

FOREST SPECIES CLASSIFICATION BASED ON THREE-DIMENSIONAL COORDINATE AND INTENSITY INFORMATION OF AIRBORNE LIDAR DATA WITH RANDOM FOREST METHOD

Haotian You*, Peng Lei, Maosen Li, Fangqi Ruan

College of Geomatics and Geoinformation, Guilin University of Technology, Guilin, China-youht@glut.edu.cn

KEY WORDS: Forest Species, Three-dimensional Coordinate Information, Intensity Information, Airborne LiDAR Data, Random Forest

ABSTRACT:

Forest species is a basic parameter of forest ecosystem. The accurate identification of forest species can not only improve the estimation accuracy of other forest structural parameters, but also have important significance for forest resource monitoring and management. As an active remote sensing technology, the LiDAR could not only acquire the three-dimensional coordinate information of the object, but also acquire the intensity information. The airborne LiDAR data have been successfully used in forest species classification research. However, most of the research is based on the three-dimensional coordinate information of LiDAR data. It's fact that the parameters derived from the intensity data are closely related to the spectral reflection of forest species and could be beneficial for forest species classification, but the research with LiDAR intensity data is fewer. Therefore, it is necessary to explore the potential of LiDAR intensity data on forest species classification and test if the combined application of the three-dimensional coordinate and intensity information can improve the forest species classification accuracy. In this paper, the Moon Lake National Forest Park located in Changchun is selected as the study area, which planted with Scotch pine, Larch pine, Mongolian oak, aspen and other tree species. Two kinds of parameters are separately derived from the three-dimensional coordinate and intensity information of airborne LiDAR data. Then Random Forest is used to classify the forest species based on the above parameters. The main purposes of this study are: (1) to test if the parameters derived from the three-dimensional coordinate information of LiDAR data can be used to identify the forest species; (2) to test if the parameters derived from the intensity information of LiDAR data can be used to identify the forest species; (3) to test if the combined application of the three-dimensional coordinate and the intensity information can improve the accuracy of forest tree species identification. It was found that the classification accuracy of forest species based on structural parameters derived from the three-dimensional coordinate information was 87.54% and Kappa coefficient was 0.81. The classification accuracy based on the parameters derived from LiDAR intensity information was 89.23% and Kappa coefficient was 0.83. And the classification accuracy based on three-dimensional coordinate and intensity information was 92.35% and Kappa coefficient was 0.88. The results demonstrated that both the parameters derived from LiDAR three-dimensional coordinate and intensity information can identify forest species. The results based on LiDAR intensity information are better than that of three-dimensional coordinate information. And the combined application of the two information can improve the classification accuracy of forest species. Therefore, further research should make use of the three-dimensional coordinates and intensity information of LiDAR data to improve the accuracy of results.

1 INTRODUCTION

As the basic horizontal structural parameter of forest, forest classification information plays an important role in forest ecosystem diversity, ecological environment simulation and quantitative research of forest resources[1-3]. For quantitative

study of forest structure, the forest stand type information is indispensable. The classification results accuracy will have a great impact on the subsequent quantitative study of forest structures. This is mainly because most allometric growth models used to estimate forest structures are derived from specific forest stand species. Therefore, how to obtain high-precision classification results of forest stand species has

* Corresponding author

become the primary condition for quantitative study of forest structures.

