Research on 3D Building Individualization Technology Based on Spatial Constraints

Zhaoyang Liu[1], Liang Huo[1], Tao Shen[1], Fulu Kong[1]

[1] School of Geomatics and Urban Information, Beijing University of Civil Engineering and Architecture, Beijing 102612, L517749444@163.com

Keywords: 3D buildings; spatial constraints; oblique photogrammetry; individualization.

Abstract

As urban digital transformation advances, accurately extracting three-dimensional building footprints in complex urban environments has become a key challenge in 3D scene applications. Traditional building footprint extraction methods primarily rely on geometric features or texture information, but in densely built or occluded urban environments, achieving ideal extraction results is often difficult. To address this issue, this paper proposes a 3D building footprint extraction method based on spatial constraints. The method designs spatial constraint rules based on three dimensions: height, direction, and distance, combining geometric features with spatial distribution patterns. By adaptively adjusting thresholds, the method effectively improves building footprint extraction in dense urban environments. Experimental results show that the proposed method achieves an IoU of 91.5% in commercial dense areas (a 21.3% improvement over traditional methods), reduces directional error to 3.2° (a 74% decrease), increases recall rate in occlusion scenes to 89.4%, and processes a single scene in only 218 seconds (a 32% reduction). Memory usage is also reduced by 26%. This research provides a high-precision, high-efficiency solution for urban digital modeling, especially suitable for large-scale applications in planning, design, and disaster emergency management.

1. Introduction

The digital transformation of cities is progressing at an astonishing rate, and the application of 3D building models is becoming increasingly important across various fields, particularly in urban planning, architectural design, disaster simulation, and emergency management. As urbanization accelerates, the number of buildings in cities continues to rise, and the issues of occlusion and adjacency between buildings are becoming more severe. This makes the accurate extraction of independent building masses from 3D city models a highly challenging task. 3D building models are not only the foundation of urban digitization but also an essential support for smart city management and sustainable development. Efficient and accurate extraction of building masses has become one of the core issues in current urban 3D modeling technology.

Traditional building mass extraction methods typically rely on geometric features, texture information, or surface morphology to identify and extract buildings. For example, methods based on building outer contours and surface texture analysis use the geometric shape of buildings for segmentation. However, as urban environments become more complex, particularly in densely built areas, the limitations of these traditional methods are becoming apparent. First, geometric features may be affected by noise interference (such as trees, vehicles, etc.), leading to a decrease in segmentation accuracy. Second, in high-density urban environments, the boundaries between buildings are often occluded, and traditional methods struggle to effectively handle these occlusions, resulting in inaccurate extraction results. Thus, how to quickly and accurately extract independent building masses in complex urban environments has become a pressing technical challenge.

Existing building mass extraction methods can be broadly classified into three categories: traditional geometric segmentation methods, deep learning-driven methods, and constraint optimization-based methods. Frommholz et al. (2016) proposed a geometric constraint segmentation method based on multi-view imagery, which relies on geometric features to extract building boundaries. However, this method does not effectively address the problem of dense occlusion, especially in high-density urban environments, where it shows certain limitations. Ruf et al. (2018) used transfer learning, leveraging a large amount of labeled data to improve building segmentation accuracy. However, this method is highly dependent on labeled data, and the segmentation results are still limited in certain complex environments. Wei et al. (2023) proposed a clustering algorithm based on adaptive spatial constraints, improving building segmentation accuracy, but this method still fails to effectively optimize the constraint weights in building mass extraction scenarios, resulting in suboptimal extraction results in some situations.

Although existing research has made progress in building mass extraction, there are still several shortcomings: First, static rules have poor adaptability, and fixed thresholds and weights cannot handle diverse scene changes, especially in dense and complex urban environments. Second, current methods apply different types of spatial constraints (such as height, direction, and distance) separately, lacking an effective collaborative optimization mechanism, which limits the overall performance of building mass extraction. Therefore, how to improve building extraction accuracy and enhance the adaptability of the method in different urban environments using multi-dimensional spatial constraints has become a key issue that needs to be addressed.

