

CLASSIFICATION OF MLS POINT CLOUDS IN URBAN SCENES USING DETRENDED GEOMETRIC FEATURES FROM SUPERVOXEL-BASED LOCAL CONTEXTS

Z. Sun, Y. Xu,* L. Hoegner, U. Stilla

Photogrammetry and Remote Sensing, Technische Universität München, 80333 Munich, Germany
-(zhenghao.sun, yusheng.xu, ludwig.hoegner, stilla)@tum.de

Commission II, WG II/3

KEY WORDS: Classification, MLS, detrended geometric features, supervoxel, local context

ABSTRACT:

In this work, we propose a classification method designed for the labeling of MLS point clouds, with detrended geometric features extracted from the points of the supervoxel-based local context. To achieve the analysis of complex 3D urban scenes, acquired points of the scene should be tagged with individual labels of different classes. Thus, assigning a unique label to the points of an object that belong to the same category plays an essential role in the entire 3D scene analysis workflow. Although plenty of studies in this field have been reported, this work is still a challenging task. Specifically, in this work: 1) A novel geometric feature extraction method, detrending the redundant and in-salient information in the local context, is proposed, which is proved to be effective for extracting local geometric features from the 3D scene. 2) Instead of using individual point as basic element, the supervoxel-based local context is designed to encapsulate geometric characteristics of points, providing a flexible and robust solution for feature extraction. 3) Experiments using complex urban scene with manually labeled ground truth are conducted, and the performance of proposed method with respect to different methods is analyzed. With the testing dataset, we have obtained a result of 0.92 for overall accuracy for assigning eight semantic classes.

1. INTRODUCTION

In the past decade, automatic 3D scene analysis using LiDAR point clouds has been a challenging task in research fields of photogrammetry (Vosselman and Maas, 2010), remote sensing (Lefsky et al., 1999), computer vision (Buch et al., 2011), and robotics (Rusu et al., 2009). As a generally utilized data type, LiDAR point clouds can be acquired through different acquisition techniques such as terrestrial laser scanning (TLS), mobile laser scanning (MLS), and airborne laser scanning (ALS). ALS datasets are usually used for large scale scene description and analysis, with a relatively low point density. However, for dense and accurate 3D scene analysis and interpretation, especially in the context of urban areas, TLS and MLS, which have higher scanning density and more stable carrier platform (e.g., static scanning stations of TLS), are considerably more reliable.

Point clouds obtained via TLS, which have normally a high point density and a corresponding high spatial resolution, show a great potential of being used as datasets for interpreting 3D scenes in urban areas. However, the utilization of this kind of datasets also meets some problems in the meantime. For example, due to the fixed measuring station of one single scan, there are massive differences of point density between regions, resulting from the observation distance from the targets to measurement center. As a contrast, possessing an ultra-high point density and spatial resolution, the point cloud of MLS has a relatively evenly distributed point density, due to the movement of the scanning platform. Furthermore, multiple measurement stations are not mandatory when conducting a mobile laser scanning, so that the complex registration process can be avoided to some extent. Therefore, for further 3D analysis of large scale urban scene, especially for acquiring

3D dataset of street views and building facades, MLS is considered to be the primary choice. However, to achieve the analysis of complex 3D urban scenes, acquired points of the scene should be tagged with individual labels of different classes. Thus, assigning a unique label to the points of an object belonging to the same category plays an essential role in the entire 3D scene analysis workflow. Although plenty of studies in this field have been reported, this work is still a challenging task.

To this end, we propose a classification method designed for the labeling of MLS point clouds, focusing on detrended geometric features extracted from points of supervoxel-based local contexts. The following are the contributions that are specific to this work: 1) A novel geometric feature extraction method, detrending the redundant and in-salient information in the local context, is proposed, which is proved to be effective and efficient for extracting local geometric features from the 3D scene. 2) Instead of using individual point as basic element, the supervoxel-based local context is designed to encapsulate geometric characteristics of points, providing a flexible and robust solution for feature extraction. 3) Experiments using complex urban scene with manually labeled ground truth are conducted, and the performance of proposed method referring to different features is analyzed. In Fig. 1, we illustrate an example of the MLS point cloud we tested and the classified point cloud with different semantic labels.

2. RELATED WORK

Classification as well as semantic labeling of point cloud, which aims at assigning a unique class label to each 3D point of the input point cloud, is a very important issue for urban remote sensing. Typically, the classification of point clouds involves two core steps: feature extraction and semantic classification. Based on

