# Automatic Layered 3D Reconstruction of Simplified Object Models for Grasping

Lucian Cosmin Goron[1] and Zoltan Csaba Marton[2] and Gheorghe Lazea[1] and Michael Beetz[2]
[1] Technical University of Cluj-Napoca, Dorobantilor 71-73, 400609 Cluj-Napoca, Romania
{lucian.goron, gheorghe.lazea}@aut.utcluj.ro
[2] Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Deutschland
{marton, beetz}@cs.tum.edu

## Abstract

Most grasping systems require complete CAD models as input, but building these models automatically is not straightforward. We propose a strategy for approximating complete models of previously unknown objects from a single 3D scan of a cluttered scene. We achieve this by exploiting common symmetries in objects of daily use, and by handling over- and under-segmentation issues. The idea is to use 3D shape primitives by breaking up objects into parts than can be approximated by boxes and cylinders. To validate the reconstruction, we present a set of experimental results using real-world data-sets containing a large number of objects from different scan-points.

## 1 Introduction

We are focusing on a method to obtain complete 3D representations for typical objects found in a kitchen table cleaning scenario. The robot has to locate the table in a 3D point cloud, segment the objects that are on the table, and produce a complete 3D representation that can be used to compute grasping points. Our approach is different from similar research initiatives in the sense that we do not use databases containing predefined models in combination with machine learning classifiers to address the object reconstruction problem. Instead we are generating simplified surface models by taking advantage of shape symmetries. These surfaces can then be checked for validity, at an area unit level of desired size, by exploiting the known position of the viewpoint, and used directly for grasp planning if accepted. Our first approach was presented in [1], and in this paper we are investigating the efficient decomposition of objects into layers that, when combined, approximate the original object more accurately and facilitate the solving of the segmentation problems that occur in cluttered scenes. To demonstrate the applicability of our approach, we make use of a mobile manipulation platform from Figure 1 to acquire 3D point clouds, and show reconstruction results for different sets of unseen objects located on table planes.

We are limiting ourselves to cylindrical and box shapes and their combination in order to reconstruct the occluded parts of the objects and to ease the decision process in resolving over- and under-segmentation. Also, generating grasps is straightforward for these shapes, and they are easily parameterized given the shape's extends. Most objects of daily use have a vertical plane or axis of symmetry (e.g. mugs, boxes, bottles, jars, plates, pans, bowls, silverware, etc.), and can be approximated locally using planar and cylindrical parts. We are performing this decomposition on multiple layers to be able to model the surface better, and the verification of each surface unit provides a fine-grained solution in dealing with concavities or model inaccuracies. To exploit the vertical symmetries we are analyzing the footprints of the objects on their supporting plane and on the bases of the layers parallel to it. Lines and circles are identified in the connected components using RANSAC and merged into rectangles and circles.

**Figure 1.** B21 robotic platform equipped with a SICK LMS 400 laser mounted on PowerCube arms.

The main contributions of our approach are:

- exploiting common symmetries to produce complete and simplified models of novel objects from a single 3D scan for (parameterized) grasping applications;
- a strategy for overcoming the 3D over- and under-segmentation of objects produced by occlusion, clutter and measurement errors;
- a novel decomposition technique into adaptively sized layers for a better approximation of objects' surfaces;
- improved model fitting by using more general and robust methods than proposed previously.

The outline of the proposed method is shown in Figure 2, which presents the results of the clustering, layering and model fitting methods. Please see Section 3 for a more detailed description and Figure 7 for more results.

