

FRACTAL DIMENSION BASED SUPERVISED LEARNING FOR WOOD AND LEAF CLASSIFICATION FROM TERRESTRIAL LIDAR POINT CLOUDS

Z. Hui¹, Y. Xia^{1,*}, Y. Nie¹, Y. Chang¹, H. Hu¹, N. Li¹, Y. He¹

¹ Faculty of Geomatics, East China University of Technology, Nanchang, China – huizhenyang2008@163.com, 20432xyp@163.com, nieyunju@126.com, changyonglei@ecut.edu.cn, 15179582932@163.com, 18846133671@163.com, 1013518531@qq.com

Commission I, WG I/2

KEY WORDS: Terrestrial LiDAR, Wood, Leaf, Classification, Fractal dimension, Supervised learning

ABSTRACT:

Terrestrial Laser scanner has been widely used in the field of forestry. Wood-leaf separation is the fundamental step to most applications of forestry. This paper presented a robust supervised learning method for wood and leaf classification by developing four new feature vectors. Fractal dimension is first calculated to indicate the difference of regularity or roughness between wood and leaf. Zenith angle and variation are presented to distinguish trunks or branches from leaves. The adaptive axis direction of cylinder is adopted to calculate the local point density precisely. Experimental results show that the supervised learning method using the four feature vectors presented in this paper can achieve a good classification performance. Both accuracy and *F1* score are higher than the ones of the method using eigen value based feature vectors.

1. INTRODUCTION

Three-dimensional laser scanning (LS) technology is an active remote sensing technology which has been developing rapidly in recent years (Hui et al., 2019). LS system can actively transmit laser pulses to obtain the three-dimensional coordinate information of the target object, making it become an important data source for spatial topological analysis of vegetation (Zhang et al., 2019). With the improvement of LS measurement accuracy and sampling rate, this technology has been widely used in forestry, ecology, botany and other related fields (Wang et al., 2019). Compared with airborne or satellite-borne LS, terrestrial LS (TLS) can provide smaller light spots, achieve higher single point measurement accuracy, and also provide denser point clouds. Thus, TLS is widely used in forest structural parameters calculation, above-ground biomass (AGB) estimation, leaf area index (LAI) quantification, etc. (Calders et al., 2015. Calders et al., 2018)

The separation of wood and leaf from TLS data is a crucial step for realizing these above-mentioned applications. The traditional methods usually separate wood and leaf manually with the help of visualization software (Hosoi, Omasa, 2006). Obviously, it will be time consuming and labor intensive. Moreover, the classification results mainly depend on the experience of the staff and the quality of the point clouds obtained. To solve these problems, many automatic algorithms for separating wood and leaf have been developed in recent years. These algorithms can be divided into two categories including geometric information based and echo attributes information based (Jin et al., 2019). The geometric information based approaches implement the separation mainly relying on the different shape characteristics of wood and the leaf (Hosoi et al., 2006. Xu et al., 2007. Ma et al., 2015. Zhu et al., 2018. Vicari et al., 2019). For instance, wood is generally linear

distribution and shows cylinder shape in local areas, while leaf is more scattered and lacks linear characteristics. The echo attributes based methods mainly depend on the difference in information such as the reflection intensity, reflectance, or waveform to achieve the separation (Beland et al., 2011. Beland et al., 2014. Cote et al., 2011). In recent years, with the fast development of LS technology, several laser scanners can acquire full waveform data over vegetation areas. According to the difference in height and width of the waveform data towards wood and leaf, the accuracy of separation can be improved (Hancock et al., 2017. Danson et al., 2018. Li et al., 2018. Li et al., 2018).