*Corresponding author

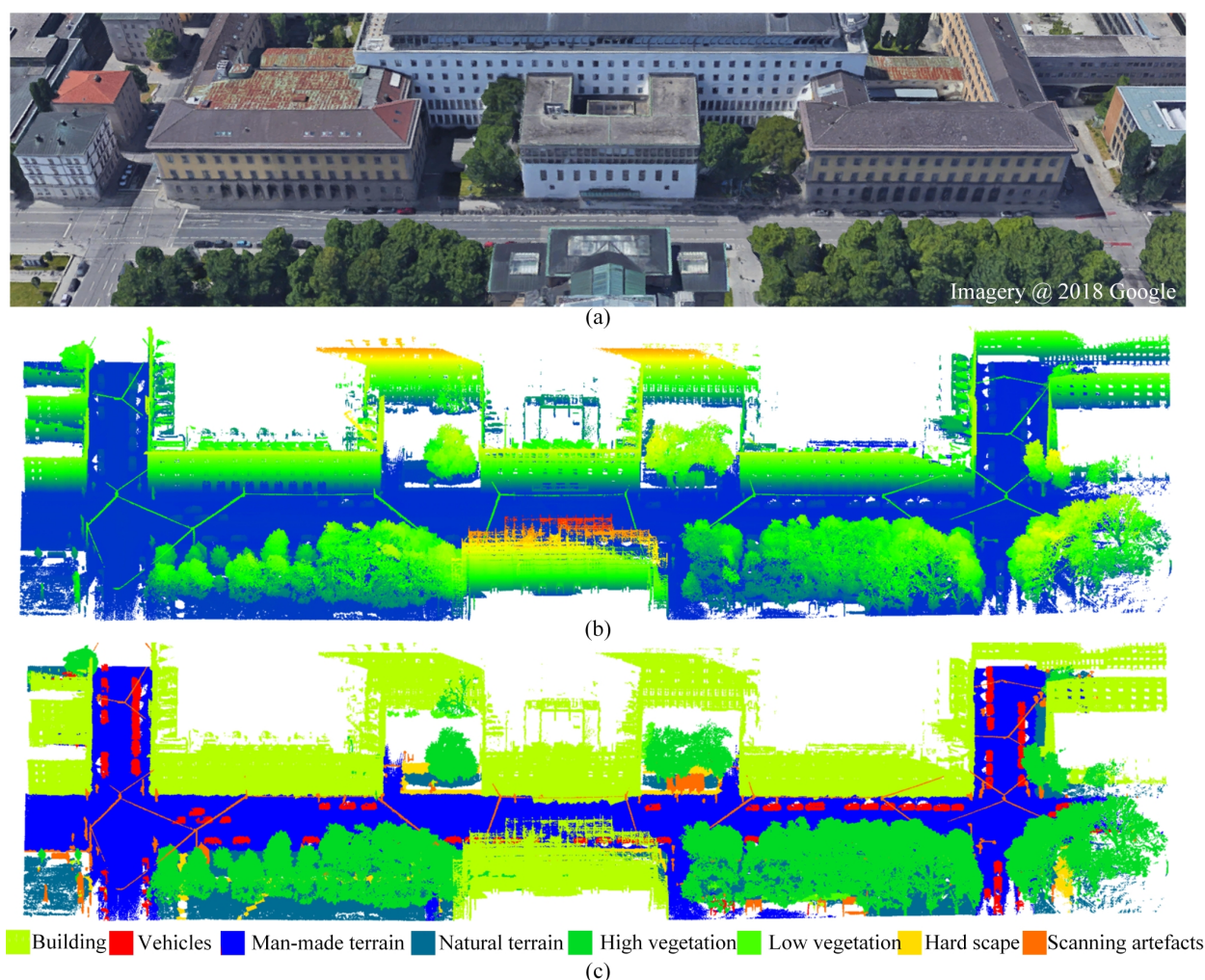


Figure 1. (a) Real scene of the TUM main entrance from Google Maps, 2018. (b) MLS point cloud colored with respect to height. (c) Classified point cloud of the test scene with eight different semantic labels .

the derived various features, classifiers can be applied to assign a label to each point.

2.1 Feature extraction

The extraction of features for further classification is a process of abstracting local geometric information of points within a local neighborhood and encapsulating them into feature vectors (Guo et al., 2014). Generally, there are two influential factors for extracting features: 1) Selection of an appropriate neighborhood for each element (i.e., point) in order to describe local geometric features; 2) Extraction of discriminative features with appropriate descriptors and abstraction of features as feature vectors.

The selection of the neighborhood is indispensable for describing the detailed information around a certain point (Weinmann et al., 2015). For different purposes, various objective details relying on all the points within the chosen neighborhood are required. Commonly used definitions of neighborhoods can be categorized into two types: single-scale neighborhoods and multi-scale neighborhoods. The first one means extracting features from a fixed scale of neighborhoods, while the second one uses flexible scales of neighborhood. Most commonly used single-scale neighborhood definitions are the spherical (Lee and Schenk, 2002) or cylindrical (Filin and Pfeifer, 2006) neighborhood. Furthermore, the neighborhood around a point can be also defined by a fixed number

of k nearest neighbors, in which the distance between two points can be either 3D distance (Linsen and Prautzsch, 2001) or 2D projective distance (Niemeyer et al., 2014). Moreover, Weinmann et al. (2015) propose an approach, which relies on individually optimized neighborhoods in 3D scenes for both feature extraction and contextual classification. Dong et al. (2017) use a feature selection strategy to improve the accuracy of 3D point clouds classification in a multi-scale neighborhood. With various form and size of neighborhoods defined, identical features from multi-scale neighborhoods are separately extracted to implement further feature encoding. Through the different classification performances, features are weighted. Yu et al. (2016) implement a multi-layer feature generation model consisting of various levels of octree partition structure to detect certain object cars. Besides, in (Wang et al., 2015), point-based hierarchical clusters are generated with a Latent Dirichlet Allocation (LDA) model, in which cluster features are derived in order to classify objects of different sizes.