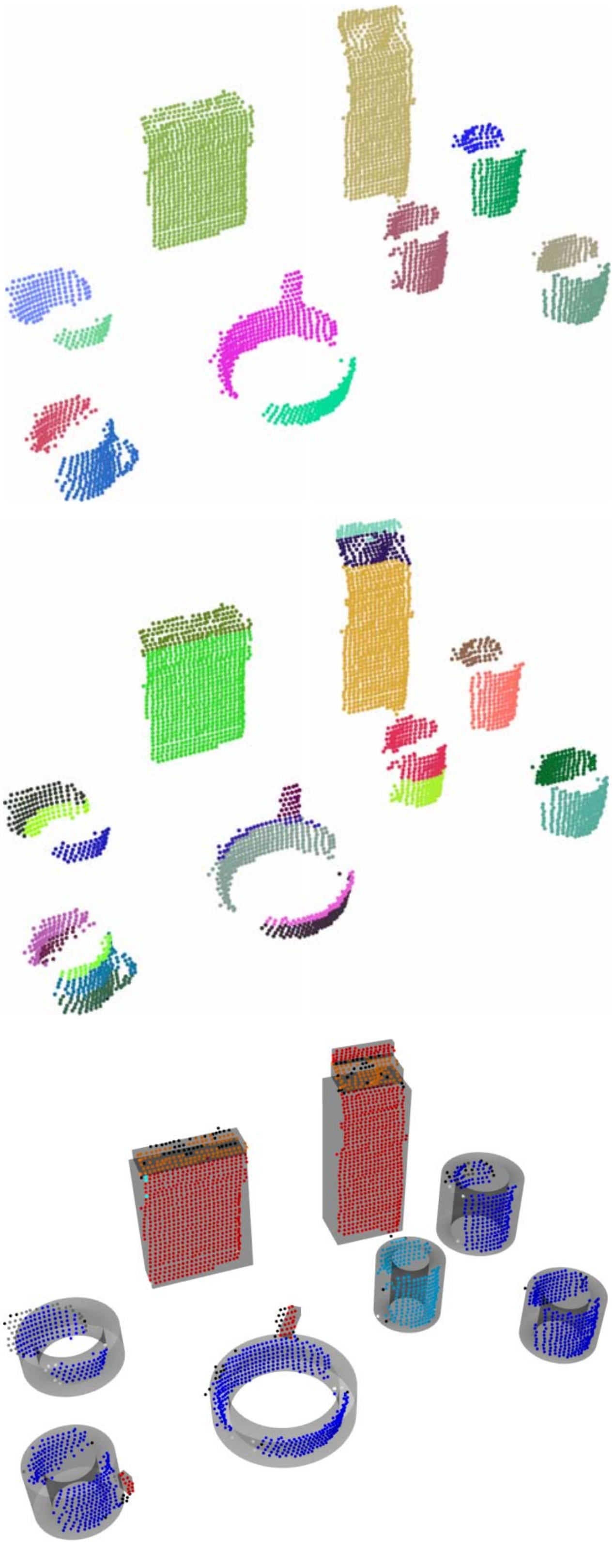

**Figure 2.** The basic steps of the algorithm illustrated from top to bottom: Extracting connected clusters of points from on top of a table's plane, adaptive layering of each region, and reconstruction using shape primitives while dealing with over- and under-segmentation as well as satisfying conditions for physical stability.

The remaining of the paper is organized as follows. In the next section an overview is given on the current approaches, followed by the detailed presentation of our method in Section 3. The experimental results are analyzed in Section 4, followed by our conclusions in Section 5.

## 2 Related Work

Usual tasks of a household assistant robot include pick and place subtasks. The generally used grasping paradigms, like [2] and [3], require a simple closed-surface representation of the objects, which is usually obtained manually using CAD or similar software. To produce these representations from sensor data rely on image or depth information. Typically a model from a database is matched to the camera image [4, 5], or a combination of shape primitives [6] and/or superquadrics [7, 8] are fitted to the 3D or 2.5D data. Alternatively, the surface is triangulated [9, 10], which is the most accurate representation of the sensed data but unfortunately produces too complicated models to be handled by the grasp planning algorithms. A hybrid approach is presented in [11], but reconstruction of boxes is not included. We are expanding this work and address the problem of connectivity in all the mentioned approaches, based on our previous work [1].