The airborne LiDAR can not only obtain the three-dimensional structure information of forest, but also obtain the intensity information of forest canopy surface. The three-dimensional information is highly correlated with the spatial distribution of forest, while the intensity information is highly correlated with the reflectivity of forest canopy. Therefore, the LiDAR data are widely used in forest species classification research. For example, Brandtberg[4] used the structural and intensity parameters extracted from airborne LiDAR data to classify three broad-leaved forest species, e.g. Quercus, Pineapple and Acer rubra, in West Virginia. And the classification accuracy is relatively low with only 64%. Holmgren et al. [5] classified the Scotch pine and Norway spruce using the structure and intensity information extracted from LiDAR data. It was found that the classification accuracy could be up to 95%. Vauhkonen et al. [6] and Korpela et al. [7] classified Norway spruce, Scotch pine and broad-leaved trees using the intensity parameters derived from LiDAR data with accuracy of 78% and 88% respectively. The results show that the LiDAR data can achieve the forest species classification. And the classification accuracy of coniferous and broad-leaved species are between the accuracy of broad-leaved species and coniferous species. Moreover, the classification accuracy of coniferous species is higher than that of broad-leaved species. Although the previous studies have explored the application of LiDAR data used for forest tree species classification, they have not systematically studied the three-dimensional spatial information and intensity information contained in LiDAR data. To find the classification differences among different information contained in LiDAR data and the influence of different information combinations on the classification results, a system research is needed.

In this paper, the height distribution and intensity parameters were derived from the pre-processed LiDAR data. Then the extracted parameters were used to classify Scotch Pine, Larch pine, Mongolia oak, Aspen and other tree species with random forest algorithm. It's expected that the results could explore the classification differences among different information and the influence of different information combinations on the classification results and provide guidance and suggestions for LiDAR forest species classification in large areas in the future.

2 STUDY AREA AND METHODS

2.1 Study area

The study area is located in Jingyuetan National Forest Park, southeast of Changchun Jilin Province. Most of the forest is planted forest including Larch pine, Scotch Pine, Aspen and Mongolia oak and other tree species. Among them, the crown of Scotch Pine is relatively large and mostly irregular dome or flat top with longer needles ranging from 4 cm to 9 cm (Figure 1 (a)). The crown of Larch pine is relatively small and mostly conical with relatively short needles ranging from 1.5 cm to 3 cm (Figure 1 (b)). The crown of Aspen is round, and the leaves are triangular oval or nearly round. The length and width of leaves are approximately the same with 3-6 cm (Figure 1 (c)). The crown of Mongolia oak is oval. The leaves are obovate with the length of 7-19 cm and the width is about 3-11 cm (Figure 1 (d)).



(a)



(b)



(c)



(d)

Figure 1 Graphic representation of coniferous forest: a) Scotch pine, (b) Larch pine, (c) Aspen, and (d) Mongolian Oak

2.2 Field data acquisition

To better realize the classification of forest species, the forest stand species of the study area were analyzed with the RCD high-resolution images synchronized with LiDAR data before collecting the field data in 2014. It was found that Scotch Pine, Larch pine, Mongolia oak and Aspen occupied a relatively large area in the study area and therefore were regarded as the main species. At the same time, other tree species in the study area are classified into one group called others. According to the situation in the field data acquiring, the final sampling numbers of different tree species are listed as Table 1.

Table 1 Sampling numbers of different tree species

Species	Scotch	Larch	Mongolia	Aspen	Others
---------	--------	-------	----------	-------	--------

	Pine	pine	oak		
Number	156	135	20	31	10

2.3 Airborne LiDAR data collection and processing

In order to make full use of the effective information contained in airborne LiDAR data, a series of processing is needed before parameter extraction, including denoising, overlapping point removal, classification, interpolation and elevation normalization, etc. The specific processing process is shown in [8,9].

The two parameters are derived from the above pre-processed LiDAR data. One is the height distribution parameters related to the three-dimensional spatial information of LiDAR data and the other is the intensity parameters related to the intensity information of LiDAR data. Detailed description of the parameters is shown in Table 2 and Table 3 [10].

