Frontiers in Environmental Research
ISSN: 2960-026X
DOI: 10.61784/fer240113

APPLICATION ANALYSIS AND PROSPECTS OF TREE SPECIES CLASSIFICATION BASED ON FORESTRY REMOTE SENSING

Herman R. Mason
Department of Environmental Science and Policy, University of California, Davis.

Abstract: Tree species classification is an important application field in forestry remote sensing. It has widespread application scenarios in the fields of sustainable forestry management, biodiversity research and ecological environment protection. Combining the research results in this field after 2000 and the application practice in the forestry production process in recent years, the application of multi-source data in tree species classification is summarized. From the perspectives of workflow and corresponding algorithms, this problem was analyzed and compared based on image classification and mathematical statistics. Facing the problems and challenges encountered in the application of forestry remote sensing in tree species classification, different working ideas based on semantic segmentation and instance segmentation were proposed, and the development prospects of future multi-source remote sensing data fusion acquisition and hardware processing equipment were prospected.
Keywords: Forestry remote sensing; Tree species classification; Lidar data; Hyperspectral data; Theoretical application

1 APPLICATION OF MULTI-SOURCE REMOTE SENSING IN TREE SPECIES CLASSIFICATION

Forestry undertakes major national tasks such as forest and wetland ecological protection, biodiversity protection, and desertification control. It also shoulders the responsibility of protecting resources and the environment and maintaining ecological security. Forest is an important part of the terrestrial ecosystem and has a huge carbon sequestration function, which makes forestry play a decisive role in maintaining ecological balance, and plays a special role in maintaining ecological security and responding to climate change [1]. Remote sensing technology has been widely used in various fields, and forestry remote sensing technology, as an indispensable component, can not only obtain data on forestry resource management, but also further reveal the ecological impact of forestry management. With the rapid development of emerging high- resolution remote sensing equipment and technologies such as synthetic hyperspectral, thermal infrared, lidar, and aperture radar, remote sensing has developed rapidly in quantitative and qualitative mapping of forest biological and ecological attributes. With fine spectral resolution and With the rapid development of active remote sensing and the continuous improvement of tree species identification technology, remote sensing can identify various attributes of the forest on a finer scale [2]. Tree species classification is of great significance for sustainable forestry management, biodiversity research, and ecological environment protection. However, the plant species in the forest itself have a high diversity, making tree species classification on a large scale still a big challenge. This article analyzes and compares research on tree species classification from remote sensing data sources, tree species classification processes and methods, and proposes problems faced in production practice and prospects for the future.

1.1 Tree Species Classification Based on Traditional Visible Light Multispectral and Other High-Resolution Remote Sensing

Visible light refers to the wavelength range of approximately 0.38 ~ in the electromagnetic spectrum The part between 0.76 μm is named because it can be perceived by the human eye. It generally has three bands of red, green and blue. It is the most basic and most commonly used working band in aerial remote sensing and surveying and mapping research. Although visible light data contains less spectral information, with the launch of high-resolution satellites, especially domestic high-resolution civilian satellites, and the rise of the drone industry, it has become more convenient to obtain and has higher spatial resolution [3] , coupled with the relatively mature data processing technology, visible light has become the most common forestry remote sensing data source.
Multispectral data is to obtain spectral information such as target surface brightness through multi-band detectors. Generally, there are several to dozens of bands. The wavelength range is larger than visible light. Its acquisition cost is lower than data sources such as hyperspectral and lidar. Some can meet the requirements of high timeliness at the same time.

1.2 Tree Species Classification Based on Hyperspectral and Mid-Infrared/Thermal Infrared

Hyperspectral data is currently the most widely used data in forestry remote sensing tree species classification. Through statistical analysis of articles on the application of remote sensing data to study plant species classification and identification since 2000, it was found that among 120 research cases, more than 1/3 of the studies used Hyperspectral data were used[3]. Hyperspectral remote sensing data mostly refers to visible light/near-infrared (450-970nm) and short -wave red in the electromagnetic spectrum. In the outer band (970 to 2 500 nm), many of the acquired wavelengths are very narrow. Continuous spectrum image data can contain dozens to hundreds of spectral band information. These two

bands are important wavelengths for studying plant classification. Area, the spectral information of these two bands can be used to calculate various vegetation indices, making it possible to classify tree species more accurately. Commonly used vegetation indices include normalized vegetation index (NDVI), photochemical reflectance Rate Index (PRI), Optimized Soil Adjusted Vegetation Index (OSAVI), Normalized Infrared Index (NDII), Normalized Water Index (NDWI), etc. Mid-infrared and thermal infrared, as supplementary bands of airborne or spaceborne hyperspectral, can be obtained through ground supplementary surveys using portable infrared spectrometers.
This type of data is relatively effective in the research field by classifying the huge differences in physical and chemical properties, moisture content, morphology, etc. of different tree species, but the cost is relatively high. At the same time, the common problems of spectral data such as "same objects with different spectra", "different objects with the same spectrum", and greater influence from the meteorological environment cannot be completely solved. The current conditions for acquisition and application on a large scale are not mature enough.