Work on locating previously known objects in the current scene is presented in images [12] and range data [13]. Grasping of objects modelled in the 3D object modelling centre [14] composed of a digitizer, a turntable and two cameras was presented in [15]. There are around 40 highly detailed, high-precision objects currently available. While such a system gives high quality data, it is extremely expensive and time consuming to build models of each object and it cannot be used for the online detection phase. Acquiring these models automatically from the Internet has been explored in [5], but our approach on the other hand is to model previously unknown objects on the fly, its advantages over similar approaches being presented in [1].

## 3 Reconstruction of Kitchen Table Scenes

The first stage of our method is to locate the table and the object hypotheses that are on in, as presented in our previous work [1]. The system takes a single view scan of a table scene as input, and extracts the table together with the objects that are on it. The data is cleaned using a statistical analysis of point densities, and clustering is performed to segment the different objects into regions. These regions are then sliced adaptively into layers to which later on models will be fitted. The models then undergo a correction stage where the method deals with over- and under-segmentation (i.e., when an object is split into multiple regions or when objects are clustered together). In the end the shapes are reconstructed and validated. In the next subsections we present the aforementioned stages in detail.

## 3.1 Layered Decomposition

After the points which are above the table are obtained, the sparse outliers are removed using a statistical analysis of point densities in each point's neighbourhood. The dense points are then projected along the table's normal vector to obtain the regions, which are then grouped based on a connectivity criterion [1].

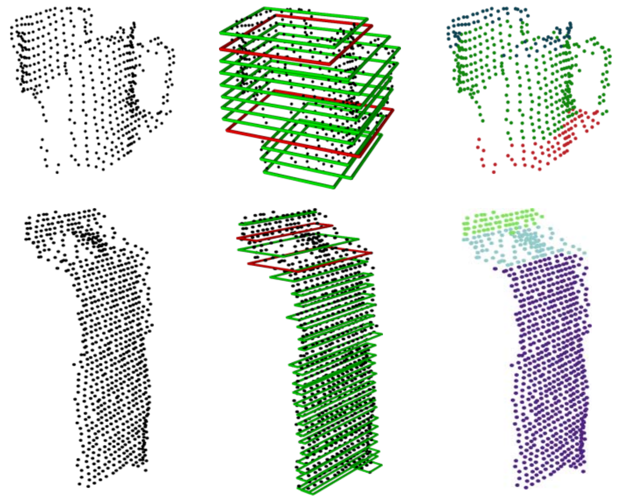

**Figure 3.** Example of a mug and a tetra pack decomposed into layers, from left to right, by columns: a) objects' point clouds from a single view scan; b) sliced regions into thin layers where red rectangles represent jumps; c) decomposed regions into adaptive layers.

Next stage is the decomposition of each region in adaptively sized layers, computed vertically along the Z-axis of every object. The method starts by slicing every object into step-layers with their height according to the scanning resolution. In our real-world data-sets this values is somewhere between $5 - 7.5\ [mm]$. To ensure a reliable decomposition without inconsistencies, the resolution is set to $10\ [mm]$. Then the bounding rectangles are computed along the principle components of the object hypothesis, and the method looks for jumps in the area of successive layers that are greater than a given threshold. In our case the threshold is approximated at $0.001\ [m^2]$ and the object is cut into the final layers where the jumps are taking place. After detecting the jumps the method will correct the layers by merging smaller ones so that in the end we will obtain adaptive layers which can be approximated with box and models. Please take a look at Figure 3.

In the 2D projection of each layer a set of boundary points are indentified, i.e. the points with a maximum angle between the vectors pointing towards their neighbours that angle matches or exceeds the opening of what would be a straight line, $180^{\circ}$ that is. In the sense of providing for every layer a contingent set of boundaries, their neighbours are also marked [1]. Now the method can start fitting 2D models to these points.

## 3.2 Hierarchical Model Fitting

The detection of shape primitives will be performed in 2D space using the boundaries identified earlier. In order to locate the vertical planes and cylinders of objects the method will fit lines and circles by the means of RANSAC. After the fitting of a line and a circle, whichever model has the most inliers is considered the best approximation and ends up being selected.

