УДК 004.67 http://doi.org/10.15407/pp2024.02-03.313

O.V. Tsaryniuk, A.M. Hlybovets

DEVELOPMENT OF HEIGHT-BASED VEGETATION SEGMENTATION METHODS: EVALUATING EFFICIENCY AND ACCURACY

Height-based vector vegetation segmentation is one of the critical aspects of spatial analysis. This segmented data is used in radio propagation modeling, environmental monitoring, and vegetation mapping. Many studies on vector vegetation segmentation focus on delineating individual tree crowns, allowing detailed data sets to be obtained. However, the high level of detail results in a substantial data volume, making it impractical to use these datasets over large areas, such as an entire country. Segmentation of large vector data sets remains a significant challenge in geospatial data creation. In our study, we developed three different segmentation methods: hexagon segmentation, convolution segmentation, and random points method. A test data fragment was processed to compare the proposed methods and accuracy and volume metrics were calculated.
Keywords: vegetation segmentation, spatial analysis, hexagonal grid, random points, convolution filters.

О.В. Царинюк, А.М. Глибовець

РОЗРОБКА МЕТОДІВ СЕГМЕНТАЦІ РОСЛИННОСТІ ЗА ВИСОТОЮ ТА ОЦІНКА ЇХ ЕФЕКТИВНОСТ І ТОЧНОСТІ

Сегментація векторної рослинності за висотою є одним із важливих етапів просторового аналізу. Цей тип даних використовується у побудові моделей розповсюдження радіосигналів, екологічному моніторингу та картографуванні рослинності. На сьогодні існує багато досліджень із сегментації векторної рослинності, що зосереджені на виокремленні індивідуальних крон дерев та дозволяють отримати деталізовані набори даних. Але наслідком високої деталізації є суттєвий обсяг, який унеможливлює використання цих даних на великих територіях, наприклад, в масштабі цілої країни. Сегментація великих масивів векторних даних досі є суттєвим викликом у сфері створення геопросторових даних. В процесі нашого дослідження ми розробили три різні методи сегментації: сегментація шестикутниками, сегментація за допомогою згорткових фільтрів та метод випадкових точок. Для порівняння запропонованих методів був опрацьований тестовий фрагмент даних та прораховані метрики точності та об’єму.
Ключові слова: сегментація рослинності, просторовий аналіз, шестикутна сітка, випадкові точки, згорткові фільтри.

Introduction

Integrating diverse datasets is a pivotal challenge in geospatial data production, particularly in vegetation analysis, where combining vector-based vegetation cover with canopy height models (CHM) is essential for depth-enhanced segmentation. This study tackles such integration, aiming to segment vegetation based on height — a crucial step for comprehensive environmental and geographical analyses. Through the lens of satellite and aerial imagery, vegetation segmentation unlocks insights into vegetation distribution, health, and variety across vast areas. We introduce and assess three segmentation methods: Hexagon Segmentation, Convolution Segmentation, and Random Point Method, prioritizing their applicability to large-scale datasets, potentially encompassing entire countries. This comparative evaluation showcases the method's precision and practicality and advances our methodological toolkit for environmental studies.

1. Literature review

Image segmentation is one of the most challenging tasks in image processing. Currently, there are numerous approaches and methods for image segmentation, such as the hexagon segmentation method Hofmann & Tiede (2014) and the Point Initialization Approach Mueller & Corcoran (2021). Most of the research in vegetation segmentation has focused on identifying individual tree crowns. This direction has been instrumental in detailed studies of forest ecosystems, as exemplified by the works of Douss et al. (2022), Li et al. (2014), Lindberg et al. (2021), and Jakubowski et al. (2013). These studies have significantly advanced our understanding of individual tree characteristics, forest structure, and biomass distribution.

In contrast to the detailed focus on individual tree crowns, our research aims to develop a method for generalized segmentation that represents large arrays of vegetation with similar (or nearly identical) heights. This approach is well-suited for segmenting vegetation over vast areas, such as entire countries, addressing the need for macro-level vegetation analysis. Such analysis is essential for regional and national environmental assessments, land use planning, and large-scale conservation efforts.