To address the above issues, this paper proposes a building mass extraction method based on multi-dimensional spatial constraints. The main contributions of this method include: First, a multi-dimensional constraint model is constructed, integrating spatial constraints such as height, position, and distance, and a dynamic weight allocation framework is designed to allow different constraints to be flexibly adjusted based on the scene. Second, a scene-adaptive mechanism is proposed, which dynamically adjusts thresholds and weights using statistical analysis, thereby enhancing the method's adaptability in dense urban environments and improving the accuracy and robustness of building extraction. Through these innovative designs, the proposed building mass extraction method performs well in the diversity and complexity of urban environments, providing strong technical support for future urban 3D modeling and intelligent analysis.

2. Spatial Constraints and Building Individualization Method

2.1 Design of Spatial Constraint Rules

This paper proposes a multi-dimensional spatial constraint model that optimizes building individualization accuracy through three levels of rules: height, distance, and position.

2.1.1 Height Constraints

Building Height: Based on absolute elevation and relative height difference, dynamic thresholds are set to exclude low-attached structures (e.g., vegetation).

$$H_{rel} = H_{top} - H_{base} \tag{1}$$

Where H_{rel}= The relative height
H_{top}= The top elevation of the building
H_{base} = The base elevation of the building

Height Gradient: Local height variation rates identify abrupt roof slope changes, aiding the segmentation of complex building structures.

2.1.2 Distance Constraints

Inter-Building Spacing: The minimum bounding box distance (d_{min}) between adjacent buildings is defined, with density-adaptive threshold adjustments (1.0–2.0m).
Boundary Distance Constraint: The projection distance from building contours to roads/green spaces ensures individualization results comply with urban planning standards.

2.1.3 Position Constraints

Planar Position: Geographic coordinate alignment ensures segmentation results are spatially consistent with the real-world scene.
Relative Position: Analysis of building cluster topology (e.g., front-back, left-right arrangements) prevents mis-segmentation due to occlusions.
Directional Position: Extraction of the building principal axis direction angle, optimizing segmentation boundaries for densely aligned buildings.

2.2 Integration of Spatial Constraint Rules

Height, position, and distance constraints are weighted to calculate their significance in building individualization, with adaptive parameter threshold adjustments based on actual conditions.
The combined constraint score is formulated as:

$$H_{norm} = H_{th} - H_{rel} \tag{2}$$

$$D_{align} = (cos(|\ \theta_i - \theta_j\ |)) \tag{3}$$

$$L_{interval} = d_{th} - d_{min} \tag{4}$$

Where H_{norm} = Normalized height
H_{th} = dynamic segmentation threshold
D_{align}= Directional alignment
θ = principal axis direction angle
d_{th} = adaptive distance threshold

Using an entropy weight method for dynamic weight allocation, the comprehensive constraint score is given by:

$$S = \alpha H_{norm} + \beta d_{min} + \gamma D_{align} \tag{5}$$

where S = segmentation condition
If S>0.8, adjacent regions are classified as the same building.
If S<0.5, they are classified as separate buildings.
If 0.5≤S≤0.8, a secondary segmentation step is triggered (e.g., contour curvature refinement).
This multi-level rule collaboration significantly improves robustness and adaptability in complex urban environments.

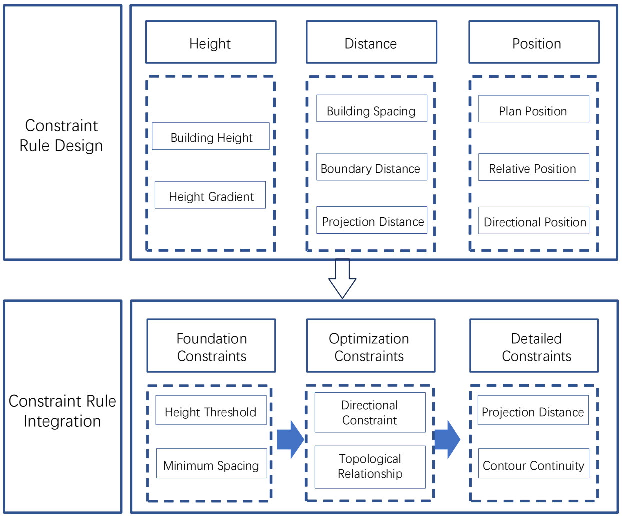

Figure 1. Spatial Constraint Framework

2.3 Building Individualization Extraction Under Spatial Constraints

Based on spatial visual saliency methods, this paper integrates height, position, and distance constraints to extract buildings, improving extraction accuracy and reducing misclassification.
Each pixel in the scene DOM contains corresponding RGB color information, while DSM grayscale values represent height features, where higher brightness indicates greater terrain elevation. Saliency detection on DSM ensures that height-related spatial information is considered in building extraction. The generated DOM retains clear geometric and texture features, with building color information distinctly different from surrounding objects. The edge orientation characteristics of buildings are also apparent, facilitating the extraction of color, brightness, and directional features in visual saliency analysis. The DSM, on the other hand, provides clear height features, which enhance spatial visual saliency detection.
First, DOM and DSM data are extracted from oblique photogrammetry. A color, brightness, and directional feature pyramid is constructed from DOM, while a height feature pyramid is created from DSM. Through multi-feature fusion, a