As detailed in our previous work [1] a circle can approximate a rectangle much easier than a single line. Therefore it would be of great use to know which regions, or better said layers, are quadrilateral so that only line models would be accepted. Until recently, this was the reason why we used a quadrilateral approximation technique based on the principle component analysis. By improving the model fitting and the circle decomposition procedures, the quadrilateral approximation became redundant.

Another improvement is that after the selection of a model, region growing is performed to its inliers. In the case of lines, the points from the biggest cluster are validated as inliers, and in the case of circles, decomposition occurs if its inliers are clustered in more than one or two regions. So circles might end up being checked against a set of parallel and perpendicular lines. The circle model is rejected only if a subset of pair-wise parallel and perpendicular lines present more inliers than the circle itself. The same connectivity criterion was used, for the clustering of the models' inliers, as for the objects' region growing.

In the following iterations, the above mentioned steps are repeated for the remaining points, until their number drops below a minimum threshold for which a robust detection cannot be ensured anymore.

## 3.3 Model Correction by Merging and Splitting

To improve the quality of reconstruction the detected shape primitives need to be corrected by the merging and splitting of layers and regions. This way the method is overcoming the over- and under-segmentation problems.

### 3.3.1 Merging Strategy

The over-segmentation problem is inherent in point clouds taken from a single viewpoint, which contain multiple objects which are split into different regions.

Merging can occur between two circles:

- from the same layer
- from different layers of the same region
- and from different regions.

In the first two cases the circles are merged according to (5.1), thus meaning that one of the two circles overlaps the other one half-way through:

$$max(r_1, r_2) \geq d, \qquad (5.1)$$

where $r_1$ and $r_2$ are the radius values and $d$ is the distance between the centers. The general idea being that inliers of 2D circles which present a high overlap are likely to belong to the same 3D cylinder. By merging them the reconstruction is simplified and also corrected. The weighted average (5.2) is computed for the two models, where the weights are the number of inliers of each circle:

$$\begin{aligned} cx_0 &= (cx_1 n_1 + cx_2 n_2)/(n_1 + n_2) \\ cy_0 &= (cy_1 n_1 + cy_2 n_2)/(n_1 + n_2) \\ r_0 &= (r_1 n_1 + r_2 n_2)/(n_1 + n_2) \end{aligned} \tag{5.2}$$

where $(cx_i, cy_i, r_i)$ is the 2D centre and radius, while $n_i$ are the number of inliers for the one approximated and the two initial circles respectively. Keep in mind that a re-fit with RANSAC would be redundant, because a search for circles was already performed using the same points, and yielded the two models separately.

However, in the last case the merge is conditioned by (5.3), in the sense of overlapping one with another:

$$r_1 + r_2 \geq d, \tag{5.3}$$

where $r_1$, $r_2$ and $d$ are the parameters mentioned above. Here, a re-fit is plausible since their combined inliers were not already considered by the sample consensus algorithm, returning accurate results even for small overlaps.

Merging also can occur between lines that form parallel and/or perpendicular groups inside a layer, since they are most probably part of an object with quadrilateral features for which the boundaries were identified.

### 3.3.2 Splitting Technique

Splitting is required for those scenes where objects are very close positioned or even touching. These objects are inevitably clustered together and layered as an independent region. As a consequence, the detected models are most likely to have inliers located on more than one object, which is resulting in a faulty reconstruction (Figure 4.b).