Our study on vegetation segmentation will leverage canopy height model (CHM) data with a 10-meter resolution, as developed by Liu et al. (2023). This CHM data is crucial for our methodology as it provides a detailed representation of vegetation height across large areas. Using a 10-meter resolution matrix allows for a fine-grained analysis of vegetation structure, making it manageable for large-scale applications like country-wide segmentation.

2. Methodology

We developed three distinct methods to address the challenge of segmenting vegetation based on height. We aimed to understand the complexity of accurately determining vegetation at different altitudes on large datasets. A series of specific metrics were selected to assess the effectiveness and appropriateness of these approaches. These metrics serve as a foundation for evaluating each method's performance, ensuring a balanced analysis between the innovative aspects of our methodologies and their practical outcomes.

The following metrics were used for comparison:

Accuracy (1). This is the ratio of correctly identified pixels, TruePixels (2) to the total number of pixels. It is a straightforward measure of how accurately a model classifies or segments pixels.

$$Accuracy = \frac{TruePixels}{TotalNumberofPixels} \quad (1)$$

Where: *Total Number of Pixels* is the sum of all pixels within all vegetation segments.

$$TruePixels = \sum_{i=0}^{n} \left(\left| h_{input}(p) - h_{output}(p) \right| \leq 3 \right) \quad (2)$$

Where: $h_{input}(p)$ is the height associated with pixel p in the input data, $h_{output}(p)$ is the height associated with pixel p in the output data, as determined by the segmentation process.

Volume. This metric is expressed in the number of vertices after segmentation. It reflects the segmentation's complexity and detail. A more significant number of vertices usually implies a more detailed segmentation but negatively affects the display speed and processing.

The Hexagon segmentation method involves creating a hexagonal grid with uniform hexagons (each side is 100 meters long) and generalizing the height matrix to a 3-meter interval. The vegetation vector is clipped according to the hexagon grid to form segments. Heights from the height matrix are then assigned to each segment, with the most frequent height value in the segment being selected (using the MODE function). Adjacent segments with the same height are merged.

Statistics are computed for each height value and the number of coordinates necessary for comparing the methods.

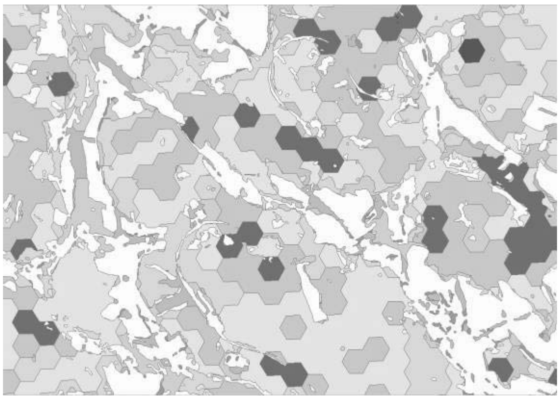

Fig. 1. Result of the hexagon segmentation method

Like the first method, the **Convolution Segmentation Method** also generalizes the height matrix to a 3-meter interval. The matrix is then generalized using a convolutional filter. Several iterations with different convolutions (7x7, 9x9) are conducted using the "Majority" operation, selecting the most frequently occurring value, as in the first method. The generalized matrix is then converted into vector polygons and intersected with the vegetation vector.

Final statistics, including accuracy and volume, are calculated similarly to the first method.

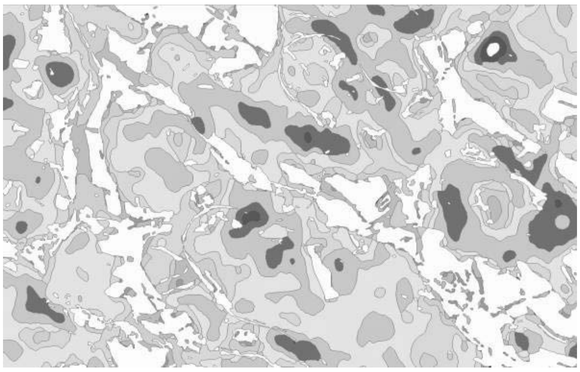

Fig. 2. Result of the convolution segmentation method

The Random Point Method is based on creating random points within a vegetation polygon using several approaches: 1) Generation of random points across the bounding box of the polygon; 2) Generating points along the central line of the polygon; 3) Extracting the central point of the polygon. Utilizing different approaches for point generation ensures an even coverage of all types of polygons with points. The next step involves using the ArcGIS procedure 'Generate Subset Polygons' to construct Thiessen polygons for a given set of points.