Table 2 The height parameters extracted from LiDAR data

Parameters	Physical meaning
H25、H50、H55、H75、H99	The quantile height of 25%、50%、55%、75% and 99%
Hmax	Maximum height of point cloud data
Hmean	Mean height of point cloud data
HR1	The difference between 75% and 25% quantile height
HCV	Coefficient of variation of height of quadrate point cloud data
HEntropy	Height entropy of point cloud data
Hkurtosis	Height kurtosis of point cloud data
Hskwness	Height skewness of point cloud data
LAD	Leaf area density distribution
LADCV	Coefficient of variation of leaf area density distribution
LADkurtosis	Kurtosis of leaf area density distribution
LADskwness	Skewness of leaf area density distribution
NLPI	Laser penetration index calculated by the number of point cloud data

Table 3 The intensity parameters extracted from LiDAR data

Parameters	Physical meaning
------------	------------------

I _{mean}	Mean intensity of point cloud data
ICV	Coefficient of variation of intensity of point cloud data
I _{skewness}	Intensity skewness of point cloud data
I _{kurtosis}	Intensity kurtosis of point cloud data
V _{I_{mean}}	Mean intensity of vegetative point cloud data
VICV	Coefficient of variation of intensity of vegetative point cloud data
V _{I_{skewness}}	Intensity skewness of vegetative point cloud data
V _{I_{kurtosis}}	Intensity kurtosis of vegetative point cloud data
N _{I_{mean}}	Mean intensity of ground point cloud data
NICV	Coefficient of variation of intensity of ground point cloud data
N _{I_{skewness}}	Intensity skewness of ground point cloud data
N _{I_{kurtosis}}	Intensity kurtosis of ground point cloud data
ILPI	Laser penetration index calculated by point cloud intensity

2.4 Random forest classification method

To obtain more accurate classification results, the random forest algorithm is used in this paper. Random forest is a relatively new non-parametric "ensemble learning" algorithm which consists of a set of decision trees. It first establishes a single decision tree by using some observation values and variables. Then it merges several independent decision trees to obtain a set of decision trees and achieves more accurate and stable prediction results. There are two main parameters in random forest algorithm: (1) *ntree* which stands for classification decision tree; (2) *mtry* stands for variables contained in each node. The experimental results show that

when *nTree* is greater than 500, the classification error tends to be stable. Therefore, the number of decision trees used in this study is the default value (500), and the number of node variables is the root mean square of the input variables.

2.5 Accuracy evaluation

To reduce the impact of accidental factors on the model and further improve the model's accuracy, the method of ten-fold cross-validation was used. The total 321 samples were divided into 10 parts on average, 9 of which were used for modeling each time, and the remaining one was used for validation until all 10 data were used for validation. The mean value of 10 times is taken as the result of the final validation of the model. User accuracy, product accuracy, overall accuracy and kappa coefficient calculated from obfuscation matrix are usually used as evaluation indexes for model accuracy validation.

3 RESULT AND ANALYSIS

3.1 Forest stand species classification results based on LiDAR height parameters

The results of forest species classification based on LiDAR height distribution parameters are shown in Table 4. By analyzing the results shown in Table 4, it is found that the overall classification accuracy of scotch pine, larch pine, mongolia oak, aspen and other tree species is 87.54% and the kappa coefficient is 0.8067. Among the results, the classification accuracy of scotch pine is the highest, the user classification accuracy is 91.67%, and the product classification accuracy is 91.08%. The classification accuracy of larch pine is the second, and the user classification accuracy is 88.15%, the product classification accuracy is 88.15%, followed by mongolia oak and aspen. The classification accuracy of other tree species is the lowest, the user classification accuracy is 70.00%, and the product classification accuracy is 100.00%.

Table 4 The results of forest stand species classification based on LiDAR height parameters

Species	Scotch Pine	Larch pine	Mongolia oak	Aspen	Others	Total	User accuracy (%)
Scotch Pine	143	12	1	0	0	156	91.67
Larch pine	13	119	0	3	0	135	88.15
Mongolia oak	1	0	17	3	0	21	80.95
Aspen	0	4	4	23	0	31	74.19
Others	0	0	0	3	7	10	70.00
Total	157	135	22	32	7	353	
Production	91.08	88.15	77.27	71.88	100		87.54

accuracy (%)

