EIGENENTROPY BASED CONVOLUTIONAL NEURAL NETWORK BASED ALS POINT CLOUDS CLASSIFICATION METHOD

Xu Zhihai [1], Yang Zhishuang [2]

[1] Airborne patrolling center of Guangdong power grid Co.,Ltd
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing of Wuhan University, Wuhan 430079, China

KEY WORDS: ALS, LiDAR, Convolutional Neural Network, Eigenentropy

ABSTRACT:

The classification of point clouds is the first step in the extraction of various types of geo-information form point clouds. Recently the ISPRS WG II/4 provides a benchmark on 3D semantic labelling, a convolutional neural network based method achieves the best overall accuracy performance in all participants who only use the geometrical and waveform based features extracted from the ALS data. Features of the point are calculated in different scales to achieve the best performance. It is not efficiency for the future use. In this paper, we use an eigenentropy based scale selection strategy to improve this method. The scale selection strategy improves the average F1 score and makes the classification method more simple and efficient.

1. INTRODUCTION

The classification of 3D point clouds has caused great attention in computer vision, remote sensing, and photogrammetry. In recent decades, airborne laser scanning (ALS) has become important in acquiring 3D point clouds. ALS point clouds allow an automated analysis of large areas in terms of assigning a (semantic) class label to each point of the considered 3D point clouds (Blomley et al., 2016; Chehata et al., 2009; Mallet et al., 2011; Niemeyer et al., 2014; Shapovalov et al., 2010). However, the relatively low point density, the irregular point distribution and the complexity of observed scenes make the accuracy of the classification result hard to improve. In our previous work (Yang et al., 2017), we presented a convolutional neural network based method to handle this problem. Although result achieves a good performance, it still remains some question. In order to find the best scale to calculate the features of the point clouds, a lot of repetitive experiments are needed. It is resource intensive and time consuming. A proper way to find the best scale should be used solving this problem.

Many presented approaches have focused on finding an optimal neighbourhood size for each individual 3D point and thus increasing the distinctiveness of derived features. For densely sampled and continuous surfaces, Mitra and Nguyen (Mitra et al., 2004) find the optimal size based on iterative schemes relating to curvature, point density and noise of normal estimation. For more cluttered surface representations, Mark (Mark et al., 2010) uses the surface variation based size selection method, Demantké (Demantké et al., 2011) uses the dimensionality based size selection method and Weinmann (Weinmann et al., 2015) uses the eigenentropy based size selection method.

In this paper, we use the eigentropy based method to select the best scale. The task of finding a suitable neighbourhood may be correspond to minimizing a eigenvalue based measure of unpredictability given by the Shannon entropy(Weinmann et al., 2015). Although the optimal neighbourhood size selection

method may cause additional computational effort, it reduce the repetitive experiments and has a significantly positive impact on 3D scene analysis. In order to verify the algorithm capability, we compare eigenentropy based result to our previous result on the ISPRS WG II/4 benchmark dataset.

The rest of this paper is organized into four Sections. Section 2 introduces our methodology. We present the information of the benchmark dataset and the experiment result in Section 3 and the concluding remarks and suggestions for future work are in Section 4.

2. METHODOLOGY

2.1 Feature Extraction

For each point in the ALS point clouds, we can get four types of features which are intensity, eigenvalue-based features, normal vector based feature and height above DTM (Yang et al., 2017).

2.1.1 Intensity: Except for the X,Y and Z coordinates, the LiDAR system can also collect the intensity of the returns, which is a measure of the amount of energy reflected back to the sensor(García et al., 2010).

In the ISRPS benchmark dataset both impervious surfaces and low vegetation are flat objects, it is hard to distinguish them only by the geometric features. The intensity values are high on the impervious surfaces and low on the low vegetation that can make these object easy to distinguish.

2.1.2 Eigenvalue-Based Features: these features are calculated from the normalized eigenvalues λ_i which are extracted from the point covariance function. The planarity and sphericity are as follows, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$:

$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \tag{1}$$

$$S_\lambda = \frac{\lambda_3}{\lambda_1} \tag{2}$$

2.1.3 Normal Vector Based Feature: These features may help us identify the planar objects such as roofs and roads. The local plane is estimated via a robust M-estimator(Xu and Zhang, 1996) in our experiment. The normal vector can be derived from the local plane and we use the variance of normal vector angle from vertical direction to discriminate planar surface from vegetation.