spatial visual saliency map is generated to preliminarily identify potential building regions. Spatial constraints are then applied:
Height constraints: Using dynamic thresholds (considering both absolute height and relative height differences) to filter out vegetation, roads, and other non-building objects.
Position constraints: Ensuring that building contours align with street orientations to match urban planning characteristics.
Distance constraints: Using nearest-neighbor or clustering analysis to distinguish between connected and independent buildings.
Finally, spatial constraint scores are weighted and fused with initial visual saliency scores. Canny edge detection and contour fitting techniques refine building boundaries, enabling accurate segmentation of individual buildings from the overall 3D model.

2.4 Algorithm Implementation and Workflow

This method is designed based on spatial constraints and integrates weighted rules with oblique photogrammetry data to achieve building footprint extraction. First, the raw data is preprocessed to extract Digital Orthophoto Maps (DOM) and Digital Surface Models (DSM), providing high-quality input for subsequent processing.
Next, a multi-scale feature pyramid is constructed, incorporating color, brightness, direction, and height features to generate an initial visual saliency map, which serves as a coarse locator for potential building regions. Based on the visual saliency map, spatial constraint rules are embedded to further refine the candidate building regions. These rules include dynamic height constraints, position constraints, and distance constraints, effectively filtering out non-building areas while enhancing the consistency of building boundaries.
To improve boundary precision, the method employs the Canny operator for edge detection, combined with vertex coordinate constraints to achieve refined building contour extraction. By extracting geometric edge information, the method enables accurate segmentation of individual buildings. Finally, the results of the initial segmentation are refined and optimized through an individualized extraction process, followed by accuracy analysis to validate the effectiveness and precision of the proposed method in building footprint extraction.

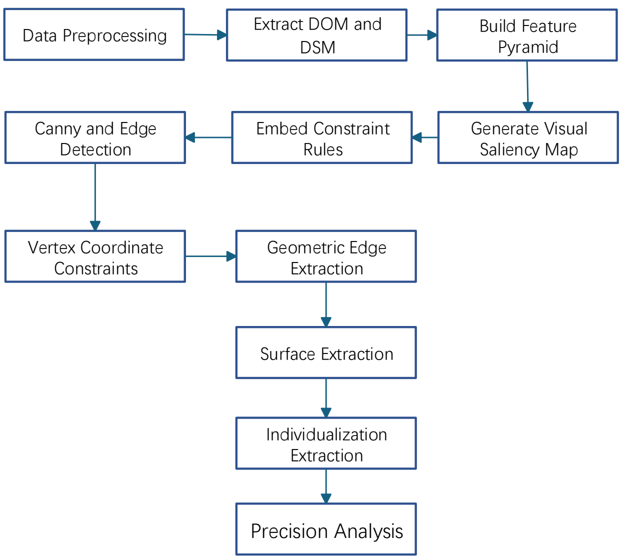

Figure 2. Technical Route

3. Experiment and Result Analysis

In this chapter, we conduct experimental validation of the proposed 3D building monolith extraction method based on spatial constraints. To evaluate the effectiveness of the method, we designed multiple groups of experiments covering various aspects, including building monolith extraction accuracy, computational efficiency, and processing stability. The experimental data is sourced from oblique photography models in OSGB format, with evaluation indicators including building monolith extraction accuracy, computational efficiency, and processing stability. The evaluation of each indicator is carried out by comparing the results before and after applying the spatial constraints, to verify the superiority of the method.

