Multi-Scale Local Context Embedding for LiDAR Point Cloud Classification

Rong Huang, *Student Member, IEEE*, Danfeng Hong, *Student Member, IEEE*, Yusheng Xu, *Member, IEEE*, Wei Yao, and Uwe Stilla, *Senior Member, IEEE*

***Abstract*— The semantic interpretation using point clouds, especially regarding light detection and ranging (LiDAR) point cloud classification, has attracted a growing interest in the fields of photogrammetry, remote sensing, and computer vision. In this letter, we aim at tackling a general and typical feature learning problem in 3-D point cloud classification—*how to represent geometric features by structurally considering a point and its surroundings in a more effective and discriminative fashion?* Recently, enormous efforts have been made to design the geometric features, yet it is less investigated to fully explore the potentials of the features. For that, there have been many filter-based studies proposed by selecting a subset of the whole feature space for better representing the local geometry structure. However, such a hard-threshold selection strategy inevitably suffers from information loss. In addition, the construction of the geometric features is relatively sensitive to the size of the neighborhood. To this end, we propose to extract multi-scaled feature representations and locally embed them into a low-dimensional and robust subspace where a more compact representation with the intrinsic structure preservation of the data is expected to be obtained, thereby further yielding a better classification performance. In our case, we apply a popular manifold learning approach, that is, locality-preserving projections, for the task of learning low-dimensional embedding. Experimental results conducted on one LiDAR point cloud data set provided by the 2018 IEEE Data Fusion Contest demonstrate the effectiveness of the proposed method in comparison with several commonly used state-of-the-art baselines.**

***Index Terms*— Geometric features, light detection and ranging (LiDAR) point cloud classification, local manifold learning (LML), multi-scale.**

I. Introduction

LAND use and land cover is of great significance for urban understanding and monitoring. To acquire sufficient information for this task, using an appropriate data set is necessary. Many recent studies have focused on the use of earth observation data sources, such as hyperspectral images and multispectral images, for classifying different materials by means of spectral absorption characteristics of various objects [1], [2]. Intuitively, the characteristics of spatial structures cannot be well encoded when using such image-based data sources for the classification task, such as a digital surface model (DSM) [3], [4]. Recently, light detection and ranging (LiDAR), acquiring rich 3-D spatial information efficiently and cost-effectively, has become an important data source for the classification of urban objects [5]–[7]. However, in the complex urban scenes, the automatic classification of 3-D points is still a challenging task due to an insufficient and inaccurate representation of the local geometry.

For classifying point clouds, the generation of distinctive features is of great importance for obtaining high-precision results. The straightforward strategy to increase the distinctiveness of feature is to introduce more candidate features [8] to get a better representative of the domain. However, superabundant information would be counterproductive in improving the performance of generated features when representing the local geometry due to possible high coherence of features. Moreover, the estimation of high-dimensional features is computationally expensive and time-consuming. To tackle this problem, the selection of features is a feasible strategy for reducing repetitive and redundant feature information [5], [9], [10]. However, it is inevitable to lose some useful information during the selection of features. An alternative strategy is used to reduce the dimensionality of features. Compared with the feature selection strategy, it automatically reduces the size of the features and simultaneously encapsulates repetitive and redundant information into distinctive domains of the feature space. Moreover, the number of neighbors of points is also a critical factor to be considered for the calculation of geometric features. Weinmann *et al.* [9], [10] proposed a promising strategy in their comprehensive review to guide the selection of neighborhood size by means of eigenentropy measurement. As a tradeoff between the computational cost and optimal size of the neighborhood, we propose to extract contextual features in a multi-scale way instead of directly estimating the size of the neighborhood. It should be noted, however, that multi-scaled features inevitably suffer from redundancy in information diversity.

The pointwise classification results may be heterogeneous due to the deficiency in modeling local neighborhood relations of a given point. A solution was proposed in [11] by applying contextual classification for smoothing, while

Manuscript received January 12, 2019; revised April 7, 2019 and June 28, 2019; accepted July 7, 2019. This work was supported by the China Scholarship Council. *(Corresponding author: Danfeng Hong.)*

R. Huang, Y. Xu, and U. Stilla are with Photogrammetry and Remote Sensing, Technische Universität München, 80333 München, Germany (e-mail: rong.huang@tum.de; yusheng.xu@tum.de; stilla@tum.de).