2.1.4 Height Above DTM: The DTM we use is generated by the commercial software package SCOP++ which uses the robust filtering theory (Kraus and Pfeifer, 1998). Based on the analysis by Niemeryer(Niemeyer et al., 2014), the height above DTM represent the global distribution for a point which helps a lot in the classification work and this feature may be the most important one. It is strongest and best discerning feature for all classes and relations. For instance, to distinguish between a relation of points on a roof or on a road level this feature has a great influence.

2.2 Neighborhood Selection

To obtain an appropriate local neighbourhood as the basis of feature extraction, we focus on the use of an eigenentropy based scale selection method.

The linearity L_λ, planarity P_λ and sphericity S_λ can be derived from the eigenvalues which represent the 1 dimensional, 2 dimensional and 3 dimensional features. The three features sum up to 1, thus conform to the two of three probability axioms. The work of finding a suitable neighbourhood scale can be changed to find the most important dimensionality which corresponds to minimizing a measure of unpredictability given by the Shannon entropy(Shannon, 1948) as:

$$E_{dim} = -L_\lambda ln(L_\lambda) - P_\lambda ln(P_\lambda) - S_\lambda ln(S_\lambda) \tag{3}$$

The eigenvalues can directly be exploited in order to estimate the order/disorder of 3D points within the local 3D neighbourhood (Weinmann et al., 2015). So we normalize the three eigenvalues by the sum $\sum_\lambda$ and the normalized

eigenvalues can be represented as $e_i = \lambda_i / \sum_\lambda$. The eigenentropy is defined from Shannon entropy as

$$E_\lambda = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3) \tag{4}$$

In our experiments, the neighborhood is defined as a sphere in a radii of r. In order to find the best scale, the interval of $r \in [r_{min}, r_{max}]$ has been sampled into 20 different scales. The optimal neighborhood size relate to the radii which minimizes the eigenentropy can then be determined.

2.3 Image Generation & Convolutional Neural Networks

2.3.1 Image Generation: A feature image generation method is used to convert the 3D point clouds to 2D images. The features we use is same as (Yang et al., 2017).

For each point in the ALS point clouds, a square window is set up and divided into 128*128 cells. For each cell centre, we find the nearest point from the point clouds. The best scale and the features are calculated based on this point, and these features are transferred into three integers within 0 to 255 as follow:

$$\begin{aligned} RED &= \left[255 * S_\lambda * \sigma_z^2 \right] \\ GREEN &= \left[Intensity * P_\lambda \right] \\ BLUE &= \left[255 * H_{above} \right] \end{aligned} \tag{5}$$

where P_λ is the planarity, S_λ is the sphericity, σ_z^2 is the variance of normal vector angle from vertical direction and H_{above} is height above DTM. These features are normalized between 0 and 1. *Intensity* denotes the echo intensity values normalized between 0 and 255.

Mapping these three integers to the cell with red, blue and green colors. Then the 128*128 cells square window will be transferred into a 128*128 RGB image.

2.3.2 Convolutional Neural Networks. The architecture of the CNN model we use is shown in Figure 1. The CNN model is implemented with Caffe deep learning netwroks (Jia et al., 2014). As shown in Figure1 the conv denotes the convolutional layers and the pool denotes the pooling layers. From the original input image, hierarchical features can be extracted by these layers. FC denotes the fully connected layers followed by the convolutional and pooling layers to do the classification task.
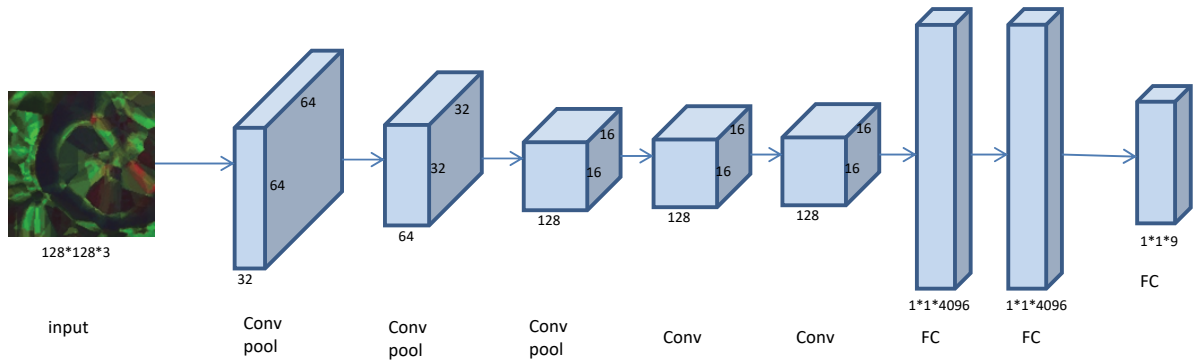
Figure 1. The architecture of the used deep CNN