For the extraction of discriminative features, local shape descriptors carrying valuable information of objects are representative solutions of parameterizing features. Apart from the original 3D coordinate, echo and intensity information, there are also possibilities to enrich point attributes for describing features. RGB colors (Al-Manasir and Fraser, 2006) and thermal information

(Weinmann et al., 2013) can be acquired from 3D point clouds. Besides point attributes, the local and global environment should be taken into consideration as well. On the basis of the spatial information of all 3D points, both the global and local features can be calculated. Therefore, an appropriate way of describing these features with mathematical formulations is normally required. To tackle this problem, some representative feature descriptors are developed in recent decades (Yu et al., 2013) such as 3D context shape (Frome et al., 2004), Signature of Histogram of Orientations (SHOT) descriptor and its variants (Salti et al., 2014), and Fast Point Feature Histogram (FPFH) (Rusu et al., 2009). However, since all the mentioned descriptors mainly rely on detailed geometric properties of objects and rarely concentrate on integrated structure features, noise can hardly be resisted, and a robust geometric topological relationship is missing. Consequently, a method aiming at characterizing the geometric structure of objects is proposed. Considering no distribution of detailed geometries and textures but analyzing eigenvalues of 3D coordinates of point clouds with a proper definition of neighborhood, essential geometric and topological relationships are obtained. In this regard, eigenvalue based geometry (Jutzi and Gross, 2009, Chehata et al., 2009) is one of the representatives. Through eigenvalues of the coordinate tensor, 3D characteristics of shapes are characterized.

2.2 Classification using extracted features

There are several typical strategies for the classification task. Point-based classification is one of the classic solutions, in which each point will obtain a label during the classification process (Weinmann et al., 2015, Hackel et al., 2016). In contrast, the segment-based classification, pre-clustering or segmenting the point cloud into primitives with homogeneities (Yao et al., 2011, Yu et al., 2015, Guinard and Landrieu, 2017), is also drawing increasing attentions due to its advantage of separating individual objects from the scene simultaneously. However, no matter what kind of strategy is used, classifiers always play an essential role for the final classification performance. Support vector machines (Lodha et al., 2006, Secord and Zakhori, 2006), adaBoost (Lodha et al., 2007), random forest (RF) (Chehata et al., 2009), conditional random field (Lim and Suter, 2009), and deep neural networks (Vetrivel et al., 2017, Landrieu and Simonovsky, 2017) are the representatives of the commonly used classifiers.

2.3 Voxel-based point clouds classification

Recently, voxel-based data structures are also becoming popular for the point cloud processing instead of using conventional points based data structure to cope with non-uniform point density and large-scale dataset. Voxel structures like octree (Vo et al., 2015) can simplify the dataset and suppress the outliers and noise with rasterized representation. It can also define neighboring relations of generated voxels as well as points within them simultaneously, facilitating the neighbor search. For exploring the potential of the voxel structure, voxels can be further clustered into supervoxels through algorithms like voxel cloud connectivity segmentation (VCCS) (Papon et al., 2013) algorithm. As a consequence, neighboring voxels are grouped together via a local graph so that a supervoxel is generated. Supervoxel structure can better and precisely preserve the boundary position of the segments, tending to an oversegmentation of the complete segments. Considering that the supervoxel structure has already pre-clustered voxels with homogeneous properties such as normal vector, spatial distance or even color information, the edges

of objects are well detected and a supervoxel is thus an appropriate neighborhood for extracting features. Moreover, supervoxels can also be further grouped into local patches considering a given neighborhood (Wang et al., 2015, Yu et al., 2015), by which geometric features in a local vicinity can be delineated in a more complete way.

3. METHODOLOGY

In Fig. 2, a general workflow of our proposed method is given, involving five major steps, namely supervoxelization and selection of local context, segment-based feature extraction, detrending of geometric features, and supervised classification. In the initial step, an over-segmentation process is implemented through the voxel cloud connectivity segmentation (VCCS) (Papon et al., 2013). Besides, for each supervoxel, a local context is defined, taking all the directly connected neighbors into consideration. In the second step, local geometric features of each supervoxel as well as its connected neighbors within the local context are calculated. Afterwards, for each supervoxel, a local tendency is estimated in the feature space based on the features of all the neighboring supervoxels in the local context. Then, the geometric features of the center supervoxel are detrended by the use of the local tendency. Finally, for the supervised classification, through a training stage, a RF classifier is learned to classify objects by the use of the detrended features in complex urban scenes. The resultant classes have eight different objects, covering man-made terrain, natural terrain, high vegetation, low vegetation, buildings, hard scape, scanning artifacts and cars.

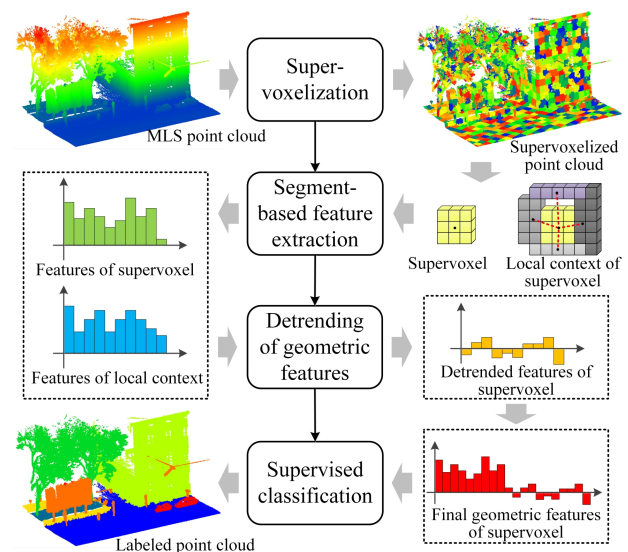


Figure 2. Workflow of our point cloud classification method.