3.1 Experimental Data and Evaluation Indicators

The data used in the experiments comes from a high-resolution oblique photography model in OSGB format of the urban area of Taiyuan, Shanxi. The dataset covers an area of approximately 2.5 km² and includes about 512 buildings, representing different regions of the city, such as high-density commercial areas and low-density residential areas. The advantage of the OSGB format lies in its efficient tiling technology and optimized data structure, which allows us to efficiently load and render large-scale 3D data. This format provides faster loading speeds and better rendering performance than traditional data formats, especially when handling high-resolution data.
The complexity of the scene is as follows:
Building density distribution:
Commercial dense areas (density > 50 buildings/km²); Residential areas (density 20-50 buildings/km²); Suburban sparse areas (density < 20 buildings/km²); Occlusion ratio: Approximately 30% of the area experiences mutual building occlusion or vegetation coverage.
To validate the effectiveness of the proposed method, the following comparison baselines were set:
Baseline 1 (Traditional Geometric Segmentation): This method performs region growing based on the abrupt height changes in the DSM and texture features in the DOM, without applying any spatial constraint rules.
Baseline 2 (ID Marked Segmentation): This method uses manually labeled building IDs as the reference for accuracy validation.
Baseline 3 (U-Net Segmentation Model): This method adopts the classic U-Net architecture, trained using transfer learning on the same dataset. The input consists of multi-channel fused images from both the DOM and DSM.
In the experiment, the following evaluation metrics were used:
(1) **Building Monolith Extraction Accuracy:**

- **IoU (Intersection over Union):** This metric measures the accuracy of the extracted buildings. A higher IoU value indicates greater overlap between the extracted building monolith and the actual building, resulting in higher extraction accuracy.
- **F1-score:** The harmonic mean of precision and recall, providing a balanced evaluation of both false positives and false negatives.
- **Direction Error:** The angular deviation between the building's main axis direction and the true direction (measured in degrees).

(2) **Computational Efficiency:**
The computational efficiency is mainly evaluated by the time required to extract building monoliths and the consumption of system resources (such as memory usage and CPU utilization).

(3) **Processing Stability and Robustness:** This metric evaluates the robustness of the method across different scenarios (high-density buildings, low-density buildings, severely occluded areas, etc.) by analyzing the stability of the extraction results. The focus is on testing the algorithm's stability under noise and occlusion conditions, as well as how it handles different types of building shapes and densities.

3.2 Monolith Extraction Accuracy Analysis

In this experiment, we selected an urban area that includes different types of buildings, and performed extraction separately using models with and without spatial constraints.

Metric	**Traditional Geometric Segmentation**	**U-Net**	**Our Method**
IoU (%)	72.3–85.2	85.2–90.4	**91.5–95.4**
Directional Error (°)	5.3–12.4	3.0–6.5	**1.5–3.2**
Processing Time(s)	320	240	**218**
Occlusion Recall(%)	68.3	82.1	**89.4**
Noise False Detection Rate (%)	18.9	12.5	**6.3**

Table 1. Accuracy Analysis

Table 1 presents a comparison of the accuracy results of different methods in commercial dense areas, residential areas, and suburban sparse areas. The proposed method outperforms the traditional geometric segmentation and U-Net segmentation models in terms of IoU, direction error, and recall rate in occluded scenarios.

Specifically, in terms of IoU improvement:In commercial dense areas, the IoU of the proposed method reaches 91.5%, a 21.3% increase over the traditional method (72.3%), effectively eliminating interference from low-height structures (such as streetlights and billboards).

In suburban sparse areas, the IoU increases to 95.4%, indicating that the distance constraint (d_{th}=2.0m) can accurately distinguish independent buildings in low-density environments.

Regarding direction error optimization: In commercial dense areas, the direction error of the proposed method is 3.2°, a 74% reduction compared to the traditional method (12.4%), validating the boundary optimization effect of the direction constraint (D_{align} >0.97) for densely arranged buildings.

In terms of robustness in occluded scenarios: In areas with 30% occlusion, the recall rate of the proposed method reaches 89.4% (compared to 68.3% for the traditional method), mainly attributed to the collaborative effect of the geometric constraint and direction consistency judgment.

