

STRUCTURAL SEGMENTATION OF POINT CLOUDS WITH VARYING DENSITY BASED ON MULTI-SIZE SUPERVOXELS

Yuan Li, Bo Wu*

Dept. of Land Surveying & Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
(y-uan.li@connect.polyu.hk, bo.wu@polyu.edu.hk)

KEY WORDS: Mobile laser scanning, Point clouds, Structural segmentation, Multi-size supervoxels, Varying point density, MRF

ABSTRACT:

Ground objects can be regarded as a combination of structures of different geometries. Generally, the structural geometries can be grouped into linear, planar and scatter shapes. A good segmentation of objects into different structures can help to interpret the scanned scenes and provide essential clues for subsequent semantic interpretation. This is particularly true for the terrestrial static and mobile laser scanning data, where the geometric structures of objects are presented in detail due to the close scanning distances. In consideration of the large data volume and the large variation in point density of such point clouds, this paper presents a structural segmentation method of point clouds to efficiently decompose the ground objects into different structural components based on supervoxels of multiple sizes. First, supervoxels are generated with sizes adaptive to the point density with minimum occupied points and minimum size constraints. Then, the multi-size supervoxels are clustered into different components based on a structural labelling result obtained via Markov random field. Two datasets including terrestrial and mobile laser scanning point clouds were used to evaluate the performance of the proposed method. The results indicate that the proposed method can effectively and efficiently classify the point clouds into structurally meaningful segments with overall accuracies higher than 96%, even with largely varying point density.

1. INTRODUCTION

Mobile laser scanning (MLS) and static terrestrial laser scanning (TLS) point clouds can intuitively present the three-dimensional (3D) geometric characteristics of ground objects with abundant details. In general, these detailed structures of objects can be grouped into linear, planar and scatter shapes that can be described by the eigenvalues derived from the covariance matrix of local neighbourhoods (Hackel et al., 2016; Landrieu et al., 2017; Liu and Boehm, 2015; Weinmann et al., 2014, 2017; Yang et al., 2015). Accurate computation of the shape descriptors is essential for both the structural segmentation and semantic classification of terrestrial obtained point clouds (Qiao et al., 2010). However, as the point densities of such point clouds always vary dramatically with the scanning distance changing, a local neighbourhood with optimal size should be necessary for accurate measurement of shape characteristics. For instance, in the work of Weinmann et al. (2017), the three shape descriptors corresponding to linearity, planarity and scattering were computed point-wisely from the local optimal neighbourhoods defined by a minimum entropy function (Demantke et al., 2011). Similarly, in addition to the three shape descriptors, Landrieu and Simonovsky (2018) also computed the verticality descriptor of points from the local optimal neighbourhoods and then partitioned the points into superpoints based on the four local descriptors through global energy optimization via ℓ_0 -cut (Landrieu and Obozinski, 2017). However, due to the great data volume of the point clouds, the point-wise feature computing and labelling processes are time consuming and require vast amounts of memory space for establishing the fundamental graphical model.

To efficiently compute the shape descriptors and decompose the objects into different structural components, supervoxels (Lin et al., 2018; Papon et al., 2013) can be introduced as the primitives for structural segmentation. In fact, supervoxels have recently been widely used for the interpretation of large-scale scenes (Dong et al., 2018; Kang and Yang, 2018; Luo et al., 2018; Zhu et al., 2017) for three reasons. Firstly, the local homogeneity can be well preserved within the supervoxels, secondly, supervoxels provide explicit adjacent relationships rather than vague neighbouring relationships among unordered points, and thirdly, exploiting supervoxels instead of individual points can significantly reduce the computation load and time. However, many state-of-the-art supervoxel segmentation methods generate supervoxels with fixed resolutions, such as the widely adopted voxel cloud connectivity segmentation (VCCS) (Papon et al., 2013) and this will lead the supervoxels located in areas with sparse densities to contain insufficient points for accurate feature computation. In order to generate supervoxels with adaptive resolutions, Lin et al. (2018) proposed a toward better boundary preserved (TBBP) method that formalized the supervoxel segmentation as a subset selection problem. Instead of supervoxel resolution, this method used the number of supervoxels as a direct constraint to select representative points, so that the final supervoxels would be free of size constraint and well preserve the object boundaries. However, the adaptability to the boundaries of this method also causes a problem that the supervoxels are quite sensitive to linear features and weak at presenting the scatter properties of objects.