3.1 Supervoxelization and selection of local context

To organize the entire point cloud into a supervoxel structure, the space is firstly divided into a small 3D cubic grid by means of octree partitioning, which splits each node into eight equal child nodes, in order to generate the octree-based voxel structure. Compared with others point-based neighborhoods, for example, kd-tree based points structure, when using voxels as basic processing unit under an octree structure, there is no need to handle problems like uneven density resulting from mobile laser scanning. In fact, the voxelization process can also be regarded as a down-sampling process, so that the computational cost is

drastically reduced. Besides, the octree structure is achieved by the approximate nearest neighbor (ANN) (Muja and Lowe, 2009) searching algorithm, which largely increases the efficiency of the neighbor search procedure. For the further generation of supervoxel structures the VCCS algorithm is adopted to cluster voxels in terms of geometric as well as spectral distance between seed and candidate voxels (Papon et al., 2013). In our work, only normal vectors and spatial distance are considered during supervoxelization, which shows a better performance at preserving real boundary of objects than that implemented at the voxel-level.

Considering that the supervoxelization itself is essentially an over-segmentation process, thus using merely one supervoxel can hardly represent the geometric characteristic of an object. To solve this problem, for each supervoxel, we define a local context to capture the contextual information of each supervoxel. Here, the context is defined by the first order neighbors (Wang et al., 2015), namely the directly connected neighbors of each supervoxel. In Fig. 3, we provide an illustration about the defined local context of the supervoxel. On the basis of voxel structures, supervoxel adjacency can be found in a similar way. Once the neighboring voxels belonging to different clustered supervoxels, the corresponding supervoxels are marked as "directly connected". Considering supervoxels having limited but flexible extension, the numbers of connecting are varying. Supervoxel-based contexts composed of directly connected adjacent supervoxels are similar to a variant of flexible scale of neighborhood.

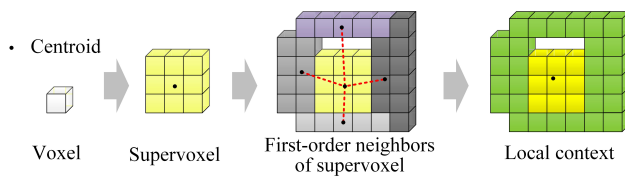


Figure 3. Local context of the supervoxel.

3.2 Segment-based feature extraction

Considering a large amount of the 3D point clouds containing merely spatial coordinates, we focus thus on the geometric features. Since supervoxel neighborhood and its adjacent graph are already aware, an appropriate representation of the local geometry is necessary. Therefore, local 3D shape features (Chehata et al., 2009; Weinmann et al., 2015) are introduced to tackle this problem. The respective derived eigenvalues λ_i with $i \in \{1, 2, 3\}$ within a certain neighborhood can be used to explore and quantize local 3D shape. The eigenvalues are firstly normalized into e_i with $i \in \{1, 2, 3\}$ by

$$e_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3}, i = 1, 2, 3 \quad (1)$$

Then, the linearity L_λ , planarity P_λ , scattering S_λ , omnivariance O_λ , anisotropy A_λ , eigenentropy E_λ as well as local curvature C_λ can be derived according to the method presented in (Weinmann et al., 2015), which are nowadays commonly utilized in 3D point clouds processing. Further more, height, derived normal direction, and intensity are also introduced as additional information for feature extraction. To be specific, in Table 1, we provide the details about the entire feature vectors we used.

3.3 Detrending of geometric features

3.3.1 Local tendency of supervoxel-based context Although supervoxel structure has already pre-clustered voxels at

Features	Definition	Category
Linearity	$L_\lambda = \frac{e_1 - e_2}{e_1}$	Eigen feature (Weinmann et al., 2015) (Chehata et al., 2009)
Planarity	$P_\lambda = \frac{e_2 - e_3}{e_1}$	
Scattering	$S_\lambda = \frac{e_3}{e_1}$	
Omnivariance	$O_\lambda = \sqrt[3]{e_1 \cdot e_2 \cdot e_3}$	
Anisotropy	$A_\lambda = (e_1 - e_3)/e_1$	
Eigenentropy	$E_\lambda = -\sum_{i=1}^3 e_i \ln(e_i)$	
Local curvature	$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3}$	Height-feature (Maas, 1999)
Height mean	$\frac{1}{n} \sum_{i=1}^n Z_i$	
Height difference	$Z_{max} - Z_{min}$	Spatial feature (Rabbani et al., 2006)
Normal vectors	N_x	
	N_y	
	N_z	
Intensity	Mean intensity of all points	Radiometric feature (Aijazi et al., 2013)

Table 1. List of totally used features.

a lower level, supervoxels tend to oversegment objects into fragmented pieces, which results in the dissimilarity between features of different patches belonging to identical object. Hence, the decision trees may not be well trained. To tackle this problem, Wang et al. (2015) utilize the first-order graph around a single supervoxel and generalize this graph into a local reference frame (LRF), which shows an impressive performance at car detection.

However, for the complex 3D scene interpretation, there are usually various kinds of objects to be detected and the accurate boundaries between objects are necessary to be identified in the meantime. Besides, according to the analysis conducted in (Guinard and Landrieu, 2017), for local descriptors, even for the same kind of objects, the contribution of each vector in the generated feature histogram are varying. This will result in ambiguities of the generated features for two different kinds of objects, for example, the natural ground surface and man-made ground surface. Both of these two objects have quite similar geometric characteristics (e.g., linearity, planarity, and normal vectors), and the only obvious difference between them is the smoothness and roughness of their surfaces. For the achieved features histogram, we can conduct a procedure enhancing the useful features vectors with a better saliency and suppressing the trivial feature vectors.