Several existing experiments have shown that the accuracy of separation methods using echo attributes information such as reflection intensity is lower than the accuracy based on the geometric information method. This is because the reflection intensity of the laser pulse is related to various factors such as distance, incident angle, and roughness of the surface of the object. Thus, it is difficult to achieve radiation calibration of the reflection intensity. Although the use of full waveform data information can improve the accuracy of wood and leaf separation, not all laser scanners can acquire full waveform data. Compared with these sensor specific methods, the approaches based on geometric information containing three-dimensional coordinates of point clouds are more applicable. Almost all the existing geometric methods used eigen values and vectors to classify wood and leaf. Since eigen values indicate the magnitude of variance present in coordinates, while eigen vectors reflect the direction of the variance. To further develop geometric features used for wood-leaf classification, this paper try to develop new feature vectors calculated from three-dimensional coordinates. Two publicly available datasets are used to test the new developed features. Experimental results

* Corresponding author

show that the proposed method can classify wood and leaf effectively.

2. METHODOLOGY

2.1 Fractal dimension feature vector

In Euclidean geometry, objects are generally seen as regular shape and their corresponding geometric feature can be determined as integer dimensions, such as one, two, three etc. However, in the real-world, there are many complicated and irregular objects (e.g., coastlines or snowflakes). Their geometric morphology cannot be described by integer dimensions. To better describe the complexity and roughness of these objects, fractal theory has become a new branch of modern mathematics. Nowadays, this theory has been widely used in many areas, such as signal analysis and image processing (Yang et al., 2015).

In fractal theory, fractal dimension is an important index to describe the fractal morphology. The values of fractal dimensions will indicate the irregularity and roughness of the complex objects. Fractal dimension can be calculated in different ways. This paper adopts the box counting method since its principle is simple and this method is easy to implement. The box-counting dimension is defined as Equation (1).

$$Dim = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)} \quad (1)$$

where ε = the side length of the cube

$N(\varepsilon)$ = the number of the occupied cubes by point clouds

The fractal dimension Dim can be calculated when ε approaches 0. However, in terms of point clouds, the side length of cube cannot get infinitely close to 0. Moreover, ε is generally discrete. To better describe the box-counting dimension, Equation (1) can be changed to the form as Equation (2).

$$\log(N) = -Dim \times \log(\varepsilon) + b \quad (2)$$

where b = a constant

Obviously, a series of different side lengths will lead to a series of different numbers of occupied cubes. By applying the least square fitting between $\log(\varepsilon)$ and $\log(N)$, the box-counting dimension Dim can be obtained.

In terms of forest point clouds, wood and leaf own different fractal morphology. To separate wood and leaf effectively, the fractal dimension for every point can be calculated according to the following steps:

- i Organize the raw point clouds using the k-dimensional tree.
- ii Traverse every point to acquire its neighbors $Set^r\{p_i\}$ within r radius.

iii Voxelize the neighboring point sets with a series of cube side lengths $\varepsilon = \{0.1, 0.2, \dots, r\}$ as shown in Figure 1 and calculate the corresponding number of occupied cubes $N(\varepsilon)$.

iv Calculate the box-counting dimension Dim according to Equation (2) using the least square fitting and repeat steps ii and iii until all the points are traversed.

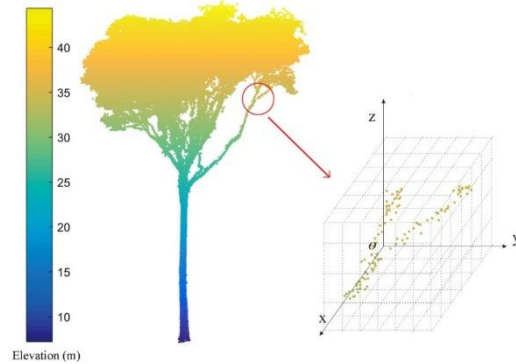


Figure 1. Schematic diagram of fractal dimension calculation

2.2 Zenith angle and variation feature vectors

In general, wood and leaf own different grow features. For instance, trunks tend to grow straight up, while leaves tend to diverge. In other words, zenith angles of trunk points are generally close to 90° , while the zenith angles of other points do not own this characteristics as shown in Figure 2. Moreover, no matter trunks or branches their zenith angle variations are generally smaller than the ones of leaves. Thus, wood and leaf can be separated according to zenith angle and variation feature vectors.