D. Hong is with the Remote Sensing Technology Institute (IMF), German Aerospace Center, 82234 Wesseling, Germany, and also with the Signal Processing in Earth Observation (SiPEO), Technische Universität München (TUM), 80333 München, Germany (e-mail: danfeng.hong@dlr.de).

W. Yao is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: wei.hn.yao@polyu.edu.hk).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2927779

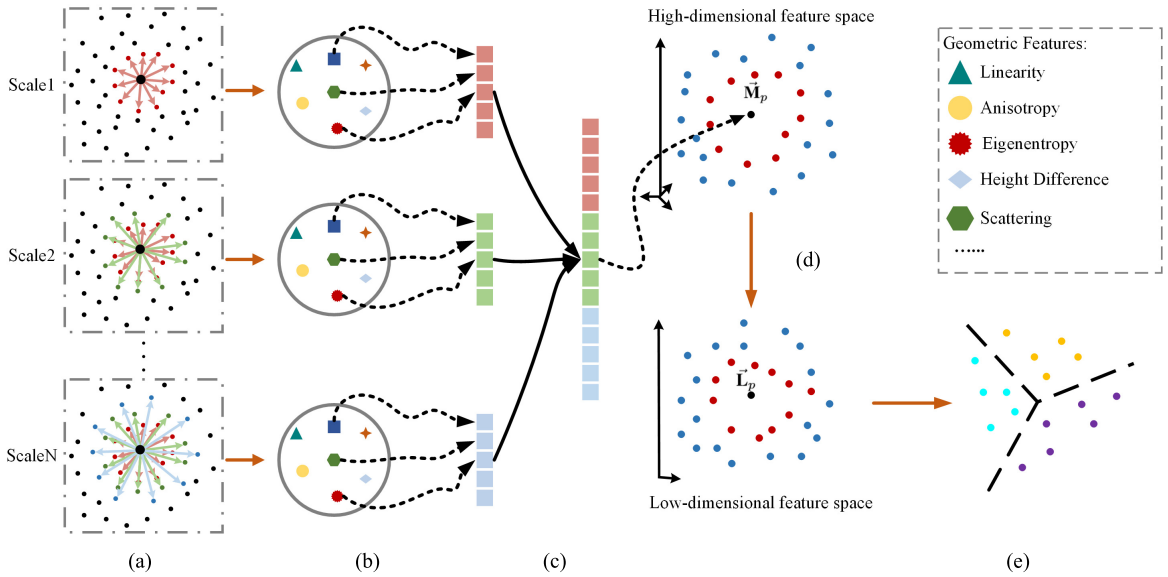

Fig. 1. Workflow of the proposed point cloud classification strategy. (a) Scaling. (b) Geometric space. (c) Concatenation. (d) Embedding. (e) Classification.

Landrieu *et al.* [12] globally optimized the classification results by considering contextual information as a post-processing step.

To this end, we present a workflow for land use and land cover classification using LiDAR point clouds, combining multi-scale feature extraction with manifold learning (ML)-based dimensionality reduction (DR). Inspired by promising results achieved in the context of hyperspectral DR [13], [14], we also make use of ML for the DR of features. The main reason to select the ML methods behind the feature extraction step lies in the consistent motivation, that is, they both model the locally contextual information. Owing to the massive data amount of the point cloud, we use a linear ML method, locality-preserving projections (LPP). To be specific, we developed a novel feature reduction method for point cloud feature representation, using ML strategies. We evaluate our classification method by performing a qualitative and quantitative analysis with two comparable and representative approaches. We also conduct experiments using LiDAR point clouds provided in the 2018 GRSS data fusion contest (DFC) [15], to test and analyze the performance of our feature extraction method in a highly complicated urban scene.