The input data is usually followed by the convolutional layers. Features can be detected regardless of their position by these convolutional layers. A pointwise nonlinear activation operation $g(\bullet)$ is usually performed subsequently. A pooling layer is followed by the activation operation which can select the dominant features. Pooling layers can improve the robustness of translation and reduce the number of network parameters. The whole process is shown as：

$$H^{k} = pool\left(g\left(H^{k-1} * W_{k} + b_{k} \right) \right) \qquad (6)$$

where H^{k} denotes the kth output data, W_{k} denotes the kth convolutional operation kernels and b_{k} denotes the kth bias values. Rectified linear units are used as our nonlinear activation operation followed as：

$$g(x) = \max(0, x) \qquad (7)$$

3. EXPERIMENTAL RESULT

3.1 Dataset

The ISPRS WG II/4 benchmark test provides the ALS data from Vaihigen used for the 3D reconstruction and 3D labelling challenge. We use this benchmark dataset to evaluate our method. The dataset is acquired with Leica ALS50 system over Vaihigen, a small village in Germany (Cramer, 2010). The mean flying height is 500m and the field of view is 45 degrees. The point density of the dataset is around 8 points/m^2. The points have not been rasterized or post processed. In total 9 classes have been defined and each point in the training dataset is labelled accordingly. The reference labels are provided by the authors of (Niemeyer et al., 2014).

As shown in Figure 2, the training area contains 753876 labeled points with the XYZ coordinates, intensity values, number of returns and reference labels. 9 classes are defined which are Power-line, Low Vegetation, Impervious Surfaces, Car, Fence/Hedge, Roof, Façade, Shrub, Tree. The testing area is shown in Figure 3, which contains 411722 unlabeled points with XYZ coordinated, intensity values and the number of returns.
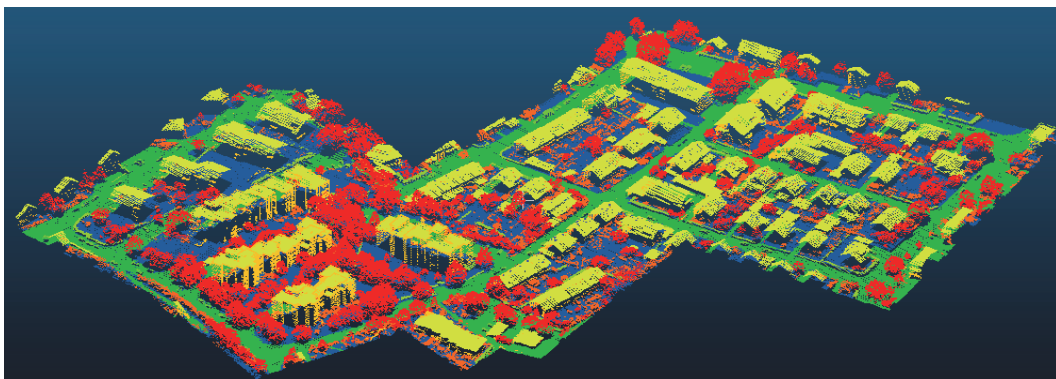