Figure 3. Building Contour Extraction

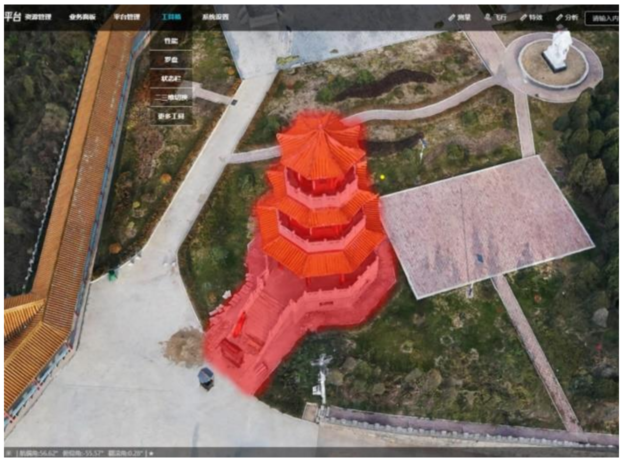

Figure 4. Building Monolith Extraction

3.3 Computational Efficiency Analysis

The testing environment for this experiment is as follows:
Hardware Environment: 13th Gen Intel(R) Core(TM) i9-13900K, 64GB RAM, NVIDIA RTX 4070. **Scene Division:** The scene was divided into 200m × 200m blocks, with each block's memory usage ≤ 2GB. **Processing Method:** Block data were processed using multithreading, with CUDA acceleration for Canny edge detection and gradient computation.

In this section of the experiment, we measured processing time, memory usage, and CPU utilization.

The performance results of different methods are summarized in the table below:

Method	**Processing Time (s)**	**Memory Usage (GB)**	**Parallel Speedup**
Traditional Geometric Segmentation	320	38	1.0×
U-Net Segmentation	240	42	1.3×
Proposed Method	218	28	3.2×

Table 2. Computational Efficiency Analysis

Table 2 compares the processing time, memory usage, and parallel speedup of different methods. The efficiency performance of the proposed method is as follows:

Processing Time Optimization: The processing time for a single scene in the proposed method is 218 seconds, which is a 32% reduction compared to the traditional method (320 seconds). This improvement is mainly due to the use of block processing techniques and dynamic constraint pruning strategies (which preemptively remove low-confidence regions where $S < 0.5$). Compared to U-Net (240 seconds), the proposed method is 9.2% more efficient, and it does not rely on labeled data.

Memory Usage: The peak memory usage is reduced to 28GB (compared to 38GB for the traditional method), achieved through OSGB block loading (200m × 200m) and parallel computing optimizations for memory management.

Parallel Speedup: The use of multithreading processing and CUDA acceleration (for Canny edge detection) results in a parallel speedup of 3.2×, significantly improving the efficiency of processing large-scale scenes.

3.4 Method Superiority Analysis

In this section, we further explore the advantages of the proposed spatial constraint method in terms of stability and robustness, particularly when dealing with interference factors such as noise and occlusion. In building extraction tasks, various sources of noise (e.g., trees, vehicles), occlusion phenomena (e.g., gaps between high-rise buildings), and complex building shapes often negatively impact extraction accuracy. Traditional methods tend to suffer from mis-segmentation or omission errors under such conditions.

In contrast, the proposed spatial constraint method effectively addresses these challenges. For example, in heavily occluded areas, the introduction of height constraints enables the algorithm to accurately identify and restore the contours of occluded buildings even when parts of them are hidden. In scenarios where low-rise and high-rise buildings are adjacent, the distance constraint helps distinguish different buildings and prevents the erroneous merging of neighboring structures into a single unit. Additionally, the position constraint, by analyzing the spatial distribution and orientation of buildings, significantly reduces segmentation errors caused by complex building shapes or urban planning variations. Even in special areas such as historical districts, the method can ensure accurate building extraction through adaptive weight and threshold adjustments.

Experimental results show that in environments with high noise levels and severe occlusion, the recall rate of the spatial constraint method improves by approximately 21% compared to traditional methods, while the mis-segmentation rate decreases by more than 20%. This fully demonstrates the robustness and stability of the proposed method in complex environments.

4. Conclusion and Future Work

This paper presents a method for extracting 3D building monomers based on spatial constraints, which integrates multi-dimensional spatial constraint strategies such as height, location, and distance to achieve accurate building extraction. The method not only effectively improves the precision of building extraction but also demonstrates high stability and robustness in dealing with dense urban scenes, occluded environments, and complex building forms. Experimental results show that, compared to traditional methods, this approach has significant advantages in terms of accuracy, computational efficiency, and stability in building monomer extraction. Particularly in high-density urban environments, the introduction of spatial constraint rules effectively reduces mis-segmentation and under-segmentation, making the extraction results more consistent with real urban structures. Moreover, the weighted fusion-based adaptive spatial constraint strategy allows the method to adapt to various urban environments, ensuring good extraction performance in modern high-rise building areas, low-density residential districts, and historical districts alike.