To generate supervoxels that are adaptive to varying point density and, simultaneously, have better presentations of the

* Corresponding author

local geometric properties of different shapes, this paper proposed a coarse-to-fine supervoxel segmentation method based on an octree structure, which is built under constraints about the minimum number of points occupied by a supervoxel and the minimum resolution of supervoxels. Then three shape descriptors corresponding to the linear, planar and scatter properties of each supervoxel are computed and used to establish an energy function. Instead of an unsupervised labelling process as in Landrieu and Simonovsky (2018), we adopt a supervised labelling strategy by using Markov random field (MRF) (Li, 2009) and graph-cut algorithm (Boykov et al., 2001), so that each supervoxel will be assigned with a geometrically meaningful label. Adjacent supervoxels with the same labels are finally clustered into the same structural components through region growing. Two experiments with different datasets were carried out to evaluate the performance of the proposed method and discussions and conclusions were drawn at last based on the experimental results.

2. STRUCTURAL SEGMENTATION OF POINT CLOUDS USING MULTI-SIZE SUPERVOXELS

2.1 Overview of the Approach

As shown in Figure 1, the proposed method includes two steps. In the first step, the input MLS point cloud is partitioned into multi-size supervoxels through a coarse-to-fine seed selection process and a supervoxel expansion process, followed by recovering the adjacent relationships between supervoxels. In the second step, three shape descriptors corresponding to linearity, planarity and scattering, are firstly computed and used to build an MRF model. The global energy of the MRF is optimized to obtain a spatially smooth structural labelling configuration for all the supervoxels. Finally, a region growing algorithm is developed to merge the adjacent supervoxels with the same structural labels into a larger segment, called structural components. The ground objects are finally presented by connected structural components with different geometric characteristics.

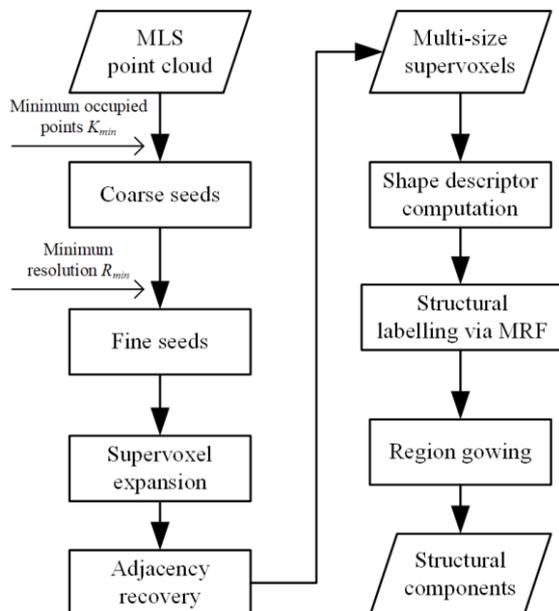


Figure 1. Overview of the proposed method.

2.2 Generation of Multi-size Supervoxels

The multi-size supervoxel segmentation is an extension of the VCCS (Papon et al., 2013). But unlike VCCS, which selects supervoxel seeds with a unified resolution, we selected seeds from multi-scale resolutions with a coarse-to-fine strategy. After the selection of seeds with multi-scale resolutions, we expand the supervoxels using a similar distance metric as VCCS. However, instead of expanding the supervoxels based on an adjacency octree used by VCCS, we expand the supervoxels based on the k-NN search to guarantee that the adjacent relationships will not be interrupted by varying point densities. During the expansion, the adjacencies between supervoxels are also determined by the neighbouring relationships between occupied points.

2.2.1 Coarse-to-fine Seed Selection: Given a point set $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$, we would like to generate a set of supervoxels $\mathcal{V} = \{V_1, V_2, \dots, V_m\}$, where each $V_i = \{p|p \in \mathcal{P}\}$ ($V_i \in \mathcal{V}$) should contain sufficient points for local feature computation. Simultaneously, the resolution of V_i should not be too small so that it can effectively capture the local geometries at an appropriate scale. In other words, there are two constraints for the supervoxel sizes. First, $K_i \geq K_{min}$, K_i is the number of points contained in V_i , and $R_i \geq R_{min}$, R_i is the euclidean size (the maximum size of the bounding box) of V_i (K_{min} and R_{min} are user-defined thresholds, for example $K_{min} = 20$, $R_{min} = 0.3$ m). With these two constraints, we propose the following coarse-to-fine seed selection method.

First, based on the constraint K_{min} , the maximum seed resolution R_{max} can be defined as $R_{max} = \max\{\max\{d(p, \mathcal{N}_p)\}|p \in \mathcal{P}\}$, where \mathcal{N}_p is the K_{min} nearest neighbors of p . Then, with R_{max} a coarse octree is built and the coarse seeds $Seed_0$ are selected as the points in \mathcal{P} that are closest to the centroids of the leaf nodes in the coarse octree, as shown in Figure 2.