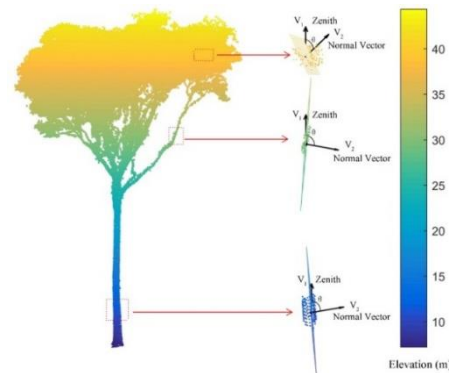


Figure 2. Zenith angles of different part of point clouds

The zenith angle is the angle between the normal vector and the vertical direction in the local region of the points. It can be calculated according to Equation (3).

$$\theta = \arccos \frac{V_1 \cdot V_2}{\|V_1\| \|V_2\|} \quad (3)$$

where V_1 = the zenith direction vector

V_2 = the normal vector

θ = the zenith angle

As mentioned above, the zenith angle variation of wood is general smaller than that of leaf. The zenith angle variation $std(\theta)$ is defined as Equation (4).

$$std(\theta) = \sqrt{\frac{\sum_{i=1}^n (\theta_i - \bar{\theta})^2}{n}} \quad (4)$$

where $\bar{\theta}$ = the mean value of the zenith angle variation
 n = the number of neighboring points

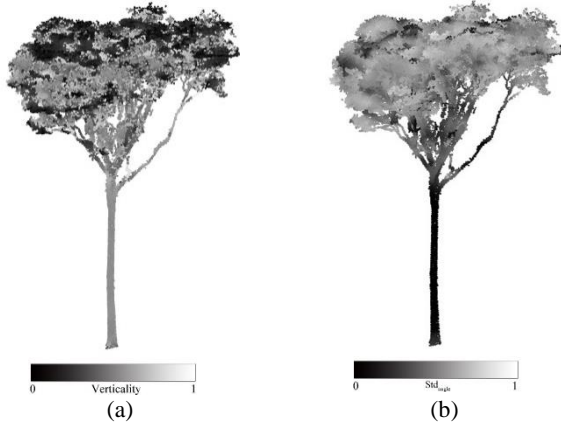


Figure 3. Illustrating the zenith angle and variation for the single tree. (a) zenith angle; (b) zenith angle variation. The color scales at the bottom of each figure indicate the range of angle.

From Figures 3 (a) and (b), it can be found that the zenith angle and variation of wood are clearly different from the ones of leaf. Thus, these two features will contribute to the wood-leaf classification results.

2.3 Points distribution feature vector

Compared with leaf points, wood points are usually cylindrical in distribution. Thus, woods can be separated from leaves by counting the number of points within the cylinder. Note that the axis direction of the cylinder should be adaptive to local point clouds as shown in Figures 4 and 5. The number of points within the cylinder can be seen as local density of point clouds which can be calculated according to Equations (5) and (6).

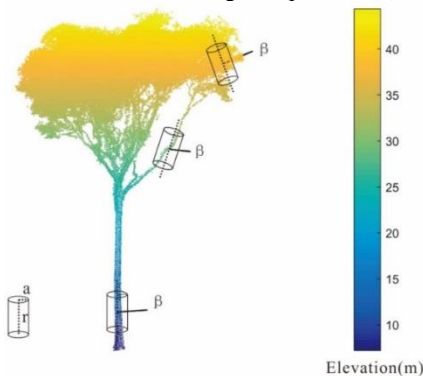


Figure 4. Local point density calculation using the cylinder with adaptive axis direction