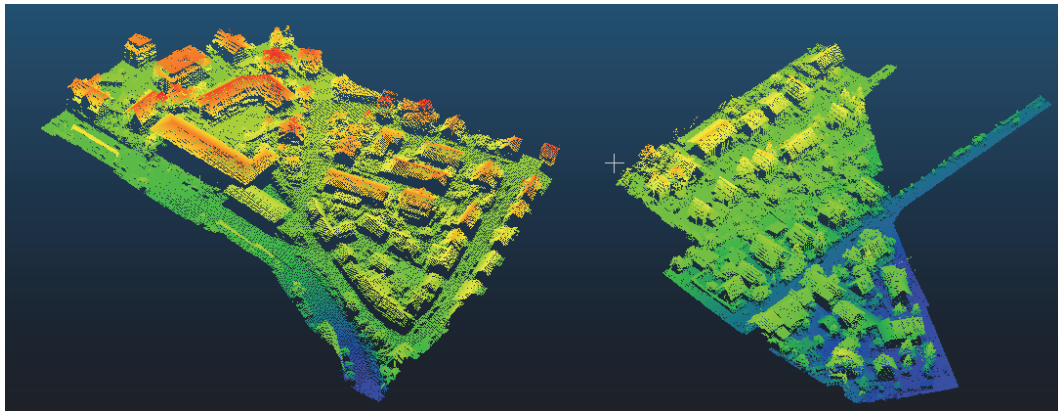Figure 2. Training set: the color encoding shows the assigned semantic class labels

Figure 3.Testing set: The color blue to red represents the z values of the points.

3.2 Accuracy Evaluation

For each class, we compute precision and recall value as follows:

$$precision = \frac{tp}{tp + fp} \qquad (8)$$

$$recall = \frac{tp}{tp + fn}$$

where tp denotes the true positive value, fp denotes false positive value and fn denotes false negative values. F1 score for each class is also calculated as:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \qquad (9)$$

The overall accuracy and the average F1 scores are used to evaluate the performance of our experiment result.

3.3 Experiments & Discussion

As for convolutional neural networks parameters, the batch size is 128, the base learning rate is 0.01, the momentum is 0.9 and the weight decay is 0.0005. As for the feature image generation parameters, the width of cell is 0.05m and the neighbourhood size is calculated with eigenentropy scale selection method. The classification result is shown in Figure 4. We compare our experiment result with Yang's (Yang et al., 2017), four different neighborhood result as shown in Table 1. The $R_{opt,\lambda}$ denotes the eigenentropy based scale.

Class	$R_{0.5}$	$R_{0.75}$	$R_{1.0}$	$R_{2.0}$	$R_{opt,\lambda}$
Power	31.3	26.1	24.7	20.3	**43.3**
Low Vegetation	80.2	80.9	**81.8**	81.2	81.7
Impervious Surfaces	87.8	90.2	**91.9**	89.2	91.1
Car	59.3	**71.2**	69.3	64.2	68.3
Fence/Hedge	11.2	12.7	14.7	13.6	**25.9**
Roof	93.1	94.4	95.4	94.9	**95.7**
Facade	34.1	38.1	40.9	41.7	**43.3**
Shrub	36.4	37.7	38.2	38.3	**45.2**
Tree	77.5	77.9	**78.5**	78.1	73.2
Overall Accuracy	79.7	81.2	**82.3**	81.3	82.2
Average F1	60.5	61.3	61.6	60.7	**63.7**

Table 1. The recall value, the overall accuracy of each class and average F1 value for different neighborhood scale in [%]. Bold numbers show the highest values within the different scale.
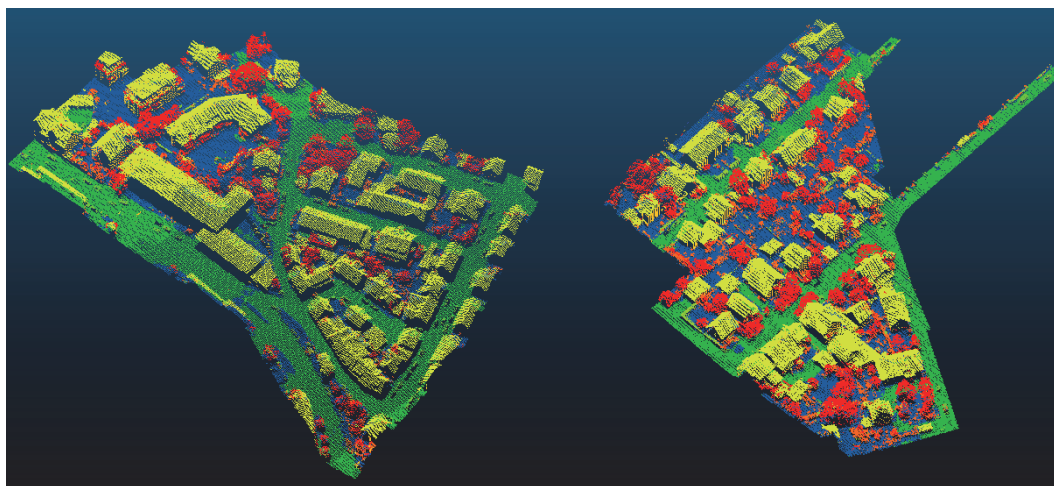
Figure 4. The classification result of the eigenentropy based method, the color encoding
shows the assigned semantic class labels.

