Efficient object search on a mobile robot using semantic segmentation

Wolf D., Bajones M., Prankl J., Vincze M.
Vienna University of Technology, Automation & Control Institute
{wolf, bajones, prankl, vincze}@acin.tuwien.ac.at

Abstract

We propose an efficient semantic segmentation framework for indoor scenes, tailored to the application on a mobile robot. In that scope, a segmentation method especially needs to be fast and robust. We developed a 3D point cloud processing framework based on Randomized Decision Forests, achieving competitive results at sufficiently high frame rates. We show our method's capabilities on two datasets and also applied it on a mobile robot in order to develop an intelligent object search procedure.

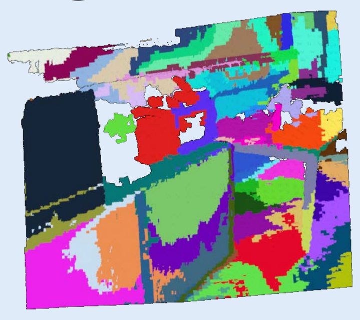
Segmentation pipeline

Input Point Cloud

RGB-D Point Cloud

→

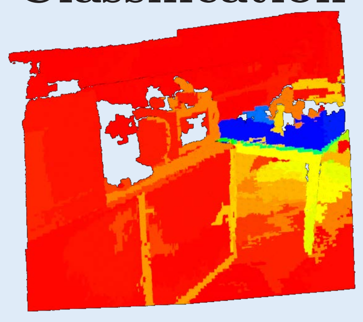
Over-segmentation



Supervoxel Clusters

→

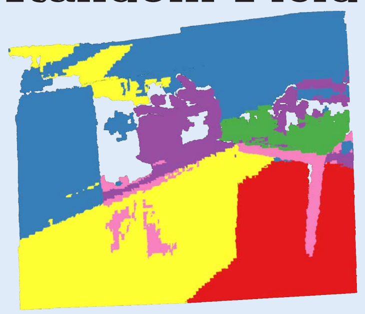
Feature Extr. Classification



Conditional Label Probabilities

→

Markov Random Field



Final Labeling

Oversegmentation

- Using Supervoxel Clustering by Papon et al. [2].
- Takes color, surface normals and geometric regularity into account.
- Point cloud subsampled with 1 cm Octree resolution (for speed, accuracy sufficient for application).
- Maximum cluster size: 50 cm

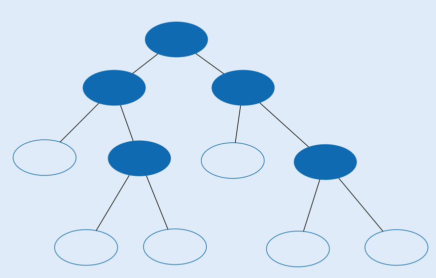
Features

An efficient-to-compute feature vector is calculated for each cluster:

Feature Type	**Dim.**
Angular (mean normal angle & std. dev. w.r.t. ground plane)	2
Height (centroid, top and bottom point)	3
Spatial (point-, linear- and surfaceness)	3
Color (CIELAB mean values and std. dev.)	6
Total Dimensionality	**14**

Classification

We use a *Randomized Decision Forest (RDF)* [3], trained on the popular NYU Depth Dataset v2 [4], to classify the feature vector of each cluster. RDFs have some **key advantages** for our task:

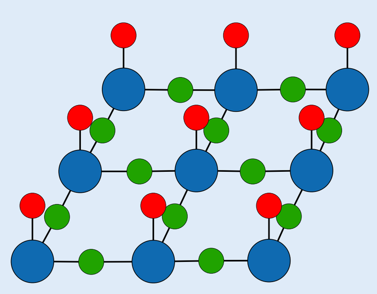


- They can cope with a large variety of different feature types and ranges.
- Training and classification is highly parallelizable for speed.
- As the classification result, they return the conditional label probabilities, which can in turn be used in the next step.