Inspired by the Difference of Gaussian operator for edge detection in the field of image processing, we developed a strategy of estimating the local tendency of 3D geometry in a local context for each supervoxel, and then remove the effect of this local tendency, in order to get the salient information of the objects representing distinctive details and structures. The local tendency of the supervoxel context also plays an essential role at precisely assigning supervoxels near real boundaries of objects semantic labels. In Fig. 4, we show a 1D illustration about the estimation of the local tendency for the geometry of an object. It is clear that after the removal of the local tendency, two geometric shapes with similar structures become more distinguishable.

This operation can also be regarded as an "high-pass" filtering which dislodges background geometric information in a local vicinity and preserve only those "high frequency" components. Considering that the eigenvalue based geometric features essentially reflect the geometric structure of the objects, namely those relatively low frequency components, better distinctiveness can be achieved if we can combine these two kinds of components together for describing the geometry of the objects.

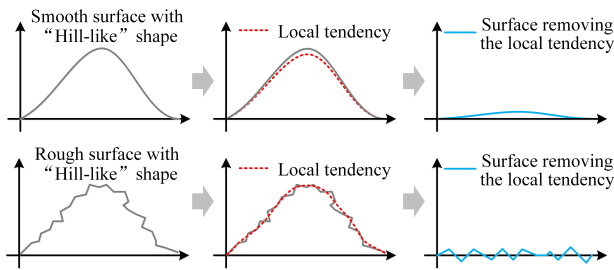


Figure 4. Local tendency of the geometric shapes.

3.3.2 Detrended geometric features The removal of the local tendency of each supervoxel is achieved in the feature space. Here, the feature histogram of the supervoxel SV_0 itself is noted as V_{SV} , while the feature histogram representing the local tendency is given by V_{LT} , which is estimated by all the points in the local context. Thus, the detrended feature histogram V_D is derived by a difference operation:

$$V_D = V_{SV} - V_{LT} \quad (2)$$

The final feature histogram V_F is defined by a weighted combination of the V_{SV} and V_D :

$$V_0 = \{V_{SV}, k \cdot V_D\} \quad (3)$$

Here, k stands for the weight given to the local tendency, which is estimated by the number of supervoxels in the local context. In Fig. 5, an illustration of the detrending of geometric features is given. Finally, a 26 dimensional feature histogram is achieved for supervised classification.

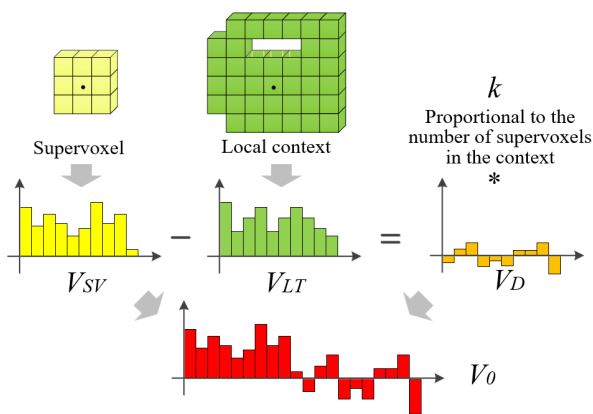


Figure 5. Generation of feature histogram.

3.4 Supervised classification

Once the final geometric features of all the supervoxels in the whole point datasets are calculated, we use a supervised classification strategy with classic RF algorithm (Breiman, 2001) to discriminate supervoxel as well as the points within it with different semantic labels. The RF classifier is a combination of tree-structured classifiers which are created by a randomizing vector sampled independently from input vectors (i.e., feature histogram), and each decision tree will vote for the most likely labels to the sample of input vectors (Breiman, 2001). However, the RF classifier splits at each node with random subset of features which makes it insensitive to overfitting problems due to the strong law

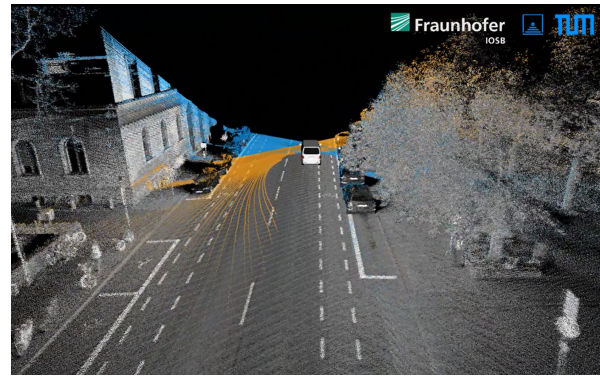


Figure 6. Two oblique mounted laser scanners of the MLS system.

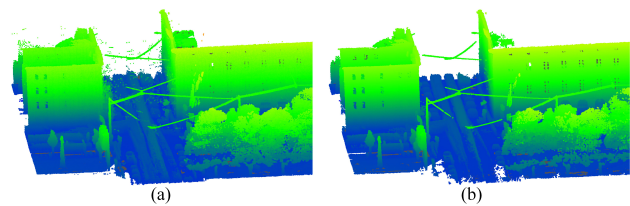


Figure 7. (a) Raw point clouds. (b) Preprocessed point clouds.

of large numbers as the number of trees increases. In the training, the bagging method is used for each feature combination to generate a training dataset by drawing N examples with random replacement, where N is the size of the original training set.