TABLE I
LIST OF USED FEATURES

Category	Features	Definition
Surface	Local density	$D_k = \frac{3n}{4\pi {r_{seed}}^3}$
Statistics [16]	Omnivariance	$O_k = \sqrt[3]{e_1 \cdot e_2 \cdot e_3}$
	Anisotropy	$A_k = (e_1 - e_3)/e_1$
	Eigenentropy	$E_k = -\sum_{i=1}^{3} e_i \ln(e_i)$
	Local curvature	$C_k = \frac{e_3}{e_1+e_2+e_3}$
	Summation	$\sum_k = e_1 + e_2 + e_3$
	Center	$\hat{\mathbf{X}}_k = \frac{1}{k+1}\sum_{i=0}^{k} \mathbf{X}_i$
Dimensionality [9]	Linearity	$L_k = \frac{e_1-e_2}{e_1}$
	Planarity	$P_k = \frac{e_2-e_3}{e_1}$
	Scattering	$S_k = \frac{e_3}{e_1}$
Height [17]	Height mean	$\bar{H}_k = \frac{1}{n}\sum_{i=1}^{n} Z_i$
	Height difference	$\Delta H_k = Z_{max} - Z_{min}$
Orientation [18]	Normal vectors	N_x
		N_y
		N_z
	Verticality	$1 - N_z$

II. METHODOLOGY

The proposed method for the feature embedding of LiDAR point cloud includes two major stages: 1) the extraction of point-based multi-scale geometric features and 2) the reduction of the dimensionality of features. The workflow of the proposed method is shown in Fig. 1.

A. Extraction of Point-Based Multi-Scale Geometric Features

Feature extraction is a crucial part in point cloud classification, and its performance plays an important role in the quality of the classification results. However, it remains a challenging task to extract sufficient information from raw LiDAR points.

In this letter, we construct a set of geometric features including surface features, statistical features, dimensionality features, height features, and orientation features. Specifically, they consist of local density D_k, omnivariance O_k, anisotropy A_k, eigenentropy E_k, local curvature C_k, sum of eigenvalues $\sum_k$, geometric center $\hat{\mathbf{X}}_k$, linearity L_k, planarity P_k, scattering S_k, height mean $\bar{H}_k$, height difference ΔH_k, normal vector $\mathbf{n}_k$, and verticality V_k (see Table I).

These features are used to represent the local geometric shape based on the eigenvalue decomposition of the 3-D structure tensor formed by a specific neighborhood of the point p [9].

The geometric features are also influenced by the neighborhood scale. Thus, to consider different local contexts, the geometric features of the query point are calculated using three different neighborhood sizes k. Here, $k = k_0$, k_1, and k_2, and in this letter, $k_0 = 10$, $k_1 = 20$, and $k_2 = 30$ (determined based on the point density). Then, the geometric features are

concatenated to form the multi-scale feature vector to represent the 3-D LiDAR point cloud scene.

B. Low-Dimensional Embedding of Multi-Scale Features

1) Locality-Preserving Projections: In this section, we briefly review a widely used local ML (LML) method, namely LPP. LPP embeds the high-dimensional data or feature into a low-dimensional subspace in which the topological structure of high-dimensional data is locally preserved. Given a set of data samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{L\times N}$ with L dimension by N pixels, LPP linearly learns a mapping $\mathbf{A}$ to find the corresponding low-dimensional embedding $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{D\times N}$ ($D \ll L$). This process can be modeled in the following:

$$\sum (\mathbf{y}_i - \mathbf{y}_j)^2 \mathbf{W}_{ij} \tag{1}$$

where $\mathbf{W}_{ij}$, which is an adjunct matrix, can be defined as

$$\mathbf{W}_{ij} = \begin{cases} \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/\sigma^2) \\ 0, \end{cases} \tag{2}$$

where σ denotes the standard deviation of the Gaussian kernel function.