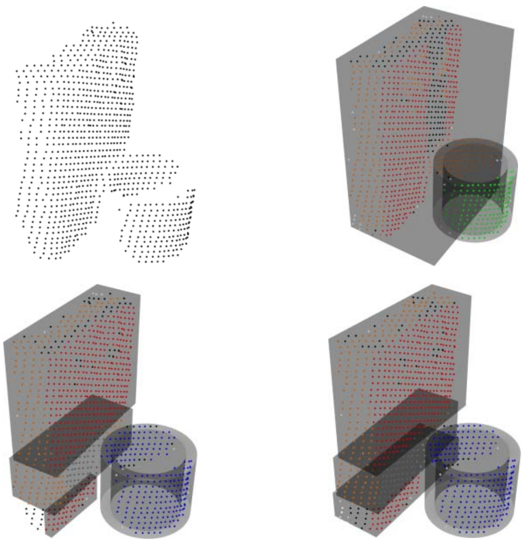

**Figure 4.** Dealing with under-segmentation, from left to right, top to bottom: a) two closely positioned objects which end up being clustered together; b) before using the splitting strategy; c) after using the splitting strategy; c) expanding the shapes in the occluded areas.

So, after the merging procedures are done, the method is inspecting every 2D model, line or circle, and its corresponding 3D shape, box or cylinder, for inconsistencies regarding to their inliers. The first step in splitting regions is to identify which points are located inside the 3D space of two different shapes. The indentified points are then assimilated by only one shape's model. Since a circle can be broken up into quasi-linear parts, line models will always yield in front of circles in our case. These circles were already identified based on solid evidence, thus if other points are found lying on them they are most probably belonging to the same object.

## 3.4 Reconstruction and Validation

The corrected lines and circles are now transformed into boxes and hollow cylinders respectively, using the minimum and maximum heights of their inliers, together with the distances of the inliers to the model as thickness.

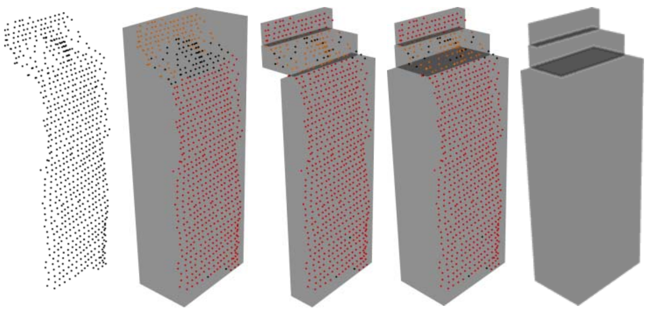

**Figure 5.** Example of reconstruction using box shapes, from left to right: a) scanned tetra pak; b) reconstruction before using layered decomposition; c) reconstruction after using the layering procedure; d) expanding shapes in the occluded areas; e) CAD-like model used for grasping.

For some boxes the reconstruction is limited by the single viewpoint scanning combined with the decomposition into layers, as in Figure 5.c. The method is dealing with this side-effect of layering by expanding the boxes in the occluded parts of the observed volume on the assumption that this will make the resulting model of the object physically more stable. The strategy is to stretch a box within the 2D space to the limits of the shapes, so that the lower one is holding the upper one, as it can be seen in Figure 5.d. This is applied iteratively to each box starting from top to bottom, inside every region. Just by visual comparison the difference in reconstruction is significant.

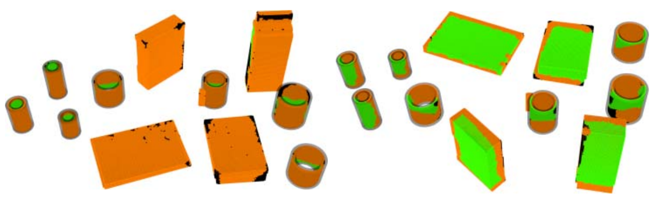

**Figure 6.** Verification of object models viewed from the upper-front (a) and lower-back (b). The orange points are marked as verified, green ones as invisible and the black points as void.

In order to validate these models, we generate grid points with a user defined resolution on the sides of the boxes and on the surface of the cylinders. These points are either *verified* by measurement, points that are *visible* from the current viewpoint, or points that are *void*, meaning that they are in visible free space, thus the models of the objects should not include those parts of the shapes [1]. Please see Figure 6 for an example.