4. EXPERIMENTS

4.1 Datasets

The testing area is the Arcisstrasse along the main entrance of Technical University of Munich (TUM) city campus, which covers about an area of around $29000 m^2$ and has been already displayed in Fig. 1a. This dataset is original acquired by Fraunhofer Institute of Optonics, System Technologies and Image Exploitation (IOSB) (Gehring et al., 2017). The used point clouds are acquired by two Velodyne HDL-64E mounted at an angle of 35° on the front roof of the vehicle. Fig. 6 provides sketch about how the two scanners are mounted (Gehring et al., 2017). The original raw point clouds are also preprocessed by a statistical outlier removal for down-sampling and noise suppressing. The number of points after preprocessing is around 50 million. In Fig. 7, we illustrate a comparison between the raw and preprocessed point clouds.

With thousands of scans acquired by the laser scanners along the Arcisstrasse (Hackel et al., 2016), a scene containing various kinds of objects is obtained through the combination of point cloud of all the scans. For the evaluation process, we also generate an accurate manually labeled point cloud for the experimental dataset as ground truth.

In our experiment, voxel resolution in the voxelization is set to 0.3 m and seed resolution for supervoxelization is set to 1.5 m. Besides, the number of trees for training RF classifier is set to 200. We use only 50% of the input points cloud for training, while the other half for the evaluation. The performance of our method is further evaluated with true positive (TP), false positive

(FP), true negative (TN), false negative (FN) from the confusion matrix according to

$$\text{precision} = \frac{|TP|}{|TP| + |FP|} \quad (4)$$

$$\text{recall} = \frac{|TP|}{|TP| + |FN|}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

To further investigate the effectiveness of the proposed detrended geometric features, we also compared our method with the other two reference methods. The first one is a variant of the method reported in (Aijazi et al., 2013), following their strategy we use geometric features listed in Table 1 instead of the original RGB colors, surface normals, shape and geometric centers used in their work. The other one is the method of encoding features proposed by (Yu et al., 2016) which designs a feature region with local adjacent graph for supervoxels and forms a grouped patch as the basic unit. Instead of using time consuming FPFH descriptor (Rusu et al., 2009) for estimate feature vectors, we adopt our geometric features listed in Table 1. From the classifiers used for supervised classification, only the RF classifier is used aiming at eliminating the influence of using different classifiers.

4.2 Results and discussion

Using the proposed segment-based classification for assigning eight semantic classes (i.e., man-made terrain, natural terrain, high vegetation, low vegetation, buildings, hard scape, scanning artefacts, and vehicles) we obtain an overall accuracy of 0.92. The statistics of the classification results are given in Table 2.

As seen from the table, the proposed detrended geometric features can outperform other reference methods with regard to the overall accuracy. For the F_1 measures of all kinds of objects, our method reveals better performance. Especially for the objects of low vegetation and hard scape, which are easily to be mixed up with the natural ground surface and building facades, our proposed method achieved much better results with the F_1 measures larger than 0.6.

As stated in Section 2.3, compared with the classical point-based classification methods, one of the major advantages of the segment or primitive based classification method is that they are more insensitive to outliers and noise existing in the dataset, benefiting from the pre-clustering process. However, the use of supervoxel structures also has some drawbacks. For example, the resolution of boundaries of the supervoxels will be directly related to the resolution of the voxels due to the volumetric process, so that all the boundaries obtained between different objects will always appear to be a zigzag shape. Thus, the labels of some details near the edges will be blurred. In other words, the selection of the size of the voxels is a trade-off between the suppression of noise and uneven density and the preservation of details. The larger the voxel, the more details will be smoothed. In Fig. 8, we show a zoom in view of a part of the testing scene. It is clear that for the regular boundaries, like the right-angle sides of the corners formed by the smooth wall surfaces, the supervoxels can find the boundaries accurately. However, when it comes to the irregular edges, for example, the edges between the wall and the ground surface, due to the existence of the French windows, the boundaries found by supervoxels are biased. In the final classification

Class	Precision	Recall	F_1
Man-made terrain	0.861	0.811	0.835
Natural terrain	0.924	0.914	0.919
High vegetation	0.841	0.987	0.908
Low vegetation	0.056	0.100	0.072
Buildings	0.640	0.834	0.725
Hard scape	0.887	0.022	0.044
Scanning artefacts	0.216	0.105	0.141
Vehicles	0.012	0.013	0.012
Overall accuracy	0.825		

Table 3. Evaluation of using Bildstein dataset.

results, these parts also appears zigzag connection influencing final accuracy.

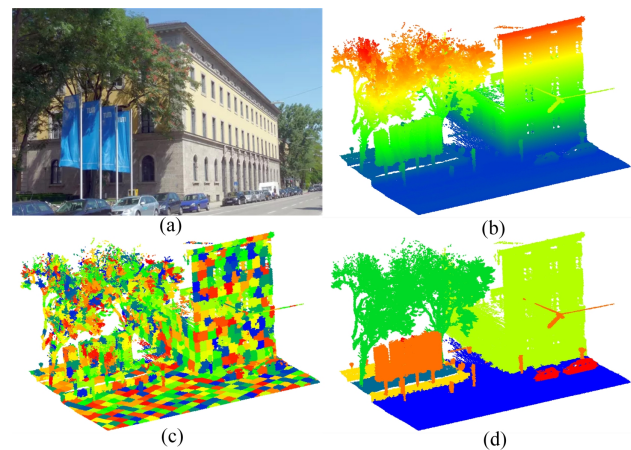


Figure 8. (a) Real scene. (b) MLS point clouds. (c) Supervoxels of the point clouds. (d) Classified sub-scene of test scene.