To enhance the model's interpretability and transferability, (4) can be approximately modeled in a linearized way. Suppose $\mathbf{y}^T = \mathbf{a}^T\mathbf{X}$, where $\mathbf{a}$ is a linear projection; thus, (4) can be simplified as follows:

$$\begin{aligned} \sum (\mathbf{y}_i - \mathbf{y}_j)^2 \mathbf{W}_{ij} &= \sum (\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j)^2 \mathbf{W}_{ij} \\ &= \sum \mathbf{a}^T\mathbf{x}_i\mathbf{D}_{ii}\mathbf{x}_i^T\mathbf{a} - \sum \mathbf{a}^T\mathbf{x}_i\mathbf{W}_{ij}\mathbf{x}_j^T\mathbf{a} \\ &= \mathbf{a}^T\mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^T\mathbf{a} = \mathbf{a}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a} \end{aligned} \tag{3}$$

where $\mathbf{D}$ is a diagonal matrix and $D_{ij} = \sum_j \mathbf{W}_{ij}$. $\mathbf{L}$ is the Laplacian matrix computed by $\mathbf{L} = \mathbf{D} - \mathbf{W}$. To avoid the trival solution, a necessary constraint formulated by $\mathbf{y}^T\mathbf{D}\mathbf{y} = 1$ is forced in the process of solving (6), and its linearized version can be written as

$$\mathbf{a}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a} = 1. \tag{4}$$

Accordingly, the variable $\mathbf{a}$ can be estimated by minimizing the following objective function:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} (\mathbf{a}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}). \tag{5}$$

The solution of (8) can be equivalently obtained by solving the following generalized eigenvalues decomposition problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a}. \tag{6}$$

2) Geometric Primitive Embedding: Multi-scaled feature modeling can provide the geometric information more sufficiently by considering the local neighboring context of a given point. To some extent, the redundancy of the features hinders the classification performance from further going better. For this reason, feature selection has been used to, to some extent, handle this issue by filtering some features based on the inter-correlations between features or the coherence between features and given labels. It should be noted that, however, the new feature space constructed by those selected features still lies in the original space. Beyond the original feature space, we propose to use the LML-based method (LPP in our case) on the extracted geometric primitive to learn a compact low-dimensional feature representation by locally embedding the neighboring information.

Compared with other DR methods that maximize or preserve the specific information, such as principal component analysis (PCA) and linear discriminant analysis (LDA), LML-based method is capable of mining the underlying data structure and considering the correlation between points. More specifically, for each point in the multi-scale feature domain, there is a strong spatial correlation between points, particularly their neighborhoods. Intuitively, LPP learns the low-dimensional representation by constructing a neighboring graph for each sample in the high-dimensional feature domain. This might further enhance the connections between the original features of the points and their neighbors.

Compared with other nonlinear LML's approaches, LPP can explicitly project out-of-samples into the learned subspace, owing to its linearized technique. Moreover, a large number of samples are usually used for the model's learning in practice. For the two purposes, the multi-scale geometric features are fed into the LPP, leading to the geometric primitive embedding as the input of the final classifier.

III. Experiments

A. LiDAR Point Cloud Data Set

The LiDAR point cloud provided by the 2018 IEEE GRSS DFC data set is a classification-related benchmark data set acquired by the National Center for Airborne Laser Mapping (NCALM) using an Optech Titan MW (14SEN/CON340) with an integrated camera (an LiDAR sensor operating at three different laser wavelengths, namely 1550, 1064, and 532 nm), including 20 land use and land cover categories. Here, experiments are carried out using LiDAR point clouds collected by laser wavelength 1550 nm.

B. Data Preprocessing

To better represent data and reduce the abundant points caused by spatial correlation, we apply a downsampling procedure to reduce the number of points for efficiently testing our proposed method. The downsampling is conducted by selecting the point whose ground coordinates are nearest to the center of a pixel so that we can assign the features of this point to the pixel. Thus, we can construct a feature map whose resolution is the same as the resolution of the ground truth map. The feature for each pixel corresponds to the multi-scale features of the point assigned to this pixel.