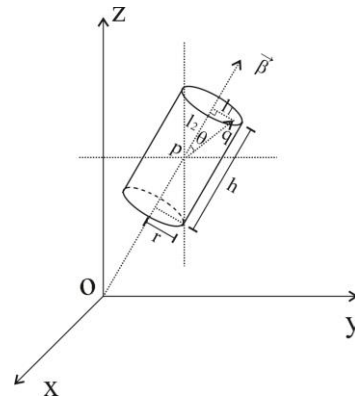


Figure 5. Sketch map of the local point density calculation

$$Density(p) = Num[q_i \in cyl(p, r, h, \beta)] \quad (5)$$

$$cyl(p, r, h, \beta) = \begin{cases} \theta = \arccos\left(\frac{\overline{pq} \cdot \overline{\beta}}{|\overline{pq}| |\overline{\beta}|}\right) \\ l_1 = dis(p, q) \cdot \sin \theta \leq r \\ l_2 = dis(p, q) \cdot \cos \theta \leq \frac{h}{2} \end{cases} \quad (6)$$

where $cyl(\bullet)$ = the cylinder

$Num[\bullet]$ = the number of points within the cylinder

r = the radius of the cylinder

h = the height of the cylinder β = the axis direction of the cylinder

$dis(p, q)$ = the distance between points p and q

$Density(p)$ = the local point density

In this paper, the direction of the cylinder is set as the vertical direction of the normal vector calculated by applying the principal component analysis (PCA) method to the neighboring points. In so doing, the calculated local point density can represent the geometric morphology accurately.

3. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of these four new developed vectors, this paper adopts two isolated tree point clouds for training and testing, respectively. These two datasets are provided by Moorthy et al. (2019). Both of the two datasets are manually labelled using CloudCompare by an experienced person. The two datasets contain x, y, z coordinates and label information, in which 0 represents a leaf point and 1 represents a wood point as shown in Figure 6. Tree 1 (Figure 6 (a)) is used for training since its structure is more complicated, while Tree 2 (Figure 6 (b)) is used for testing. This paper down sampled the two datasets to improve the computational and memory efficiency. Random forest is adopted for the supervised learning since its performance is better than other machine learning methods, such as neural networks, naive bayes, etc. (Moorthy et al., 2019).

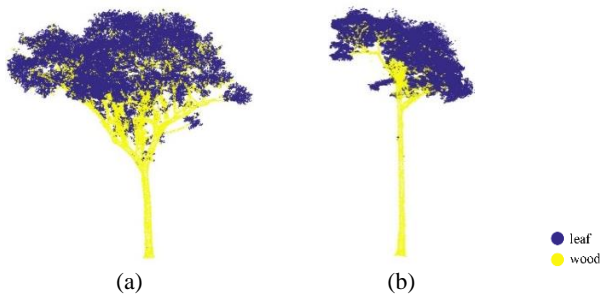


Figure 6. Two isolated tree point clouds with labeling information. Blue point represents leaf, while yellow point represents wood. (a) Tree 1 used for training; (b) Tree 2 used for testing.

Accuracy and F_1 score are used to assess the performance of the proposed method. Accuracy is the proportion of all predictions that are correctly classified. F_1 score is the harmonic mean of both precision and recall. Accuracy and F_1 score can be calculated according to Table 1.

True Positive	False Negative	Accuracy	Precision (P)
A	B	$(A+D)/(A+B+C+D)$	$A/(A+C)$
False Positive	True Negative	Recall (R)	F_1 Score
C	D	$A/(A+B)$	$2*P*R/(P+R)$

Table 1. Confusion matrix used in pointwise assessment of separation results

In the confusion matrix, A and D are correctly classified points, while B and C are wrongly classified points. Obviously, when leaf is defined as the positive class, F_1 score for leaf class can be obtained. When wood is defined as the positive class, F_1 score for wood class can be obtained.

To objectively evaluate the performance of the proposed method, this paper also acquired the wood-leaf classification results using the feature vectors based on eigen values. These eigen value based vectors are planarity, linearity, scatter, surface variation and entropy. The classification results of the two methods are shown in Figures 7 (a) and (b). From the two figures, it can be found that the proposed method owns a better classification result than that of the method using eigen value based feature vectors. What's more, the proposed method can detect the trunk more accurate.