To further exploit the potential of our proposed feature extraction method, we also test our classification method on the popular Semantic 3D dataset published by ETH Zürich (Hackel et al., 2017) for a preliminary validation. It is notable that the Semantic 3D dataset is a TLS dataset, so that the density of points varies with the distance from the observation station to the objects. In Tables 3 and 4, we show a evaluation result of using this dataset. In this experiment, the scan we used for training is Bildstein 3,5. For evaluation, the scan of Bildstein 1 and Domfontain are used, involving all the eight classes of objects. As seen from the tables, it is apparent that our proposed method can still achieve good results, with the overall accuracies reaching 0.825 and 0.934, respectively. It is noteworthy that in this experiments it is failed to classify the objects of vehicles and low vegetation. One of the possible explanations is the limited number of the training samples of these kinds of objects. This is also a common problem for all the segment based classification methods. For the objects with sufficient training samples, for example, buildings, man-made terrain, and natural terrain (see Fig. 9). In Fig. 10, we can see that the testing scene of Domfontain consists of mainly buildings, man-made terrain, and high vegetations. As seen from the figure, it is clear that for our major concerns (i.e., buildings, roads, and tree), the results of the proposed method still reveals promising potential of our feature extraction method.

5. CONCLUSION

In this work, we presented a classification method based on detrended geometric features capturing both the "high and low fre-

	Method 1			Method 2			Our method		
Class	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
Man-made terrain	0.951	0.932	0.941	0.917	0.936	0.926	0.952	0.943	0.947
Natural terrain	0.781	0.905	0.838	0.799	0.883	0.839	0.839	0.911	0.873
High vegetation	0.858	0.934	0.895	0.895	0.960	0.927	0.931	0.972	0.951
Low vegetation	0.697	0.333	0.450	0.692	0.267	0.385	0.787	0.496	0.609
Buildings	0.878	0.842	0.860	0.915	0.887	0.901	0.933	0.931	0.932
Hard scape	0.707	0.391	0.503	0.864	0.365	0.513	0.844	0.495	0.624
Scanning artefacts	0.489	0.190	0.273	0.663	0.303	0.416	0.747	0.345	0.472
Vehicles	0.798	0.766	0.782	0.840	0.801	0.820	0.827	0.848	0.837
Overall accuracy	0.868			0.893			0.921		

Table 2. Evaluation of using TUM dataset.

Class	Precision	Recall	F_1
Man-made terrain	0.990	0.992	0.991
Natural terrain	—	—	—
High vegetation	0.857	0.515	0.644
Low vegetation	0.150	0.091	0.114
Buildings	0.910	0.982	0.944
Hard scape	0.838	0.348	0.491
Scanning artefacts	0.310	0.490	0.380
Vehicles	0.379	0.020	0.038
Overall accuracy	0.934		

Table 4. Evaluation of using Domfountain dataset.

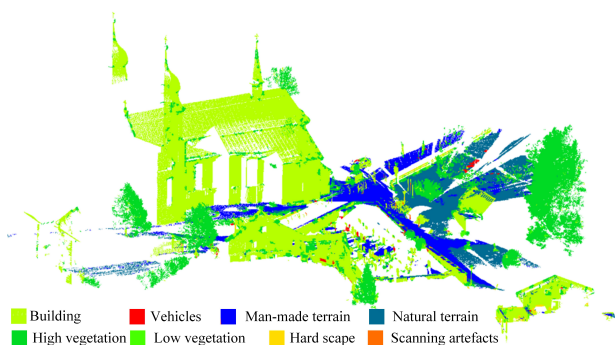


Figure 9. Classification results of the scene Bildstein.

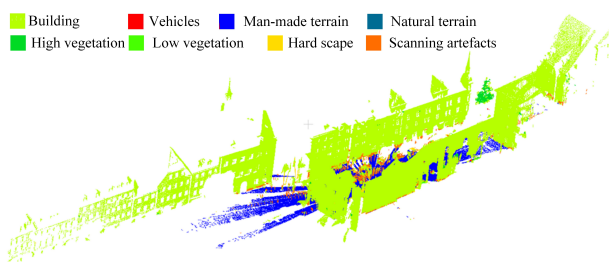


Figure 10. Classification results of the scene Domfountain.

quency” components of the highly complicated objects in an urban scene. To be specific, a novel geometric feature extraction method, detrending the redundant and in-salient information in the local context, is proposed, which is proved to be effective and efficient for extracting local geometric features from the 3D scene. Experiments using complex urban scenes with manually labeled ground truth are conducted, and the performance of the proposed method tuning with different features is tested and analyzed. In our testing dataset, a result of 0.92 for the overall accuracy is achieved for assigning eight semantic classes. In our

future work, different from using current hand-crafted features, we will investigate the possibility to include automatic geometric feature extraction methods. Many related works using auto-encoder (Elbaz et al., 2017) as well as neural networks have also been reported (Landrieu and Simonovsky, 2017). Besides, using complete segments instead of over-segmented supervoxels is also a promising solution for dealing with large-scale dataset.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Marcus Hebel and Joachim Gehring for their valuable help and support to this work. This work was carried out within the frame of Leonhard Obermeyer Center (LOC) at Technische Universität München (TUM) [www.loc.tum.de]