C. Results of 2018 GRSS DFC LiDAR Data

For this data set, we adopted a random sampling strategy to select training and test samples. We randomly select 500 samples for each class as training samples and the rest of the samples with labels are assigned as test samples. Labels are selected as training, and test samples can be seen in Fig. 2. We compare the classification results on multi-scale features with that of the single-scale neighborhood, and the classification results on dimensionality-reduced data using LPP with those using some benchmark DR methods (PCA and LDA) and original geometric features (OGF). Since our main focus

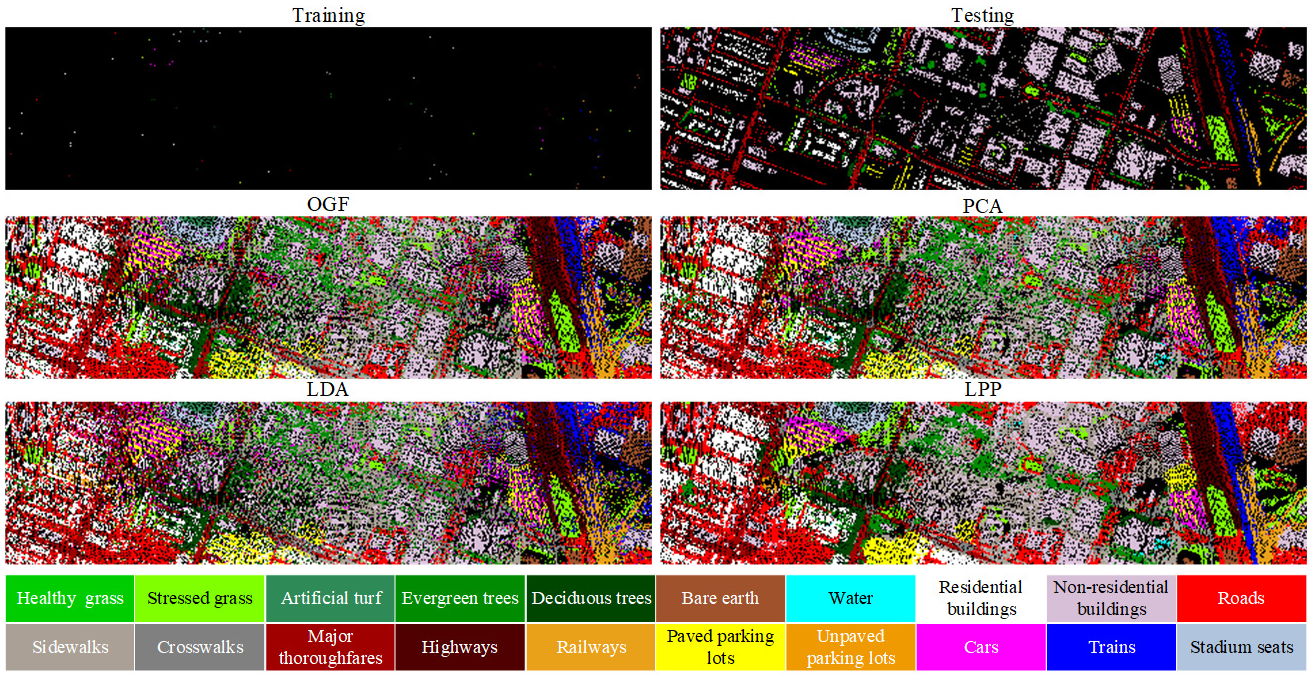

Fig. 2. Classification maps of the different algorithms obtained using an RF classifier on the DFC2018 data set (the map of training samples is dilated for better illustration, and the black points in classification maps are pixels without projected LiDAR points).

TABLE II

CLASSIFICATION ACCURACY FOR DIFFERENT FEATURE EXTRACTION METHODS USING DFC2018 DATA SET

Method	Neighborhood size	Classification Accuracy	
		OA(%)	AA(%)
Single-scale	k_1=10	74.98 ±0.42	60.70 ±0.65
Dual-scale	k_1=10, k_2=20	78.54±0.39	66.32±0.90
Multi-scale	k_1=10, k_2=20, k_3=30	79.50±0.34	67.84±0.78

TABLE III

CLASSIFICATION ACCURACIES USING OPTIMAL PARAMETERS WITH RF CLASSIFIER FOR DIFFERENT DR METHODS ON THE DFC2018 DATA SET

Method	Optimal Parameters	Classification Accuracy		Running Time (s)
		OA(%)	AA(%)	
OGF	/	79.50±0.34	67.84±0.78	/
PCA	d = 10	80.32±0.28	67.62±0.35	0.09
LDA	d = 19	70.71±0.57	56.36±0.69	0.15
LPP	d = 4, k = 15	86.66±0.42	77.62±0.80	1.78

is to assess the discriminative performance of the learned geometric features, we, therefore, use a classic classifier—random forest whose number of trees is selected to be 100 by cross-validation. Moreover, ten replications were performed for selecting training and test samples.