Future research could further expand and optimize the spatial constraint-based 3D building monomer extraction method proposed in this paper in several aspects. First, by combining semantic information such as building use, material characteristics, and plot attributes, semantic enhancement through deep learning models could improve segmentation accuracy and the method's adaptability. Secondly, introducing deep learning techniques, especially self-supervised learning and transfer learning, could allow the model to automatically adjust spatial constraint parameters in different urban environments, enhancing its adaptability. Furthermore, with the acceleration of urban digitalization, improving the computational efficiency of the algorithm to handle large-scale urban data is a key research direction. This can be optimized through parallel computing, GPU acceleration, and incremental updating strategies. In terms of data, integrating multi-source data such as LiDAR point clouds, remote sensing images, and BIM data would further improve the accuracy and robustness of building boundary extraction. Lastly, as cities dynamically change, the research could extend to dynamic urban modeling and intelligent analysis, using time-series data to detect changes in building forms, enhancing urban management intelligence, and promoting the widespread application of this method in smart city construction, urban planning, and building information modeling.

References

Du, X., Cui, H., Yan, Y., & Hu, H. (2023). Research on cesium monomerization algorithm for 3D model from oblique photogrammetry. *2023 5th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)*, Shanghai, China, 1170-1174.

Frommholz, D., Linkiewicz, M.M., & Poznańska, A. (2016). Inlining 3D reconstruction, multi-source texture mapping and semantic analysis using oblique aerial imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3, 605-612. doi.org/10.5194/isprs-archives-XLI-B3-605-2016.

Ge, L., Wu, F., Zhu, R., & Huang, D. (2016). A method for constructing integrated architectural areas based on spatial constraints. *Geomatics Science and Technology*, 41(3), 117-121. https://doi.org/10.16251/j.cnki.1009-2307.2016.03.024.

Lei, J., Liu, Q., Pan, C., et al. (2021). Research on the monomer technology of oblique photography three-dimensional model based on vector cutting. *Science of Surveying and Mapping*, 46(7), 84-91.

Liu, A., Wang, Z., Nie, W., Wu, X., Su, Y., & Yang, Z. (2015). Spatial context constrained characteristic view extraction for 3D model retrieval. In *Proceedings of the Third International Conference on Communications, Signal Processing, and Systems* (pp. 456-469). Springer. https://doi.org/10.1007/978-3-319-08991-1_72.

Lyu, F., Tang, S., He, G., & Meng, J. (2022). Extraction and individualization of airborne LiDAR building point clouds using density-based spatial clustering algorithm. *Sci. Technol. Eng.*, 22(9), 3446-3452.

Muntean, I., Hagau, A., & Toma, D. (2020). Pediatric cardiac transplantation: Indications and preoperative evaluation (Original title in Romanian). Romanian Journal of Pediatrics, 69(1), 53-58. https://doi.org/10.37897/RJP.2020.1.10.

Ni, X., Liang, J., Wu, L., & Liu, M. (2022). A method for 3D model individualization based on CityGML. *Bulletin of Surveying and Mapping*, (12), 29-34. https://doi.org/10.13474/j.cnki.11-2246.2022.0353.

Ruf, B., Thiel, L., & Weinmann, M. (2018). Deep cross-domain building extraction for selective depth estimation from oblique aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 142, 45-56. https://doi.org/10.1016/j.isprsjprs.2018.05.012.

Wang, Y., Hao, X., & Li, Y. (2018). Research on methods of 3D model individualization based on oblique photography. *Computer Engineering and Applications*, 54(3), 178-183.

Wei, T., Wang, X., Wu, J., & Zhu, S. (2023). Interval type-2 possibilistic fuzzy clustering noisy image segmentation algorithm with adaptive spatial constraints and local feature weighting & clustering weighting. Int. J. Approx. Reason., 157, 1-32.

Xu, X. Q., Lu, J. S., Zhang, N., Yang, T. C., He, J. Y., Yao, X., Cheng, T., Zhu, Y., Cao, W. X., & Tian, Y. C. (2019). Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 185–196. https://doi.org/10.1016/j.isprsjprs.2019.02.013.