REFERENCES

- Aijazi, A. K., Checchin, P. and Trassoudaine, L., 2013. Segmentation based classification of 3d urban point clouds: A super-voxel based approach with evaluation. *Remote Sensing* 5(4), pp. 1624–1650.
- Al-Manasir, K. and Fraser, C. S., 2006. Registration of terrestrial laser scanner data using imagery. *The Photogrammetric Record* 21(115), pp. 255–268.
- Breiman, L., 2001. Random forests. *Machine learning* 45(1), pp. 5–32.
- Buch, N., Velastin, S. A. and Orwell, J., 2011. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems* 12(3), pp. 920–939.
- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(Part 3), pp. W8.
- Elbaz, G., Avraham, T. and Fischer, A., 2017. 3d point cloud registration for localization using a deep neural network auto-encoder. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 2472–2481.
- Filin, S. and Pfeifer, N., 2006. Segmentation of airborne laser scanning data using a slope adaptive neighborhood. *ISPRS journal of Photogrammetry and Remote Sensing* 60(2), pp. 71–80.
- Frome, A., Huber, D., Kolluri, R., Bülow, T. and Malik, J., 2004. Recognizing objects in range data using regional point descriptors. *Computer vision-ECCV 2004* pp. 224–237.

- Gehring, J., Hebel, M., Arens, M. and Stilla, U., 2017. An approach to extract moving objects from mls data using a volumetric background representation. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Guinard, S. and Landrieu, L., 2017. Weakly supervised segmentation-aided classification of urban scenes from 3d lidar point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Guo, Y., Bennamoun, M., Soheli, F., Lu, M. and Wan, J., 2014. 3d object recognition in cluttered scenes with local surface features: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(11), pp. 2270–2287.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K. and Pollefeys, M., 2017. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-1-W1, pp. 91–98.
- Hackel, T., Wegner, J. D. and Schindler, K., 2016. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Jutzi, B. and Gross, H., 2009. Nearest neighbour classification on laser point clouds to gain object structures from buildings. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(Part 1), pp. 4–7.
- Landrieu, L. and Simonovsky, M., 2017. Large-scale point cloud semantic segmentation with superpoint graphs. *arXiv preprint arXiv:1711.09869*.
- Lee, I. and Schenk, T., 2002. Perceptual organization of 3d surface points. *International archives of photogrammetry, remote sensing and spatial information sciences* 34(3/A), pp. 193–198.
- Lefsky, M. A., Cohen, W., Acker, S., Parker, G. G., Spies, T. and Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests. *Remote sensing of environment* 70(3), pp. 339–361.
- Lim, E. H. and Suter, D., 2009. 3d terrestrial lidar classifications with super-voxels and multi-scale conditional random fields. *Computer-Aided Design* 41(10), pp. 701–710.
- Linsen, L. and Prautzsch, H., 2001. Local Versus Global Triangulations. In: *Eurographics 2001 - Short Presentations*, Eurographics Association.
- Lodha, S. K., Fitzpatrick, D. M. and Helmbold, D. P., 2007. Aerial lidar data classification using adaboost. In: *3-D Digital Imaging and Modeling, 2007. 3DIM'07. Sixth International Conference on*, IEEE, pp. 435–442.
- Lodha, S. K., Kreps, E. J., Helmbold, D. P. and Fitzpatrick, D. N., 2006. Aerial lidar data classification using support vector machines (svm). In: *3DPVT*, pp. 567–574.
- Maas, H.-G., 1999. The potential of height texture measures for the segmentation of airborne laserscanner data. In: *Fourth international airborne remote sensing conference and exhibition/21st Canadian symposium on remote sensing*, Vol. 1, pp. 154–161.
- Muja, M. and Lowe, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)* 2(331–340), pp. 2.
- Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing* 87, pp. 152–165.
- Papon, J., Abramov, A., Schoeler, M. and Worgotter, F., 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2027–2034.
- Rabbani, T., Van Den Heuvel, F. and Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. *International archives of photogrammetry, remote sensing and spatial information sciences* 36(5), pp. 248–253.
- Rusu, R. B., Blodow, N. and Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. In: *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, IEEE, pp. 3212–3217.
- Salti, S., Tombari, F. and Di Stefano, L., 2014. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding* 125, pp. 251–264.
- Secord, J. and Zakhor, A., 2006. Tree detection in aerial lidar and image data. In: *Image Processing, 2006 IEEE International Conference on*, IEEE, pp. 2317–2320.
- Vetrivel, A., Gerke, M., Kerle, N., Nex, F. and Vosselman, G., 2017. Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Vo, A.-V., Truong-Hong, L., Laefer, D. F. and Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp. 88–100.
- Vosselman, G. and Maas, H.-G., 2010. *Airborne and terrestrial laser scanning*. CRC Press.
- Wang, Z., Zhang, L., Fang, T., Mathiopoulos, P. T., Tong, X., Qu, H., Xiao, Z., Li, F. and Chen, D., 2015. A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification. *IEEE Transactions on Geoscience and Remote Sensing* 53(5), pp. 2409–2425.
- Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3d point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 5, pp. W2.
- Weinmann, M., Jutzi, B., Hinz, S. and Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 286–304.
- Yao, W., Hinz, S. and Stilla, U., 2011. Extraction and motion estimation of vehicles in single-pass airborne lidar data towards urban traffic analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3), pp. 260–271.
- Yu, X., Liang, X., Hyypä, J., Kankare, V., Vastaranta, M. and Holopainen, M., 2013. Stem biomass estimation based on stem reconstruction from terrestrial laser scanning point clouds. *Remote sensing letters* 4(4), pp. 344–353.
- Yu, Y., Li, J., Guan, H., Jia, F. and Wang, C., 2015. Learning hierarchical features for automated extraction of road markings from 3-d mobile lidar point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8(2), pp. 709–726.
- Yu, Y., Li, J., Wen, C., Guan, H., Luo, H. and Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 113, pp. 106–123.