1) Performance Comparison and Analysis Between Multi-Scale Feature Extraction and Single-Scale Feature Extraction: The representation of the neighborhood of points is a crucial issue directly influencing the classification correctness. Thus, the neighborhood for each point is of great significance. Table II shows the classification accuracy obtained by using different neighborhood scales for feature extraction.

It should be noted that two measures for the classification accuracy are used for performance evaluation, namely overall accuracy (OA) and average accuracy (AA). The multi-scale feature extraction outperforms the other two feature extraction methods. Compared with the single scale and dual scale, multi-scale feature extraction increases the OAs by 4.52% and 0.96%, respectively. For AAs, on the other hand, the corresponding increases are 7.14% and 1.52%, respectively. Overall, with the increase of scales of neighborhoods, the performance of classification also increases. It means that multi-scale features provide a more distinctive representation of local context.

2) Performance Comparison and Analysis Between LPP and Classical DR Methods: Table III lists the OAs and AAs of four different methods with optimal parameters determined by cross-validation on the training set.

The LML method outperforms other methods. Compared with OGF, PCA, and LDA, LPP increases the OA by 7.16%, 6.34%, and 15.95%, respectively. For the AA, on the other hand, the corresponding increases are 9.78%, 10%, and 21.26%, respectively. The classification maps are shown in Fig. 2. These results demonstrate the effectiveness of this LML method and imply that it successfully contributes to extracting robust and discriminative low-dimensional features.

Moreover, we also provided the classification results with our method on the test scene given in the DFC2018, yielding the 40% OA. To the best of our knowledge, the resulting accuracy is reasonable to a large extent, since our task is point cloud classification without considering intensities.

3) Sensitivity Analysis of Parameters: The sensitivity of the parameters is examined by varying the number of neighbors, the size of reduced dimensionality d, and the variance of Gaussian kernel in weight determination σ for LPP.

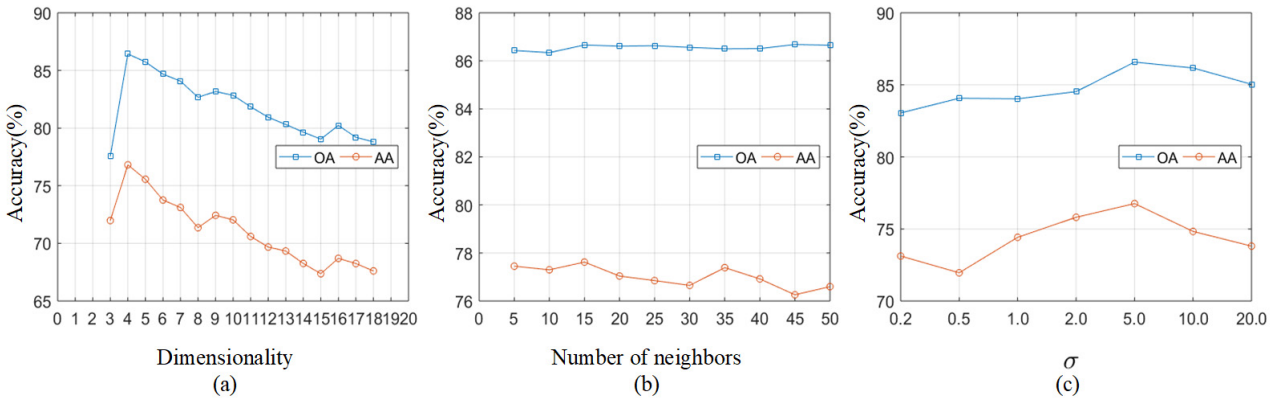

Fig. 3. Sensitivity analysis of the proposed method on three parameters. (a) Dimensionality. (b) Number of neighbors. (c) Variance of Gaussian kernel.