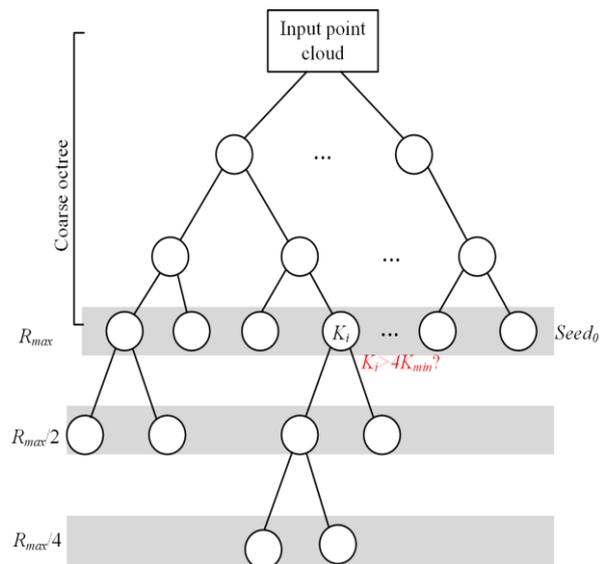


Figure 2. Coarse-to-fine seed selection based on an octree structure.

We then traverse all the coarse seeds in $Seed_0$. If the number of occupied points K_i corresponding to a coarse seed i is larger

than $4K_{min}$, this coarse seed will be further split into deeper octree structures with a corresponding resolution of $R_{max}/2^n$ (n is the difference between the current octree depth and the depth of the coarse octree). The new seeds are then selected as the points in \mathcal{P} closest to the centroids of the new leaf nodes. The traversing will be repeated until all the leaf nodes in the octree contain points less than $4K_{min}$ or the supervoxel resolution difference $|R_{max}/2^n - R_{min}|$ stops getting reduced.

2.2.2 Supervoxel Expansion and Adjacency Recovery:

After the coarse-to-fine seed selection, a set of seeds $Seed = \{s_1, s_2, \dots, s_m\}$ with corresponding multiple resolutions $\mathcal{R} = \{r_1, r_2, \dots, r_m\}$ is obtained. At the same time, each seed points s_i corresponds to a set of points $\{p\} \subseteq \mathcal{P}$ occupied by the corresponding octree leaf node.

The supervoxels are then iteratively expanded using a similar strategy as in VCCS (Papon et al., 2013). In each iteration, the K nearest neighbors of point p_{new} , which is newly assigned to a supervoxel in last iteration, will be assigned to their nearest supervoxels based on a distance metric, until there are no points left. The distance metric $D(p, V)$ is a combination of the distances in coordinate, colour (if exists) and normal (previously estimated through principle component analysis) spaces. But as the distances between supervoxel seeds selected by our method vary with point density, we define the coordinate distance measurement as

$$d(p_i, V_j) = \frac{d_{Eu}(p_i, c_j)}{r_j} \quad (1)$$

where r_j is the corresponding seed resolution of V_j , c_j is the centroid coordinates of V_j , and d_{Eu} is the euclidean distance between p_i and c_j .

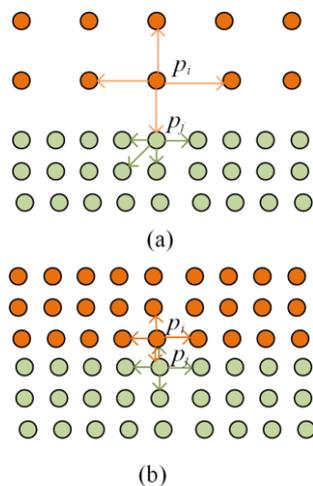


Figure 3. Recovery of adjacency between supervoxels and points. The red and green points refer to two different supervoxels. (a) shows two points are not adjacent because p_i is not the K_{min} nearest neighbours of p_j and (b) shows two adjacent points that are mutually in each other's K_{min} nearest neighbourhoods.

In each iteration, the K-NN search is adopted during the expansion of the supervoxels. For two points $p_i \in V_m$ and $p_j \in V_n$, if p_i and p_j are adjacent points, expansion operation (reassigning p_j to V_m) will be performed between V_m and V_n if $D(p_i, V_m) > D(p_i, V_n)$. p_i and p_j are only regarded as adjacent when they are mutually in each other's K_{min} nearest neighbors and their corresponding supervoxels are also regarded as adjacent as illustrated in Figure 3. If two points are adjacent but assigned to different supervoxels, the corresponding supervoxels will also be considered as adjacent.