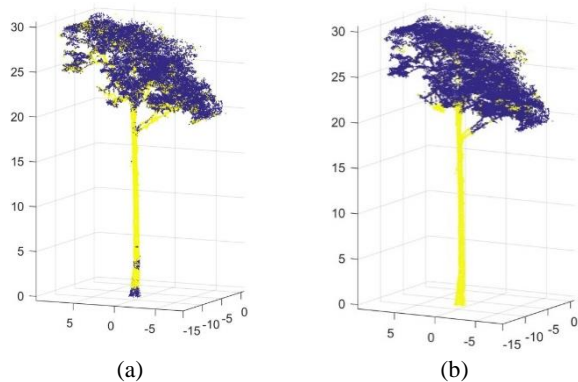


Figure 7. wood-leaf classification results using different vectors. (a) Classification result using the vectors based on eigen values; (b) Classification result using the proposed method.

Table 2 shows the accuracy and F_1 score calculated using the two methods. It is obvious that all the three indexes of the proposed method are higher than the ones of the method using eigen value based vectors. Thus, it can be concluded that the new developed vectors in this paper can separate wood and leaf effectively. From Table 2, we can also find that F_1 score for wood of the two methods is much lower than F_1 score for leaf. It is because that many small branches are wrongly classified as leaves as shown in Figure 7. The classification results can be better when more training datasets are involved.

	Accuracy	F1 Score (wood)	F1 Score (leaf)
Feature vectors based on eigen values	0.829	0.563	0.894
Feature vectors based on the proposed method	0.877	0.588	0.928

Table 2. Comparison of accuracy and F1 score

4. CONCLUSION

Wood-leaf separation is a crucial step for the applications of TLS in the field of forestry. In this paper, four new geometric feature vectors calculated from three-dimensional coordinates are presented. Considering the difference of regularity or roughness between wood and leaf, fractal dimension is developed for the classification. Zenith angle and variation are presented to distinguish trunks or branches from leaves. The adaptive axis direction of cylinder is adopted to calculate the local point density precisely. Experimental results show that these four new developed feature vectors outperformed the five feature vectors based on eigen values. The proposed method achieved a higher performance in terms of both accuracy and F_1 score. However, the F_1 score for wood is much lower than that for leaf. It is because that many branch points are wrongly classified as leaf points. How to improve the wood classification accuracy will be focused in the future research.

ACKNOWLEDGEMENTS (OPTIONAL)

The authors would like to thank the Funds from National Science Foundation (NSF) (41801325, 41962018), the Science Foundation of Jiangxi Province (20192BAB217010), the China Post-Doctoral Science Foundation (2019M661858), Education Department of Jiangxi Province (GJJ170449), Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology (DLLJ201806), East China University of Technology Ph. D. Project (DHBK2017155) for their financial support.

REFERENCES

Beland, M., Baldocchi, D. D., Widlowski, J., Verstraete, M. M., Richard, A. F., 2014. On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of