Kappa coefficient

0.8067

3.2 Forest stand species classification based on LiDAR intensity parameters

Based on LiDAR intensity data, the classification results of forest stand type are shown in Table 5. Through the analysis of the results in Table 5, it is found that the overall classification accuracy of scotch pine, larch pine, mongolia oak, aspen and other tree species is 89.23% and the kappa coefficient is 0.8325. Compared with the classification result of LiDAR height distribution parameters, the accuracy is improved by 1.69%. That is, all the user classification accuracy of scotch pine, larch

pine and aspen increased, and the result of larch pine increased the most (4.44%). The user classification accuracy of mongolia oak and other tree species decreased by 28.57% and 10.00%, respectively. The results show that the overall classification accuracy of LiDAR intensity data is higher than that of LiDAR height distribution parameters. It's demonstrated that LiDAR intensity is more suitable for the classification of scotch pine, larch pine and aspen species. The LiDAR height distribution parameters is more suitable for the classification of mongolia oak and other tree species.

Table 5 The results of forest stand species classification based on LiDAR intensity parameters

Species	Scotch Pine	Larch pine	Mongolia oak	Aspen	Others	Total	User accuracy (%)
Scotch Pine	149	7	0	0	0	156	95.51
Larch pine	10	125	0	0	0	135	92.59
Mongolia oak	0	0	11	8	2	21	52.38
Aspen	0	2	5	24	0	31	77.42
Others	0	1	2	1	6	10	60.00
Total	159	135	18	33	8	353	
Production accuracy (%)	93.71	92.59	61.11	72.73	75.00		89.23
Kappa coefficient							0.8325

3.3 Forest stand species classification based on the combination of LiDAR height and intensity parameters

The classification results of forest stand species classification based on the combination of LiDAR height and intensity parameters are shown in Table 6. According to Table 6, it's known that the overall classification accuracy of scotch pine, larch pine, mongolia oak, aspen and other tree species is 92.23% and the kappa coefficient is 0.8813. Compared with the

results of LiDAR intensity data, the accuracy is improved by 3.12%. Compared with LiDAR height distribution parameters, the classification accuracy is improved by 4.81%. The results showed that the comprehensive application of LiDAR height and intensity parameters can improve the classification accuracy of forest stand species, especially for mongolia oak and other tree species.

Table 6 The results of forest stand species classification based on the combination of LiDAR height and intensity parameters

Species	Scotch Pine	Larch pine	Mongolia oak	Aspen	Others	Total	User accuracy (%)
Scotch Pine	149	7	0	0	0	156	95.51
Larch pine	8	127	0	0	0	135	94.07
Mongolia oak	0	0	18	2	1	21	85.71
Aspen	0	2	5	24	0	31	77.42
Others	0	0	0	2	8	10	80.00
Total	157	136	23	28	9	353	
Production accuracy (%)	94.90	93.38	78.26	85.71	88.89		92.35

accuracy (%)

Kappa coefficient

0.8813

3.4 Forest stand species classification result of study area

Through the analysis of all above results, it's known that the classification results based on LiDAR height and intensity parameters are the best, and the overall accuracy is 92.23%. Therefore, the model with LiDAR height and intensity parameters was used to achieve the of forest stand type classification in the study area. The final classification results of the study area are shown in Figure 2.