# 4 Experimental Results

The results for the reconstruction of different table scenes can be seen in Figure 7. The colour coding used in the second column of Figure 7 is as follows: *red* and *green* points represent the lines and respectively the circles inliers; *light blue* means that circles from different layers were merged; *dark blue* means the merging of regions; and *orange* are the inliers of the lines which decomposed a circle.

# 5 Conclusions

We have presented a method to automatically reconstruct 3D objects using simplified approximating models that can be used in parameterized grasping applications. The objects are broken up in different layers to improve the fitting, and methods are presented to tackle the over- and under-segmentation problems.

While this approach will not produce the most fine-grained representations, the different parts of the model can be validated, and it offers a good compromise for fast grasp planning in the case of previously unknown objects.

# 6 Literature


[1] Marton, Z.C., Goron, L.C., Rusu, R.B., Beetz, M.: *Reconstruction and Verification of 3D Objects Models for Grasping.* In: Proceedings of the 14th International Symposium on Robotics Research (ISRR), Lucerne, Switerland (2009)

[2] Miller, A., Allen, P.K.: *Graspit!: A Versatile Simualtor for Robotic Grasping*. IEEE Robotics and Automation Magazine, vol. 11, no. 4 pp. 110-122 (2004)

[3] Harade, K., Kaneko, K., Kanehiro, F.: *Fast Grasp Planning for Hand/Arm Systems Based on Convex Model*. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, pp. 1162-1168 (2009)

[4] Collet, A., Berenson, D., Srinivasa, S.S., Ferguson, D.: *Object Recognition and Full Pose Registration from a Single Image for Robotic Manipualtion*. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan (2009)

[5] Klank, U., Zia, M.Z., Beetz, M.: *3D Model Selection from an Internet Database for Robotic Vision*. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan (2009)

[6] Schnabel, R., Wahl, R., Klein, R.: *Efficient RANSAC for Point-Cloud Shape Detection*. Computer Graphics Forum, vol. 26, no. 2, pp. 214-226 (2007)

[7] Biegelbauer, G., Vincze, M.: *Efficient 3D Object Detection by Fitting Superquadrics to Range Image Data for Robot's Object Manipulation*. In: Proceedings of the IEEE International Conference Conference on Robotics and Automation (ICRA), Rome, Italy (2007)

[8] Zhang, Y., Koschan, A., Abidi, M.: *Superquadric Representation of Automotive parts Applying Part Decomposition*. Journal of Electronic Imaging, Special Issue on Quality Control by Artificial Vision, vol. 13, no. 4, pp. 411-417 (2004)

[9] Marion, R., Markus, V.: *Grasping of Unknown Objects from a Table Top*. In: Workshop on Vision in Action: Efficient strategies for cognitive agents in complex environments (2008)

[10] Marton, Z.C., Rusu, R.B., Beetz, M.: *On Fast Surface Reconstruction Methods for Large Noisy Datasets*. In: Proceedings of the IEEE Conference on Robotics and Autoamtion (ICRA), Kobe, Japan (2009)

[11] Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M.: *Close-range Scene Segmentation and Reconstruction of 3D Point Clouds Maps for Mobile Manipulation in Human Environments*. In: Proceedings of the 22nd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, USA (2009)

[12] A. Collet, D. Berenson, S. S. Srinivasa, and D. Ferguson: *Object Recognition and Full Pose Registration from a Single Image for Robotic Manipulation*. In Proc of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, 2009.

[13] B. Steder, G. Grisetti, M. Van Loock, and W. Burgard: *Robust Online Model-based Object Detection from Range Images*. In Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), St. Louis, MO, USA, October 2009.

[14] *Kit object models web database* [Online]. Available: http://i61p109.ira.uka.de/ObjectModelsWebUI/

[15] Z. Xue, A. Kasper, J. Zoellner, and R. Dillmann, *An automatic grasp planning system for service robots*. In Proc. of the International Conference on Advanced Robotics, 2009. ICAR 2009. International Conference on, Munich, Germany, June 2009, pp. 1–6.