2.3 Generation of Structural Components

2.3.1 Shape Descriptors Computation: The local geometric characteristics of object structures can be roughly grouped as linearity, planarity and scattering and they can be measured using the eigenvalues derived from the covariance matrix by principle component analysis (Jolliffe, 2011). Different measurement methods of the shape descriptors based on eigenvalues are found in the work of Weinmann et al. (2014), Hackel et al. (2016) and Yang et al. (2015), but it is found that the method of Yang et al. (2015) has better ability of presenting scatter points. Therefore, in this work, the three shape descriptors f_l , f_p and f_s corresponding to the three structural categories are computed as defined in the work of Yang et al. (2015) that have the following format.

$$\begin{cases} f_l = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}} \\ f_p = \frac{\sqrt{\lambda_2} - \sqrt{\lambda_3}}{\sqrt{\lambda_1}} \\ f_s = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}} \end{cases} \quad (2)$$

where $\lambda_1 > \lambda_2 > \lambda_3$ are eigenvalues from the covariance matrix of the supervoxel.

2.3.2 Structural Labelling via MRF: The structural partition of the supervoxels can be regarded as a labelling problem, during which each supervoxel will be assign a structural label $\mathbf{y} \in \{\textit{linearity}, \textit{planarity}, \textit{scattering}\}$. To make the structural labelling spatially smooth, a graphical model is established to encode the adjacent relationships between supervoxels into the graph edges. Let \mathbf{x} be the set of nodes that corresponds to the supervoxels, and let e be the set of edges that corresponds to the adjacencies between supervoxels. An energy function based on the graphical model is given as

$$\min_{\mathbf{y}} E(\mathbf{y}) = \sum_{i \in \mathbf{x}} \Phi(\mathbf{y}_i) + \gamma \sum_{(i,j) \in e} \Psi(\mathbf{y}_i, \mathbf{y}_j) \quad (3)$$

where \mathbf{y} is a labelling configuration whose value space is $\Omega = \{\textit{linearity}, \textit{planarity}, \textit{scattering}\}$, $i \in \mathbf{x}$ corresponds to a supervoxel, $(i, j) \in e$ corresponds to the adjacency between supervoxels and γ is a constant parameter used to adjust the effectiveness of pairwise interactions.

The unary potentials presenting the fidelity to the geometric measurements are given as

$$\Phi(\mathbf{y}_i) = \begin{cases} 1 - f_l, & \text{if } \mathbf{y}_i = \text{linearity} \\ 1 - f_p, & \text{if } \mathbf{y}_i = \text{planarity} \\ 1 - f_s, & \text{if } \mathbf{y}_i = \text{scattering} \end{cases} \quad (4)$$

And the pairwise terms defined based on the Potts model (Li, 2009) are given as

$$\Psi(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{1}[\mathbf{y}_i \neq \mathbf{y}_j] \quad (5)$$

where $\mathbf{1}[\cdot]$ is a binary function that equals 0 if $\mathbf{y}_i = \mathbf{y}_j$, otherwise, 1.

The minimum global energy $\min_{\mathbf{y}} E(\mathbf{y})$ is approximated using Graph Cut with α -expansion operations (Boykov et al., 2001) to obtain spatially smooth structural labelling \mathbf{y} .

2.3.3 Clustering of Supervoxels with Structural Labels: After structural labelling, we then merge the supervoxels assigned with the same structural labels into a large segment, called structural components, by region growing (Vosselman et al., 2004). As each structural component is composed by supervoxels with the same labels, the local geometric characteristics on the component surface will be also consistent.

3. EXPERIMENTAL RESULTS

3.1 Test Data Description

Two point cloud datasets were used to evaluate the performance of the proposed method. The first one is a benchmark dataset Semantic3D (Hackel et al., 2017) (<http://www.semantic3d.net/>). The Semantic3D point cloud dataset is acquired by a static TLS system and has very high but relatively consistent point density. The second one is an MLS dataset acquired by an UltraCam Mustang mobile mapping system, which was equipped with a multi-beam rotating light detection and ranging system and a high-resolution camera. The MLS dataset presents a quite complex scene with various objects and features with a strongly varying point density.