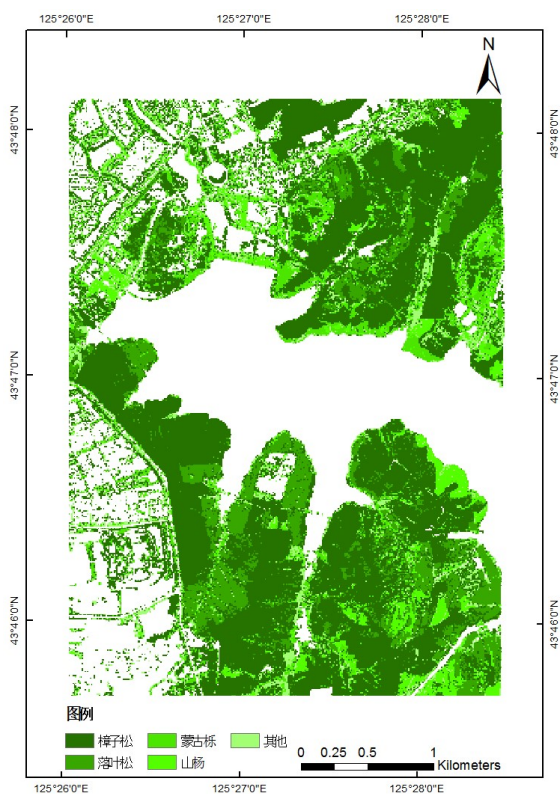


Figure 2 The study area forest stand species classification result

REFERENCES

Leathwick J R. Climatic Relationships of Some New Zealand Forest Tree Species[J]. *Journal of Vegetation Science*, 2010, 6(2):237-248.

Adekunle, V. A. J.Olagoke, A. O. Diversity and biovolume of tree species in natural forest ecosystem in the bitumen-producing area of ondo state, Nigeria: a baseline study[J]. *Biodiversity & Conservation*, 2008, 17(11):2735-2755.

Hudak A T, Crookston N L, Evans J S, et al. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data[J]. *Remote Sensing of Environment*, 2008,

4 CONCLUSION AND DISCUSSION

In this study, a series of height and intensity parameters derived from the processed LiDAR data were applied to the classification of scotch pine, larch pine, mongolia oak, aspen and other tree species, respectively. By analyzing the classification results of different LiDAR parameters, it's concluded that:

- (1) The classification result of LiDAR intensity parameters is better than that of LiDAR height parameters. This is mainly the LiDAR intensity data is closely related to the reflectance of different ground objects and contain more information about forest species, so it is more suitable for the classification of different forest stand species. While the LiDAR height data is useful for forest tree species with different height, but useless for different forest tree species with similar height information.
- (2) The accuracy of the comprehensive classification results of LiDAR height and intensity parameters is better than that of the separate classification results of the two kinds of parameters. Although LiDAR intensity information contains more information about different forest tree species, it still has some limitations for tree species with similar canopy reflectance. However, LiDAR height information has no direct relationship with canopy reflectance of different tree species, but could provide forest tree height information. Therefore, the combination of LiDAR height and intensity information can achieve the complementary advantages of different information, and realize the better result of forest stand type classification.

112(5):2232-2245.

Brandtberg T. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar. *Isprs Journal of Photogrammetry & Remote Sensing*, 2007. 61(5): 325-340.

Holmgren J , Å Persson. Identifying species of individual trees using airborne laser scanner[J]. *Remote Sensing of Environment*, 2004, 90(4):415-423.

Vauhkonen J , Korpela I , Maltamo M , et al. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape parameters[J]. *Remote Sensing of Environment*, 2010, 114(6):1263-1276.

Gullison R E , Frumhoff P C , Canadell J G , et al. Tropical Forests and Climate Policy[J]. *Science*, 316..

You H T, Xing Y Q, Peng T, et al. Effects of Different LiDAR Intensity Normalization Methods on Scotch Pine Forest Leaf Area Index Estimation[J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(02):170–179.

You H T, Xing Y Q, Peng T, et al. Research on the Effect of Side-Overlap between Airborne LiDAR Adjacent Swaths on the Coniferous Forest Structural Parameters Estimation[J]. *Scientia Silvae Sinicae*, 2018, 54(06):109–118.

You H T. Research on forest key structural parameters estimation based on airborne LiDAR data[D]. Northeast Forestry University, 2017.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (41901370, 41801030); Guangxi Natural Science Foundation (2018GXNSFBA281075, 2018GXNSFBA281054); Guangxi Science and Technology Base and Talents Foundation(GuikAD19110064); Research Foundation of Guilin University of Technology (GLUTQD2017094, GUTQDJJ2017069).