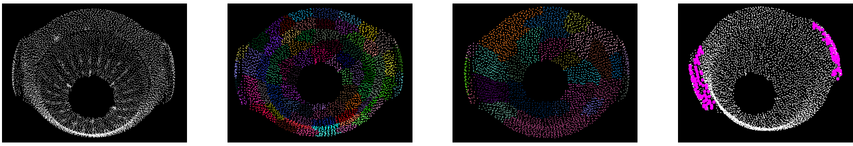
A *low-curvature region* is defined as a locally planar surface in the object. To this end, we identify convexity or concavity in the object and cut the object in those parts. We consider the high-curvature regions as breaking *local connectivity* between adjacent patches. We then propagate the connectivity information to the neighboring patches to obtain the low-curvature regions. Consequently, our method for finding low-curvature regions is twofold. First, we decompose the object into locally smooth patches based on their surface-normal distributions; secondly, we estimate convexity, concavity or flatness based on the obtained patches.

Locally Smooth Patches In this step, we want to get the patches that are locally planar, i.e. the normal vectors inside each patch are parallel. We estimate surface normals on the input point cloud, and we segment it to supervoxels based on the normal vectors. We used supervoxels [4] (available from *PCL*[1]). This method uses K-Means, starting with evenly-distributed cluster seeds over an object. The algorithm considers a fixed size for all the clusters. Due to this fact, we also merge the adjacent supervoxels if their mean surface normals are close to parallel.

Estimating Local Surface Curvatures Shape From supervoxels obtained as mentioned above, we compute the local surface curvature shape in an object by considering the curvature changes between adjacent supervoxels. We use the *HK* algorithm [6] for this aim. The *HK* algorithm uses the mean and the Gaussian curvatures to classify surface shape curvatures. The algorithm is originally based on range images, but we adapted it (as mentioned in [5]) for point clouds. The *HK* algorithm, categorizes the surface shape into five categories; flat, convex elliptic, concave elliptic, convex cylindrical, concave cylindrical.

We apply the *HK* algorithm on the estimated curvature of border voxels between two adjacent supervoxels. For noise robustness purposes, we use all the border voxels between each two adjacent supervoxels to vote for the shape of a surface. In the case that the estimated surface shape based on two adjacent supervoxels is non-flat, we consider them as disconnected. Figure 2 shows the results for object supervoxel segmentation and connectivity for an object.

[1] http://pointclouds.org/



(a) Original point cloud

(b) Segmented supervoxels denoted by different colors

(c) Merged supervoxels denoted by different colors

(d) Purple dots indicate regions of low connectivity

Fig. 2. Results for decomposed patches and connectivity for an object

Forming Low-Curvature Regions Having supervoxels and their connectivity information, we further merge them to a *low-curvature region* with connected supervoxels. We define a region starting with a random supervoxel and we grow it incrementally as long as we do not reach to a discontinuity which is caused by high-curvature. We continue this process until all supervoxels have been visited.

2.2 Probabilistic Shape Assignment

From the low-curvature regions obtained as above, we estimate the likelihood of a specific region given each particular primitive shape. As a region consists of a set of supervoxels this likelihood is further computed based on the likelihood of each supervoxel given a specific primitive shape. The primitive shape corresponding to a region is determined using Maximum-Likelihood principle:

$$m_V = \underset{m \in M}{\operatorname{argmax}}\, p(V|m), \tag{1}$$

,where m is a specific primitive shape, M is the set of all primitive shapes, V is the set of supervoxels composing a specific region, and $p(V|m)$ is the likelihood function of supervoxels V given a primitive shape:

$$p(V|m) = \prod_{t=1}^{N_m} p(v_t|m) \tag{2}$$

The likelihood of individual supervoxels v_t given a primitive shape m is defined as a probability density estimation

$$p(v_t|m) = \frac{1}{N_m} \sum_{i=1}^{N_m} k(v_t.normal|m_i.normal), \tag{3}$$

,where N_m is the number of voxels sampled in shape m, and k is a Gaussian kernel.

To match primitive shapes irrespective of their pose, we consider multiple orientations for the primitive shapes and vote for the most likely shape and orientation for a specific region.

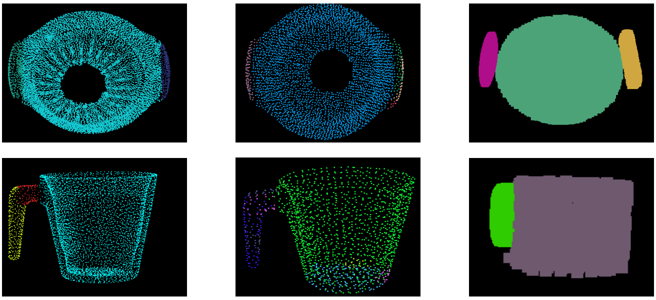


Fig. 3. Object-shape decomposition results on two example objects. Left: hand-labeled object parts. Center: Detected object connected regions. Right: primitive shapes superimposed onto the objects.

2.3 Experiments

We evaluated our approach on a dataset of *IKEA* kitchen objects. The dataset consists of 36 objects from 13 different categories (mugs, plates, bowls, etc). The input data is a point cloud obtained from the mesh file of the objects. Figure 3 shows the result of applying our method to two example objects [2]. The connected regions with the shape estimated are shown as well. We evaluated the overlap between the regions found by our method and hand-labeled parts in the objects. Quantitative results over the whole dataset provided a promising 78.3% overlap. An overlap error is asserted if an estimated part does not fully cover the corresponding hand-labeled part, or if an estimated part partially belongs to more than one hand-labeled part.

There were two main sources of these errors. First source comes from our initially estimated supervoxels. If these supervoxels already contain non-flat regions, this affects downstream steps of our method. Secondly, our surface curvature estimation depends on robust surface normal estimation. Poor estimates may result in a wrong surface shape estimation. For instance, a flat surface might be estimated as a non-flat surface. This will indicate disconnectivity between adjacent patches and hence the object will be decomposed into more regions.

3 Conclusion

We presented an unsupervised approach for 3D object decomposition into parts based on the existence of certain primitive shapes in an object. Our shape estimation is non-parametric and scale-invariant. These properties provide robustness in terms of approach-generality for multiple prototypes and also efficiency. Our results show that this is a promising method applicable to previously-unseen objects. This can be used in grasping applications to transfer grasps from primitive shapes to unseen objects.

[2] Due to the space limitation only two examples are shown

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2013-2016 (Specific Programme Cooperation, Theme 3, Information and Communication Technologies) under grant agreement no. 600918, PaCMan.

References

1. Burl, M., Leung, T.K., Perona, P.: Face localization via shape statistics (1995)
2. Burl, M.C., Weber, M., Perona, P.: A probabilistic approach to object recognition using local photometry and global geometry (1998)
3. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. Int. J. Comput. Vision 61(1), 55–79 (Jan 2005)
4. Papon, J., Abramov, A., Schoeler, M., Wrgtter, F.: Voxel cloud connectivity segmentation - supervoxels for point clouds. In: IEEE Conference on Computer Vision and Pattern Recognition CVPR (June 2013)
5. Sander, P.T., Zucker, S.W.: Inferring surface trace and differential structure from 3-d images. IEEE Trans. Pattern Anal. Mach. Intell. 12(9), 833–854 (1990)
6. Trucco, E., Verri, A.: Introductory Techniques for 3-D Computer Vision. Prentice Hall PTR (1998)