For the neighbourhood scale selection, we choose four different radii from 0.5m to 2.0m comparing with the eigenentropy method. In general, the overall accuracy has the highest value when the radius is 1m and the average F1 score has the highest value at the eigenentropy based optimal scale. More specifically, since the point density of each category varies a lot, the highest accuracy for each class may at different neighbourhood scale. The eignentropy based scale selection method solve the problem. Although the accuracy of each class may not have the highest value, the overall accuracy ranks 2nd in all experiment with 0.1% difference to the first place and the average F1 score ranks the 1st in all experiment. It seems that the eigenentropy based method may need a large amount of computation but if we choose this method we don't have to waste time find the optimal neighbourhood size, thus the time costed in the feature generation process will be saved.

4. CONCLUSION

In total, our work reveal that the eigenentropy based scale selection method do have some positive influence on the point based feature image generation thus improve the average F1 score of the CNN based semantic labeling task. In the future work, we will take more features such as echo features, point density into consideration to improve the overall accuracy and try to find a way further improve the algorithm efficiency.

REFERENCES

Blomley, R.D., Jutzi, B., Weinmann, M., 2016. 3D semantic labeling of ALS point clouds by exploiting multi-scale, multi-type neighborhoods for feature extraction. Proceedings of the international Conference on Geographic Object-Based Image Analysis, pp. 1-8.

Chehata, N., Li, G., Mallet, C., 2009. Airborne LIDAR feature selection for urban classification using random forests. Geomatics & Information Science of Wuhan University 38, pp. 207-212.

Cramer, M., 2010. The DGPF-Test on Digital Airborne Camera Evaluation – Overview and Test Design. Photogrammetrie - Fernerkundung - Geoinformation 2010, pp. 73-82.

Demantké, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality Based Scale Selection in 3d LIDAR Point Clouds. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 3812, pp. 97-102.

García, M., Riaño, D., Chuvieco, E., Danson, F.M., 2010. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. Remote Sensing of Environment 114, pp. 816-830.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding, ACM International Conference on Multimedia, pp. 675-678.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. Isprs Journal of Photogrammetry & Remote Sensing 53, pp. 193-203.

Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. Isprs Journal of Photogrammetry & Remote Sensing 66, pp. S71–S84.

Mark, P., Richard, K., Markus, G., 2010. Multi-scale Feature Extraction on Point-Sampled Surfaces. Computer Graphics Forum 22, pp. 281-289.

Mitra, N.J., Nguyen, A.N., Guibas, L., 2004. Estimating surface normals in noisy point cloud data. International Journal of Computational Geometry & Applications 14, pp. 261-276.

Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. Isprs Journal of Photogrammetry & Remote Sensing 87, pp. 152–165.

Shannon, C.E., 1948. IEEE Xplore Abstract - A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 & 623-656.

Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative Markov networks for 3D point cloud classification, Photogrammetric Computer Vision and Image Analysis, pp. 103-108.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. Isprs Journal of Photogrammetry & Remote Sensing 105, pp. 286-304.

Xu, G., Zhang, Z., 1996. Epipolar Geometry in Stereo, Motion and Object Recognition. Computational Imaging & Vision 6, pp. 208-213.

Yang, Z., Jiang, W., Xu, B., Zhu, Q., Jiang, S., Huang, W., 2017. A Convolutional Neural Network-Based 3D Semantic Labeling Method for ALS Point Clouds. Remote Sensing 9, pp. 936-949.