3.2 Structural Segmentation Results

The structural segmentation results of the two datasets with the proposed method are shown in this section and the results are also compared to the results based on supervoxel generated by other methods, including the VCCS (Papon et al., 2013) and TBBP (Lin et al., 2018). For both of these two datasets, the minimum point size K_{min} is set as 20 and the minimum seed resolution R_{min} is set as 0.3 m, which is the same with the fixed resolution in VCCS and the expected resolution in TBBP. With respect to the labelling process, the smooth parameter γ is set as 0.3 in all experiments. Below shows the segmentation results and the comparisons between different methods with the TLS and MLS datasets, respectively.

3.2.1 Results of TLS dataset (Semantic3D): Figure 4 shows the comparison between segmentation results based on supervoxels generated by different methods. Generally, for this dataset, the supervoxels generated by our method are quite similar to those generated by VCCS with fixed resolution. This is because the TLS point cloud has a very high point density and this would lead to the iterative coarse-to-fine seed selection, most of the time, stops due to the octree depth constraint, which means the most of the final supervoxel seeds have the similar resolutions approximate to R_{min} . As this dataset has a quite high point density, for most of the supervoxels generated by VCCS, the numbers of contained points would be similar with those contained in the supervoxels generated by our method, which will lead to the similarity between the final structural segmentation results.

It is also found that, compared to our method and the VCCS, the supervoxels generated from TBBP are more sensitive to linear features. The TBBP method does not constrain the sizes of supervoxels so that it can well preserve the boundaries, however, scatter structures are likely to be presented by a set of narrow supervoxels by this method as the trees shown in the black circle in Figure 4(d). Therefore, while both of the proposed method and the TBBP can generate supervoxels with adaptive resolutions, the proposed method has better ability of presenting structural with different shapes, especially scatter shapes.

The structural labelling results based on supervoxels generated by different methods were also point-wisely compared with the manually labeled ground-truth (as shown in Figure 4(b)). Table 1 shows the structural labelling accuracy ($\text{accuracy} = \text{number of correctly labeled points} / \text{number of points in ground-truth}$) corresponding to different shape categories as well as the overall accuracy. The approximate accuracies of VCCS and our method are consistent with the above analysis that our method and the VCCS generated similar supervoxels. But generally, our method has better performance than VCCS. The high linearity and low scattering accuracies of the TBBP method indicates that supervoxels generated by TBBP could be sensitive to linear features and weak for scatter shapes.

	VCCS	TBBP	Ours
Linearity	20.9	58.7	22.8
Planarity	97.9	94.4	98.4
Scattering	80.6	8.0	85.4
Overall accuracy	94.9	87.8	96.7

Table 1. Labelling accuracy (%) of the TLS test data based on supervoxels generated by different methods.