The methodology for assigning elevations to segments follows the approach established in previous methods. Each segment intersects with a generalized elevation matrix up to 3 meters. The elevation assigned to each segment is determined by the most frequently occurring pixel values within that intersection. This technique ensures consistency in elevation assignment across different segments, leveraging the established practices from prior methodologies for effective elevation mapping.

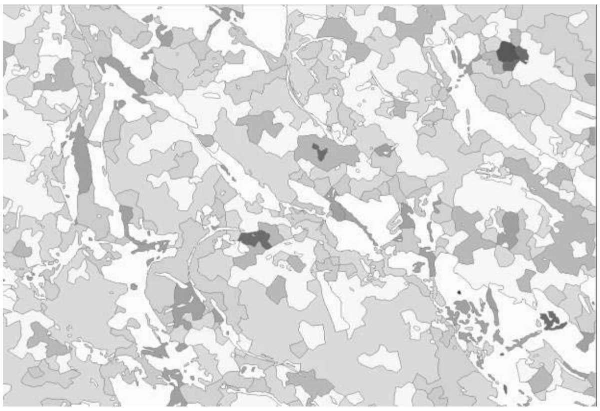

Fig. 3. Result of random point segmentation method

3. Evaluation of the quality of the proposed approaches

For this study, a test site covering an area of 430 square kilometers in the western Czech Republic was selected as the primary focus. The data concerning vegetation heights was sourced from a detailed 10-meter Canopy Height Model (CHM), as elaborated in the research conducted by Liu et al. (2023). The vegetation data itself was derived from a comprehensive vector dataset. This dataset was generated through machine learning techniques to automatically analyze high-resolution satellite imagery, a process meticulously carried out by the Visicom company.

Fig. 4. Research area location

The methods discussed in this article, as well as the analysis of the results, were implemented on PC using the Feature Manipulation Engine (FME). The obtained Accuracy and Volume results are shown in Tables 1,2,3.

Table 1

Hexagon method statistics

Vegetation Height	Accuracy %	Total pixels in CHM	Volume
0	66.22	980	
3	58.92	1020	
6	87.51	4485	
9	94.05	28114	
12	93.19	80631	
15	89.95	145203	
18	82.01	219782	558338
21	80.75	343390	
24	82.07	512259	
27	85.62	749204	
30	88.73	905916	
33	90.17	517650	
36	90.29	94701	
39	84.83	3723	

Table 2

Convolution method statistics

Vegetation Height	Accuracy %	Total pixels in CHM	Volume
0	76.29	949	
3	58.6	1256	
6	69.77	8657	
9	79.43	44705	752412
12	87.72	98143	
15	92.41	156412	
18	95.2	238215	
21	96.38	360859	
24	97.26	534555	
27	98.17	741339	
30	98.92	836759	
33	99.39	485668	
36	99.65	94605	
39	99.73	5176	

Table 3

Random point method statistics

Vegetation Height	Accuracy %	Total pixels in CHM	Volume
0	65.57	909	
3	58.26	1567	
6	83.38	6361	
9	88.01	41188	
12	88.55	87758	
15	82.67	141607	
18	82.01	213139	
21	80.75	360787	737853
24	82.07	542611	
27	85.62	780707	
30	88.73	905794	
33	90.17	516782	
36	90.29	93784	
39	84.83	4905	
42	86.36	374	

To evaluate the segmentation's accuracy, 3-meter height ranges were selected. After testing various height range options (1m, 3m, and 5m), the 3-meter range was chosen as optimal. This selection was based on its ability to accurately reflect the vegetation's true height while minimizing the amount of "noise" from individual pixels with varying heights. This compromise ensures a balance between precision and the reduction of outliers, providing a more reliable assessment of segmentation performance.

We did not consider the performance evaluation of the segmentation methods within the scope of this study. This decision was based on the understanding that performance assessments conducted on a limited test dataset would not yield representative results.

Conclusion

The findings emphasize the potential of integrating high-resolution satellite imagery and LiDAR data with advanced segmentation techniques to enhance understanding of forest ecosystems and vegetation distribution. The hexagon segmentation method provides detailed insights through a hexagonal grid, convolution segmentation leverages convolutional filters for generalized analysis, and the random points method introduces a novel segmentation approach through random point generation and Thiessen polygons.