1.3 Tree Species Classification Based on Active Ground Remote Sensing such as Lidar and Synthetic Aperture Radar

LiDAR is an active remote sensing system that emits a laser beam to detect the position, speed, shape and other characteristics of a target. Its data can be obtained from spaceborne, airborne, and ground equipment. The information of trees is reflected in point cloud data. By processing point cloud data, information such as tree height and crown width can be extracted. Ground lidar can also obtain the texture characteristics of branches and leaves. However, due to its acquisition cost, relatively long data processing cycle and the impact of point cloud density on research results, it is often used as a supplement or auxiliary to spectral identification in the field of tree species identification, and the research results are relatively good. Although there are certain exploratory results based solely on lidar data for tree species classification, its practicality remains to be further studied.
Synthetic aperture radar (SAR) is also an active remote sensing system for earth observation. The data is mostly obtained by being mounted on satellites or aircraft platforms. It has certain surface penetration capabilities and is relatively less affected by meteorological conditions. When obtaining species information, the main way is to classify tree species through the impact of tree canopy structure, water content, and surface moisture on SAR signals. As an emerging data source in the field of tree species identification, related research is still being gradually explored and has certain prospects.

2 GENERAL PROCESS AND MAINSTREAM ALGORITHMS FOR FORESTRY REMOTE SENSING TREE SPECIES CLASSIFICATION

2.1 General Process of Tree Species Classification

These steps include data acquisition, data preprocessing, single tree segmentation, feature extraction, sample labeling, execution classification, accuracy and stability testing.
Data acquisition generally includes two parts: the acquisition of remote sensing data and forestry survey data. Data preprocessing depends on the data source and processing methods. Common ones include correction registration, abnormal data processing, irrelevant item elimination, etc. Single tree segmentation is a key step in tree species classification, which has a direct and obvious impact on subsequent classification results. The effect of single tree segmentation itself is also affected by many factors, so the difficulty varies greatly from easy to difficult. Relatively speaking, the artificial forest with a simple terrain and a single tree species has a better single-tree segmentation effect. The more complex the terrain, the more diverse the tree species, and the more chaotic the natural growth conditions, the greater the challenges faced by single-tree segmentation. Effective and widely applicable segmentation The method remains to be further explored. Feature extraction is to extract corresponding features from the data according to different research methods, and some of them need to be fused or processed for classification. Sample labeling is a relatively tedious process. Some studies have proposed semi-automatic or automatic tree species identification methods, which are semi-supervised or unsupervised learning methods in machine learning, which reduce the number of samples that need to be labeled. Execution classification, as the main part of tree species classification, is classified according to the corresponding algorithm, which is mainly divided into two categories: pixel-based and object-based. Specific algorithms emerge in endlessly after the intervention of machine learning. Accuracy and stability testing are a direct reflection of classification results. Cross-validation is usually used, or a confusion matrix is used to obtain various accuracy, Kappa coefficient and other indicators. Although the quality of the method cannot be directly evaluated, it can be used as the same method. Below is a comparison of different algorithms.

2.2 Comparison of Mainstream Algorithms and Characteristics of Tree Species Classification

2.2.1 Tree species classification algorithm based on hypothesis testing

As a branch of classification and identification, the development of tree species classification algorithms is closely linked to the field of mathematical statistics. Therefore, the mainstream algorithms for tree species classification can be divided into unsupervised classification and supervised classification from the perspective of hypothesis testing (semi-automatic classification can be considered as a combined or integrated use of the two).

1) Unsupervised classification algorithm. It was more common in the early days, such as K - The advantage of mean clustering algorithm, EM algorithm, etc. is that it does not require a priori hypothesis processing of the data, nor does it require a large number of training data sets, and can even reduce the sample labeling process. However, the tree species classification accuracy is relatively average, so it is relatively difficult to use. Come less and less.

2) Supervised classification algorithm. Such as KNN algorithm, maximum likelihood estimation algorithm, etc., which are characterized by the fact that the data must go through a priori assumptions, and different numbers of training samples are required according to different needs.

2.2.2 Tree species classification algorithm based on parameter estimation

As another perspective of mathematical statistics, from the perspective of parameter estimation, tree species classification can be divided into parameterized classification algorithms and non-parametric classification algorithms.