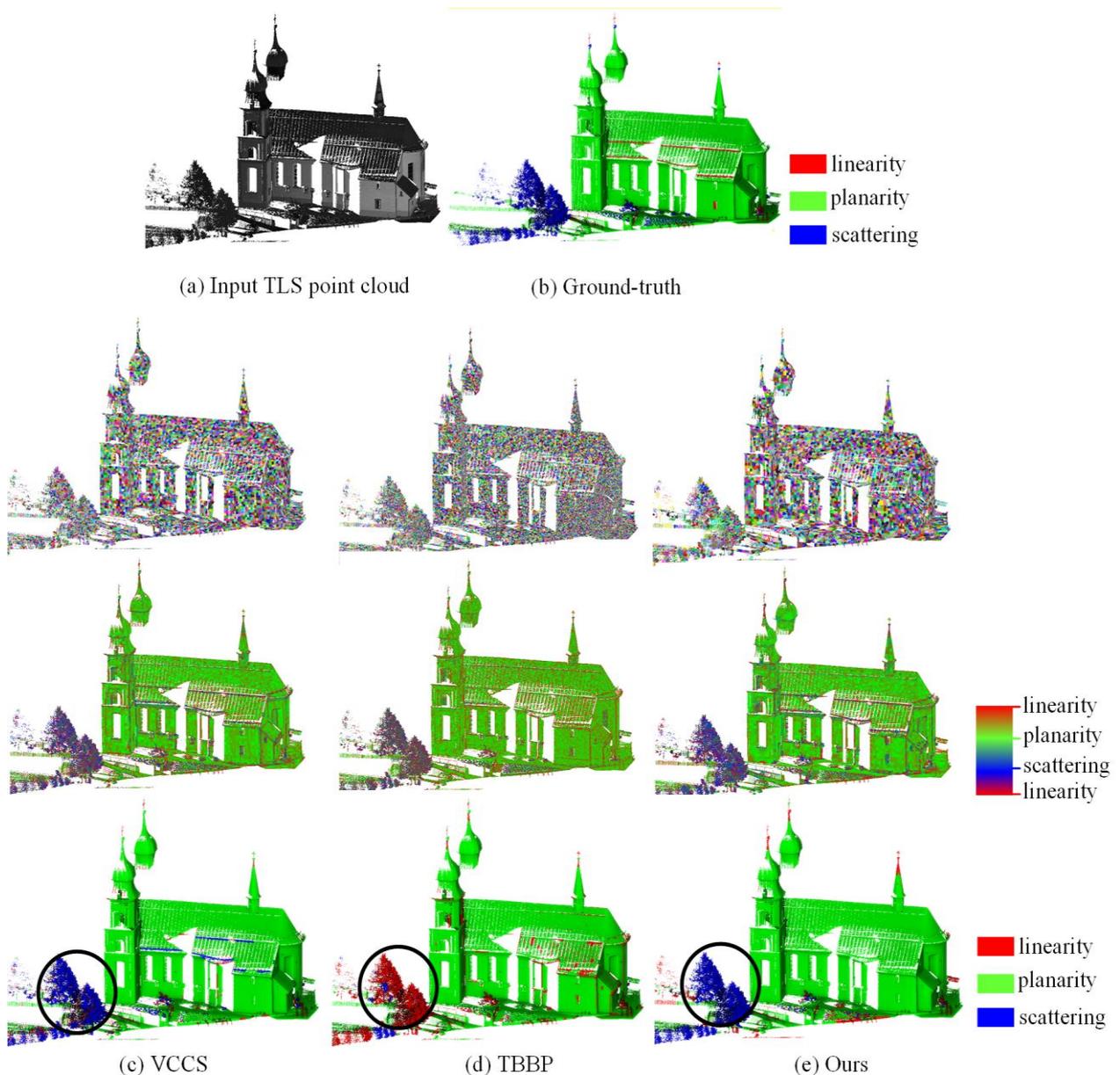


Figure 4. Comparison of structural segmentation results of the TLS dataset. (a) is the input TLS point cloud, (b) is the manually labeled ground-truth with red, green and blue referring to linearity, planarity and scattering, respectively. (c) – (e) show the segmentation results derived supervoxels generated by VCCS, TBBP and our method. In (c) – (e), from top to bottom are the random colour-coded supervoxels, visualized shape descriptors derived from supervoxels and the structural components colour-coded based on their labels.

3.2.2 Results of MLS dataset: Figure 5 shows the segmentation results of the MLS dataset based on supervoxels generated by different methods. Compared to the TLS dataset, the point density of the MLS data is much lower but varies much more strongly. Due to the great variation of MLS point density, supervoxels with quite different sizes were generated by our method and the TBBP method. Unlike the VCCS supervoxels with fixed resolution, the adaptive resolutions guarantee the supervoxels located in areas with sparse density (such as the building façade edge far away from the scanner, as shown in the black circles in Figure 5) are valid for feature computation.

The results of the MLS data set were also compared with the manually labelled results (as shown in Figure 5(b)) and the point-wise evaluation results are shown in Table 2. Compared with the VCCS, our method had higher accuracies on linear and planar structures and had a comparative accuracy on scatter structures. Although both the proposed method and the TBBP method generated supervoxels with adaptive resolutions, the TBBP method once again showed its weakness at presenting scatter structures (such as the tree crowns shown in the black circles in Figure 5) as indicated by the low accuracy shown in Table 2, while our method could effectively decompose objects into different parts with reasonable structural labels.

	VCCS	TBBP	Ours
Linearity	65.3	96.7	85.8
Planarity	97.2	94.4	97.9
Scattering	95.7	34.0	95.6
Overall accuracy	95.9	85.3	96.9

Table 2. Labelling accuracy (%) of the MLS test data based on supervoxels generated by different methods.

3.3 Efficiency Evaluation

The efficiency of the proposed method is evaluated and compared with those based on different supervoxel segmentation methods, as well as the point-wise structural segmentation strategy. In the point-wise segmentation, the features are computed with $K = 20$ nearest neighbors for each individual point and the graphical model is built based on points and their adjacent relationships, which are defined as described in Section 2.2.2.