The research contributes to environmental science by proposing a scalable and efficient methodology for vegetation analysis over large geographical areas. Utilizing canopy height model data with a 10-meter resolution demonstrates the feasibility of these methods for country-wide vegetation segmentation, highlighting their potential for regional and national environmental assessments, land use planning, and conservation efforts.

The comparative analysis reveals that each method has its merits in terms of accuracy and volume of the final segmented vector. The choice of method may depend on specific research needs, available computational resources, and the scale of the analysis. Future work should focus on refining these methodologies, exploring their application in different ecological contexts, and integrating additional data sources to enhance the accuracy and utility of vegetation segmentation for environmental monitoring and management.

Considering the rapid development and high efficiency of machine learning methods, future development of this research aims to incorporate AI-based approaches alongside the methods already compared. The introduction of the Segment Anything Model (SAM) is planned. SAM, an innovative AI-driven method, promises to enhance segmentation accuracy and efficiency by leveraging advanced machine learning algorithms capable of adapting to various vegetation and height delineation tasks. This expansion will comprehensively evaluate traditional segmentation techniques against AI-powered models, potentially setting a new benchmark in vegetation segmentation methodologies.

Additionally, plans are underway to apply the described segmentation methods to large countrywide datasets. In this context, it would be prudent to analyze each method's performance speed and calculate the computational resources required for its implementation. This comprehensive evaluation will ensure the methods' scalability and efficiency when applied to extensive data sets.

Authorship Contribution Statement

A. Hlybovets: Selection of metrics and assessment of the complexity of the proposed algorithms.

O. Tsaryniuk: Development and implementation of segmentation methods.

References

1. R. Douss, I.R Farah, Extraction of individual trees based on Canopy Height Model to monitor the state of the forest. Trees, Forests and People 8, 2022, doi: 10.1016/j.tfp.2022.100257.
2. P. Hofmann, D. Tiede,Image segmentation based on hexagonal sampling grids, South-Eastern European Journal of Earth Observation and Geomatics 3, 2014 pp. 173-177.
3. M.K. Jakubowski, W. Li, Q. Guo, M. Kelly, Delineating individual trees from lidar data: A comparison of vector- and raster-based segmentation approaches. Remote Sensing 5(9), 2013, pp. 4163–4186. doi: 10.3390/rs5094163.
4. W. Li, Z. Niu, S. Gao, N. Huang, H. Chen, Correlating the horizontal and vertical distribution of LiDAR point clouds with components of biomass in a Picea crassifolia forest. Forests 5(8), 2014, pp. 1910–1930. doi: 10.3390/f5081910.
5. E. Lindberg, J. Holmgren, H. Olsson, Classification of tree species classes in

a hemi-boreal forest from multispectral airborne laser scanning data using a mini raster cell method. International Journal of Applied Earth Observation and Geoinformation 100, 2021, doi: 10.1016/j.jag.2021.102334.

6. S. Liu, et al, The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. Science Advances 9(37), 2023, doi: 10.1126/sciadv.adh4097.
7. L. Ma, Y. Gao, T. Fu, L. Cheng, Z. Chen, M. Li, Estimation of Ground PM2.5 Concentrations using a DEM-assisted Information Diffusion Algorithm: A Case Study in China. Scientific Reports 7(1), 2017, doi: 10.1038/s41598-017-14197-z.
8. J.N. Mueller, J.N. Corcoran, A Random Point Initialization Approach to Image Segmentation with Variational Level-sets. 2021, http://arxiv.org/abs/2112.12355.
9. R. Weibel, Using Vector and Raster-Based Techniques in Categorical Map Generalization, 1999.

Одержано: 17.04.2024
Внутрішня рецензія отримана: 25.04.2024
Зовнішня рецензія отримана: 30.04.2024

Про авторів:
Царинюк Олександр Васильович,
Phd Комп'ютерні науки,
аспірант.
https://orcid.org/0000-0003-1394-2040

Глибовець Андрій Миколайович,
доктор технічних наук,
професор,
декан факультету
https://orcid.org/0000-0003-4282-481X

Місце роботи авторів:

Національний університет
«Києво-Могилянська академія»
https://www.ukma.edu.ua/