Markov Random Field

The final labeling problem can be formulated as a Conditional Random Field (CRF), where each cluster is a node $i \in \mathcal{V}$ in an interconnected graph. The connections are given by the adjacency graph $\mathcal{A}$ obtained from the clustering. The optimal label configuration then corresponds to the minimum of the energy function:

$$E(\mathcal{Y}) = \underbrace{\sum_{i\in\mathcal{V}} \phi_i(y_i)}_{\text{unary term}} + \underbrace{\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{A}_i}\phi_{i,j}(y_i,y_j)}_{\text{pairwise term}}$$



- Unary term: Models conditional label probabilities for each point; **obtained from RDF**.
- Pairwise term: Smoothes the result by assigning **less energy on consistent labelings of adjacent points**.
- Optimization solved by Loopy Belief Propagation.

Results

Dataset/Method		Floor	Wall	Ceiling	Table	Chair	Cabinet	Object	Avg.	Global
NYU v2	[6]	87.3	**86.1**	62.6	10.2	34.1	-	8.7	-	-
	Our	**96.9**	75.0	**92.3**	**59.7**	**72.2**	**58.4**	**40.0**	**71.7**	**77.2**
Our data	Our	**98.0**	**87.8**	-	**92.2**	**62.8**	**31.5**	**16.4**	**55.6**	**72.0**

First row: input point clouds; *second row:* groundtruth; *third row:* results; *first three columns:* NYU v2; *last two columns:* our dataset.

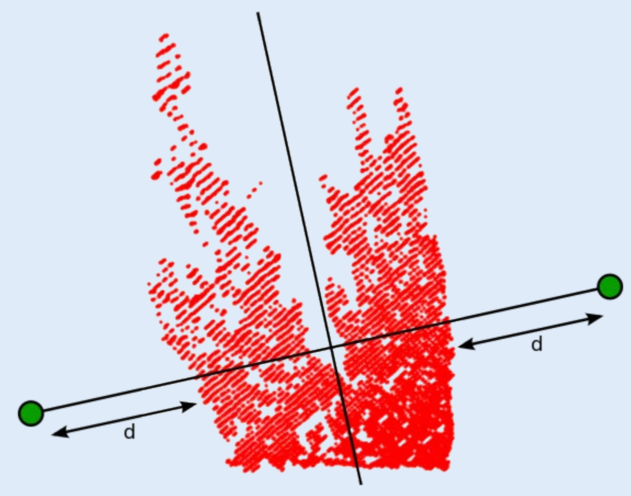
Processing Time: approx. 1 second per point cloud (640x480 pixels)
→ Most time spent for oversegmentation (≈500 ms) and MRF (≈300 ms).

Object search with a mobile robot

The HOBBIT project: Development of a socially assistive robot that helps the elderly at home [5].

One important functionality of the robot is the *Bring Object* scenario, in which HOBBIT searches for a specific object and brings it to the user. As graspable objects are most likely placed on top of planar surfaces, e.g. tables, our semantic segmentation approach enables the robot to dynamically find good locations to search for objects:

1. Cluster all points labeled *table*.
2. Project clusters to ground plane.
3. Calculate principal components of clusters and place search positions for the robot on the second principal axis (see Fig. right).
4. Eliminate unreachable search positions.
5. Rank search positions and start search.



More details about the dynamically adapted object search scenario can be found in [7].

Acknowledgements

This work has been partially funded by the European Commission under grant agreements FP7-IST-288146 HOBBIT, FP7-IST-600623 STRANDS and the Austrian Science Foundation (FWF), Project I513-N23 vision@home.

References

[1] D. Wolf, M. Bajones, J. Prankl and M. Vincze. Find my mug: Efficient object search on a mobile robot using semantic segmentation. In *ÖAGM*, 2014.

[2] J. Papon, A. Abramov, M. Schoeler, and F. Wörgötter. Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds. In *CVPR*, 2013.

[3] A. Criminisi and J. Shotton, editors. Decision Forests for Computer Vision and Medical Image Analysis. *Springer London*, 2013.

[4] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In *ECCV*, 2012.

[5] D. Fischinger, P. Einramhof, W. Wohlkinger, K. Papoutsakis, P. Mayer, P. Panek, T. Koertner, S. Hofmann, A. Argyros, M. Vincze, A. Weiss, and C. Gisinger. Hobbit - The Mutual Care Robot. In *IROS*, 2013.

[6] C. Couprie, C. Farabet, L. Najman, and Y. LeCun. Indoor Semantic Segmentation using depth information. In *ICLR*, 2013.

[7] M. Bajones, D. Wolf, J. Prankl, and M. Vincze. Where to look first? Behaviour control for fetch-and-carry missions of service robots. In *ARW*, 2014.