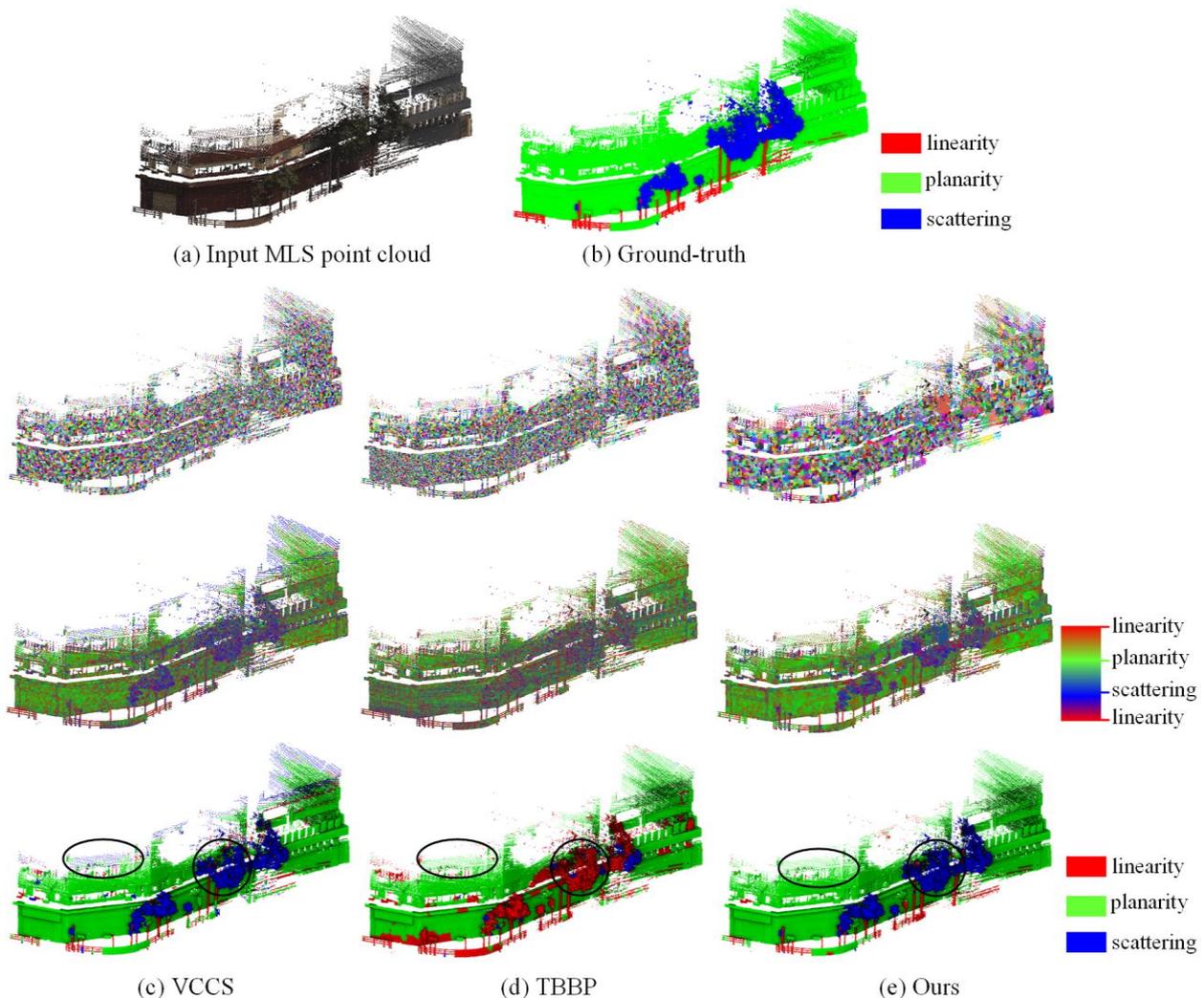


Figure 5. Comparison of structural segmentation results of the MLS dataset. (a) is the input MLS point cloud, (b) is the manually labeled ground-truth with red, green and blue referring to linearity, planarity and scattering, respectively. (c) – (e) show the segmentation results derived supervoxels generated by VCCS, TBBP and our method. In (c) – (e), from top to bottom are the random colour-coded supervoxels, visualized shape descriptors derived from supervoxels and the structural components colour-coded based on their labels.

As the point-wise labelling requires a large amount of memory, which can be out of the maximum memory of the used computer, we only chose a subset of the test TLS point cloud with about 5 million points to perform the evaluation. Table 3 shows the time consumed by different methods of the same subset test data. All the methods were implemented in C++ with a single thread in order to allow a fair comparison. The time consumed by preprocessing steps, such as normal estimation,

required in the VCCS, TBBP and our method, was included in the supervoxel generation stage in Table 3.

From Table 3 it can be seen that using supervoxels instead of individual points can significantly reduce the time consumed by feature computation and structural labelling. With respect to the time cost by supervoxel generation, although both our method and TBBP cost more time than the VCCS, as more complicated

seed (or representative points) selection strategies were adopted, supervoxels with adaptive sizes to the varying point density can be obtained by our method and TBBP. Simultaneously, compared with the TBBP, our method is considerably efficient in supervoxel generation.