1) Classification algorithm with parameters. Such as linear regression, perceptron algorithm, logistic regression algorithm, etc., because they are fast and require a relatively small amount of data, they are also used in many applications, but under-fitting is more common and the stability is correspondingly poor.

2) Non-parametric classification algorithms. For example, decision tree algorithm, SVM algorithm, random forest algorithm, neural network algorithm, etc., with the development of the times, gradually play the mainstay role in mainstream algorithms. It is characterized by strong fitability and relatively high classification accuracy, but the data requirements are correspondingly large, the time consumption will increase accordingly, overfitting is prone to occur, and the interpretability is also poor.

It is worth noting that in the same research environment and methods, different algorithms can be measured and compared in terms of accuracy, stability, time-consuming, etc., but under different research objects, locations, data sources, etc., different algorithms Algorithms are not directly comparable. Most studies on tree species classification show that under different conditions, the data itself often has a more direct impact on the algorithm effect.

3 PROBLEMS AND PROSPECTS FACED BY THEORETICAL RESEARCH AND APPLICATION PRACTICE

3.1 Application of Semantic Segmentation and Instance Segmentation in Tree Species Classification

1) Although from the perspective of research objects, tree species classification has spread from the subarctic to the tropics, and there are more than 40 tree species involved in tree species classification, and the classification accuracy has increased from 55% at the beginning to some studies. 95% or even higher, but it cannot be ignored that research with higher classification accuracy often has greater limitations, and there is also a huge difference in research cost application and actual production application cost, so it is practical, effective and feasible in production application methods are still relatively lacking.

2) The theory of tree species classification is based on image classification and recognition. Therefore, extensive tree species recognition research cannot meet the current complex environment in my country, and the weak forestry production base cannot adequately support the transformation of theoretical research into practice. perhaps we can trace the source and change the perspective to think about the application of forestry remote sensing in tree species classification, instead of just focusing on the improvement and optimization of algorithms at the research level.

3) Affected by the wave of refined classification, most of the existing tree species classification methods are based on instance segmentation, which can classify "tree a" is tree species 1, tree b is tree species 2, tree c is tree species 1, and tree d is tree species 3." But in fact, the required tree species classification satisfies "1 tree species include a and c, 2 tree species include b, and 3 tree species include d", which is the degree of semantic segmentation. Relatively speaking, the workload of semantic segmentation will be smaller than that of instance segmentation, and from the perspective of practical forestry applications, even semantic segmentation can fully satisfy the existing management and usage requirements, so perhaps such a change can solve the bottleneck of some existing methods from research to application.

3.2 Integration of Multi-Source Remote Sensing Data and Development of Hardware Equipment

The fusion of multi-source remote sensing data in tree species classification research is becoming more and more common [5]. The advantages of spectral information in reflecting the physical and chemical properties of trees and the horizontal level, and the advantages of point cloud information in reflecting the vertical structure, are integrated with each other and increasingly become tree species. Popular topics in classification development[6]. With the launch of domestic high-scoring civilian satellites and the rapid development of the drone industry, data sources have become more extensive. At the same time, in the development process of various types of spectroscopic instruments and radar instruments, they are also being integrated [7], which not only meet the requirements of different resolutions or densities, but also can quickly and directly obtain the integrated information for This will lay a good foundation for subsequent applied research.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Li Zengyuan, Qin Xianlin, Gao Zhihai, et al. Research on forestry application of high-resolution remote sensing. Satellite Applications, 2018(11): 61-65.
[2] Wu Nan, Li Zengyuan, Liao Shengxi, etc. Overview and prospects of forestry remote sensing application research at home and abroad. World Forestry Research, 2017, 30(6): 34-40.
[3] Hu Jianbo, Zhang Jian. Progress in the application of UAV remote sensing in ecology. Acta Ecologica Sinica, 2018, 38(1): 20-30.
[4] Kong Jiaxin, Zhang Zhaochen, Zhang Jian. Plant species classification and identification based on multi-source remote sensing data: Research progress and prospects. Biodiversity, 2019, 27(7):796-812.
[5] Xie Zhuli. Research on plantation tree species classification based on multi-source high-resolution remote sensing data. Hangzhou: Zhejiang Agriculture and Forestry University, 2019.
[6] Dong Wenxue. Research on remote sensing monitoring of subtropical forest tree species diversity based on airborne lidar and hyperspectral data. Beijing: University of Chinese Academy of Sciences (Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences), 2018.
[7] Lin Yi, Eetu Puttonen, Juha Hyypp. Hyperspectral lidar: three-dimensional biophysical and chemical ecometry. Remote Sensing Information, 2017, 32(1): 5-9.