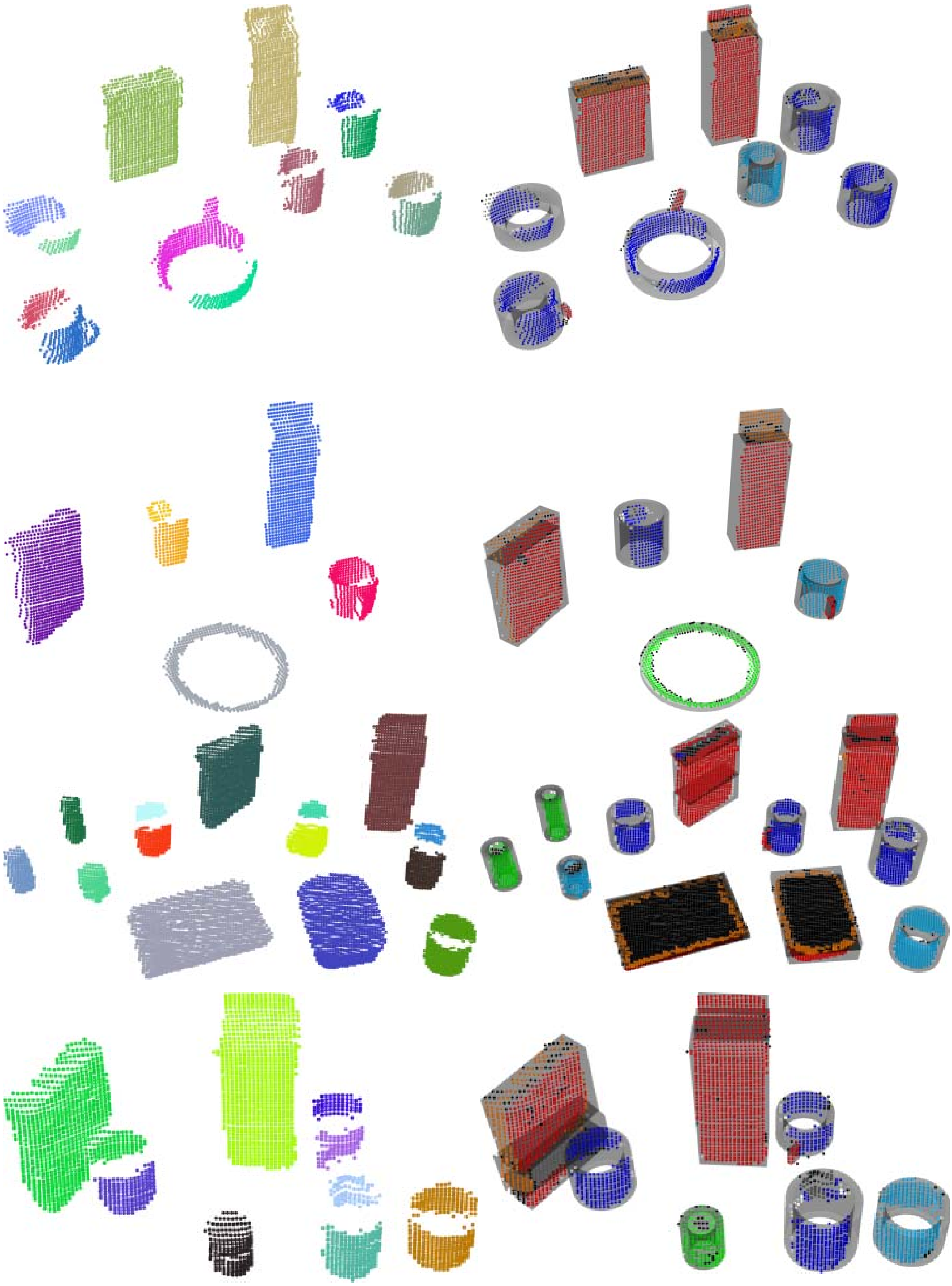

**Figure 7.** Left: clustering of the points located above the table's plane. Right: automatic layered reconstruction using boxes and cylinders.