	Point-wise	VCCS	TBBP	Ours
Supervoxel generation	-	34.4 s	50 s	43.8
Feature computation	85 s	0.5 s	0.7 s	0.4
Labelling	1247 s	1.5 s	2.5 s	0.7
Region growing	4568 s	4.9 s	6.0 s	2.3
Total	5900 s	41.3 s	59.2 s	47.2 s

Table 3. Time consumed (in seconds) of different methods with a subset of the TLS test data containing about 5 million points.

4. CONCLUSIONS

In this paper, we propose an effective structural segmentation method for MLS and TLS point clouds, which present objects with abundant details and features with varying point density. The proposed method consists of two stages. The first stage produces supervoxels whose sizes are adaptive to the variations in point density using a coarse-to-fine seed selection strategy. In the second stage, a spatially smooth structural label configuration was found the multi-size supervoxels through global energy minimization. The supervoxels were finally clustered into structural meaningful segment based on their corresponding labels and adjacent relationships. The performance of the proposed method is evaluated in terms of both effectiveness and efficiency with two different datasets. The experimental results indicate the proposed method has better ability of decomposing objects based on their local geometric characteristics even with strongly varying point density than the existing methods.

ACKNOWLEDGEMENTS

This work was supported by grants from the Hong Kong Polytechnic University (Project Nos. 1-ZE24, 1-ZEAB, and 1-ZVN6) and a grant from the National Natural Science Foundation of China (Project No: 41471345). The authors gratefully acknowledge the provision of the Semantic3D datasets and the *Hong Kong* dataset by Ambit Geospatial Ltd.

REFERENCES

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence* 23, 1222-1239.

Demantke, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based scale selection in 3D lidar point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, W12.

Dong, Z., Yang, B., Hu, P., Scherer, S., 2018. An efficient global energy optimization approach for robust 3D plane segmentation of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 137, 112-133.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. Semantic3D. net: A new large-scale point cloud classification benchmark. *arXiv Preprint arXiv:1704.03847*.

Hackel, T., Wegner, J.D., Schindler, K., 2016. Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 3(3).

Jolliffe, I., 2011. Principal component analysis, *International Encyclopedia of Statistical Science*. Springer, pp. 1094-1096.

Kang, Z., Yang, J., 2018. A probabilistic graphical model for the classification of mobile LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 143, 108-123.

Landrieu, L., Obozinski, G., 2017. Cut pursuit: Fast algorithms to learn piecewise constant functions on general weighted graphs. *SIAM Journal on Imaging Sciences* 10, 1724-1766.

Landrieu, L., Raguét, H., Vallet, B., Mallet, C., Weinmann, M., 2017. A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 132, 102-118.

Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4558-4567.

Li, S.Z., 2009. Markov random field modeling in image analysis. Springer Science & Business Media, pp. 29-39.

Lin, Y., Wang, C., Zhai, D., Li, W., Li, J., 2018. Toward better boundary preserved supervoxel segmentation for 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 143: 39-47.

Liu, K., Boehm, J., 2015. Classification of big point cloud data using cloud computing. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 553-557.

Luo, H., Wang, C., Wen, C., Chen, Z., Zai, D., Yu, Y., Li, J., 2018. Semantic labeling of mobile LiDAR point clouds via active learning and higher order MRF. *IEEE Transactions on Geoscience and Remote Sensing*, 56(7), 3631-3644.

Papon, J., Abramov, A., Schoeler, M., Worgotter, F., 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2027-2034.

Vosselman, G., Gorte, B.G., Sithole, G., Rabbani, T., 2004. Recognising structure in laser scanner point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 46, 33-38.

Qiao, G., Wang, W., Wu, B., Liu, C., Li, R., 2010. Assessment of Geo-positioning Capability of High-Resolution Satellite Imagery for Densely Populated High Buildings in Metropolitan Areas. *Photogrammetric Engineering & Remote Sensing*, 76(8), 923-934.

Weinmann, M., Jutzi, B., Mallet, C., 2014. Semantic 3D scene interpretation: a framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, 181.

Weinmann, M., Jutzi, B., Mallet, C., 2017. Geometric features and their relevance for 3d point cloud classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4, 157.

Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 99, 45-57.

Zhu, Q., Li, Y., Hu, H., Wu, B., 2017. Robust point cloud classification based on multi-level semantic relationships for urban scenes. *ISPRS Journal of Photogrammetry and Remote Sensing* 129, 86-102.