As shown in Fig. 3, the performance of LPP is in some kind sensitive to the parameters. In general, as observed from the data dimensionality point of view, the classification accuracy increases with decreasing dimensionality. When the reduced dimensionality reaches approximately four, the accuracy reaches the nearly optimal level. Compared with reduced dimensionality, LPP is much less sensitive to the other two parameters, the number of neighbors and the variance of the Gaussian kernel. As the number of neighbors gradually increases, the corresponding classification accuracy increases moderately to a peak (e.g., k is equal to around 15) and then fluctuates. A large number of neighbors may obscure the local structure, whereas a small amount of neighbor may not be sufficient to represent the local structure, causing the fluctuation of the LPP performance. The accuracy reaches a peak when the variance of the Gaussian function in weight determination is 5.0. Thus, we can determine the optimal parameters for LPP in our application for reduced dimensionality, number of neighbors, and the variance of Gaussian function to be 4, 15, and 0.5.

IV. Conclusion

In this letter, for LiDAR point cloud classification, we proposed a novel workflow that combines multi-scale geometric feature extraction with an LML method for the DR of features to provide a better representation for point cloud and further classification. The DFC18 LiDAR data set is used in our experiments. The qualitative and quantitative results reveal that our method can outperform other feature DR methods in classification and can provide an effective and distinctive geometric feature representation for our application.

Acknowledgment

The authors would like to thank the National Center for Airborne Laser Mapping (NCALM) and the Hyperspectral Image Analysis Laboratory at University of Houston (UH) and the IEEE GRSS for providing the LiDAR data set.

References

[1] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[2] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," 2019, *arXiv:1902.10858*. [Online]. Available: https://arxiv.org/abs/1902.10858

[3] P. Ghamisi and B. Höfle, "LiDAR data classification using extinction profiles and a composite kernel support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 659–663, May 2017.

[4] X. He, A. Wang, P. Ghamisi, G. Li, and Y. Chen, "LiDAR data classification using spatial transformation and CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 125–129, Jan. 2019.

[5] W. Dong, J. Lan, S. Liang, W. Yao, and Z. Zhan, "Selection of LiDAR geometric features with adaptive neighborhood size for urban land cover classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 60, pp. 99–110, Aug. 2017.

[6] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," *Remote Sens. Environ.*, vol. 158, pp. 295–310, Mar. 2015.

[7] Y. Xu, S. Tuttas, L. Hoegner, and U. Stilla, "Voxel-based segmentation of 3d point clouds from construction sites using a probabilistic connectivity model," *Pattern Recognit. Lett.*, vol. 102, pp. 67–74, Jan. 2018.

[8] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, "A novel hierarchical approach for multispectral palmprint recognition," *Neurocomputing*, vol. 151, pp. 511–521, Mar. 2015.

[9] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, Jul. 2015.

[10] M. Weinmann, B. Jutzi, and C. Mallet, "Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 181–188, Sep. 2014.

[11] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of lidar data and building object detection in urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 152–165, Jan. 2014.

[12] L. Landrieu, H. Raguet, B. Vallet, C. Mallet, and M. Weinmann, "A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 102–118, Oct. 2017.

[13] D. Hong, N. Yokoya, and X. X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, Jun. 2017.

[14] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[15] B. Le Saux, N. Yokoya, R. Hänsch, and S. Prasad, "Advanced multisource optical remote sensing for urban land use and land cover classification [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 85–89, Dec. 2018.

[16] N. Chehata, L. Guo, and C. Mallet, "Airborne lidar feature selection for urban classification using random forests," *ISPRS Arc. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38, p. W8, Sep. 2009.

[17] H.-G. Maas, "The potential of height texture measures for the segmentation of airborne laserscanner data," in *Proc. 4th Int. Airborne Remote Sens. Conf. Exhib./21st Can. Symp. Remote Sens.*, vol. 1, 1999, pp. 154–161.

[18] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 36, no. 5, pp. 248–253, 2006.