- forests with terrestrial LiDAR. *Agric. For. Meteorol.*, 184, 82-97.
- Beland, M., Widłowski, J., Fournier, R. A., Côté J., Verstraete, M. M., 2011. Estimating leaf area distribution in savanna trees from terrestrial LiDAR measurements. *Agric. For. Meteorol.*, 151(9), 1252-1266.
- Calders, K., Origo, N., Disney, M., Nightingale, J., Woodgate, W., Armston, J., Lewis, P., 2018. Variability and bias in active and passive ground based measurements of effective plant, wood and leaf area index. *Agric. For. Meteorol.*, vol. 252, 231–240.
- Calders, K., Schenkels, T., Bartholomeus, H., Armston, J., Verbesselt, J., Herold, M., 2015. Monitoring spring phenology with high temporal resolution terrestrial lidar measurements. *Agric. For. Meteorol.*, vol. 203, 158–168.
- Cote, J., Fournier, R. A., Egli, R., 2011. An architectural model of trees to estimate forest structural attributes using terrestrial LiDAR. *Environ. Modell. Softw.*, 26(6), 761-777.
- Danson, F. M., Sasse, F., Schofield, L. A., 2018. Spectral and spatial information from a novel dual-wavelength full-waveform terrestrial laser scanner for forest ecology. *Interface Focus*, 8(201700492).
- Hancock, S., Gaulton, R., Danson, F. M., 2017. Angular reflectance of leaves with a dual-wavelength terrestrial lidar and its implications for leaf-bark separation and leaf moisture estimation. *IEEE Trans. Geosci. Remote Sensing*, 55(6), 3084-3090.
- Hui, Z.Y., Jin, S. G., Cheng, P. G., Yevenyo, Y. Z., Wang, L. Y., Wang, Y. Q., Hu, H. Y., Hu, Y. J., 2019. An active learning method for dem extraction from airborne lidar point clouds. *IEEE Access*, 7, 89366-89378.
- Hosoi, F., Omasa, K., 2006. A geometric method for wood-leaf separation using terrestrial and simulated lidar data. *Photogrammetric Eng. Remote Sens.*, 81(10), 767-776.
- Hosoi, F., Omasa, K., 2006. Voxel-based 3-D modeling of individual trees for estimating leaf area density using high-resolution portable scanning lidar. *IEEE Trans. Geosci. Remote Sensing*, 44(12), 3610-3618.
- Jin, S. H., Su, Y. J., Wu, F. F., Pang, S. X., Gao, S., Hu, T. Y., Liu, J., Guo, Q. H., 2019. Stem-leaf segmentation and phenotypic trait extraction of individual maize using terrestrial lidar data. *IEEE Trans. Geosci. Remote Sensing*, 57(3), 1336-1346.
- Li, Z., Schaefer, M., Strahler, A., Schaaf, C., Jupp, D., 2018. On the utilization of novel spectral laser scanning for three-dimensional classification of vegetation elements. *Interface Focus*, 8(201700392).
- Li, Z., Strahler, A., Schaaf, C., Jupp, D., Schaefer, M., Olofsson, P., 2018. Seasonal change of leaf and woody area profiles in a midlatitude deciduous forest canopy from classified dual-wavelength terrestrial lidar point clouds. *Agric. For. Meteorol.*, 262, 279-297.
- Ma, L., Zheng, G., Eitel, J., Moskal, L., He, W., Huang, H. B., 2015. Improved salient feature-based approach for automatically separating photosynthetic and nonphotosynthetic components within terrestrial lidar point cloud data of forest canopies. *IEEE Trans. Geosci. Remote Sensing*, 54(2), 1-18.
- Moorthy, S., Calders, K., Vicari, M., Verbeeck, H., 2019. Improved supervised learning-based approach for leaf and wood classification from LiDAR point clouds of forests. *IEEE Trans. Geosci. Remote Sensing*, 1-15.
- Vicari, M. B., Disney, M., Wilkes, P., Wilkes, P., Burt, A., Calders, K., Woodgate, W., 2019. Leaf and wood classification framework for terrestrial LiDAR point clouds. *Methods Ecol. Evol.*, 10(5), 680-694.
- Wang, Y., Pyoralá, J., Liang, X., Lehtomäki, M., Kukko, A., Yu, X. W., Kaartinen, H., Hyypä J., 2019. In situ biomass estimation at tree and plot levels: what did data record and what did algorithms derive from terrestrial and aerial point clouds in boreal forest. *Remote Sens. Environ.*, 232, 11309.
- Xu, H., Gossett, N., Chen, B., 2007. Knowledge and heuristic-based modeling of laser-scanned trees. *ACM Trans. Graph.*, 26(4), 19-es.
- Yang, H. Q., Chen, W. L., Qian, T. L., Shen, D. T., Wang, J. C., 2015. The extraction of vegetation points from lidar using 3d fractal dimension analyses. *Remote Sens.*, 7(8), 10815-10831.
- Zhang, W., Wan, P., Wang, T., Cai, S. S., Chen, Y. M., Jin, X. L., Yan, G. J., 2019. A novel approach for the detection of standing tree stems from plot-level terrestrial laser scanning data. *Remote Sens.*, 11(2), 211.
- Zhu, X., Skidmore, A. K., Darvishzadeh, R., OlafNiemann, K., Liu, J., Shi, Y. F., Wang, T. J., 2018. Foliar and woody materials discriminated using terrestrial LiDAR in a mixed natural forest. *Int. J. Appl. Earth Obs. Geoinf.*, 64, 43-50.