Computers & Graphics 122 (2024) 103982

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on CEIG 2024

3D Modeling of rural environments from multiscale aerial imagery

David Jurado-Rodríguez *, Pablo Latorre-Hortelano, Luís René-Dominguez, Lidia M. Ortega

Department of Computer Science, University of Jaén., Jaén, Spain

ARTICLE INFO

Keywords:
3d modeling
Environmental preservation
Multiscale data sources
Semantic segmentation

ABSTRACT

Given the increasing attention to environmental preservation and sustainable development, the digitization of rural landscapes stands out as a pivotal strategy for effective environmental management and sustainability, land use planning, and preservation of cultural heritage. This work proposes a novel methodology for generating 3D models of rural landscapes by integrating multiscale data sources. Although Unmanned Aerial Vehicles (UAV) simplify the acquisition of multi-source data, their coverage is typically restricted to small landscapes due to their limited range and flight time. On the other hand, although the use of aerial images provides a broader view of the terrain, it is important to note that the low resolution of these images interferes with the task of accurate 3D modeling. Given these challenges, we propose a methodology that combines UAV data and high-resolution aerial imagery provided by the Spanish National Orthophoto Program (PNOA). This multi-source data integration is crucial to generating detailed and accurate 3D models of rural environments. The proposed methodology involves three steps: (1) semantic segmentation of aerial images identifying features such as vegetation, ground, and human-made structures, (2) estimation of the Digital Elevation Model (DEM), and (3) 3D modeling of rural environments using the point clouds generated from UAV images. The conducted experiments demonstrate the effectiveness of our approach identifying and representing previously mentioned features. Thus, this work presents advances in 3D representation techniques for real scenarios, contributing to the coordination of land utilization and environmental sustainability in rural landscapes.

1. Introduction

In current times, the digitization and 3D modeling of rural environments have become essential for environmental management. The ability to design detailed and precise 3D representations of rural environments facilitates the process of analyzing and evaluating the state of natural and cultural heritage, simulation of natural phenomena, biodiversity conservation, and the promotion of sustainable ecosystems.

However, the 3D modeling of natural ecosystems faces significant challenges. One of the main obstacles lies in the limited availability of detailed and complete 3D geospatial data, especially in extensive and remote areas. Although the use of UAV flights has enabled the collection of multisensory data, their capacity is restricted to specific areas, making it impossible to cover large land areas. This limitation hinders the creation of 3D models on a regional or national scale, which is crucial for government-level planning and decision-making.

In this context, the capture of high-resolution multiscale aerial images emerges as a promising solution. Entities such as the National Aerial Orthophotography Plan (PNOA), the European Earth Observation Open Science Data Hub (EEOSDA), or the Copernicus program represent valuable alternatives to complement the information acquired through drones. These images, captured from aerial or satellite platforms, offer a spatial and temporal resolution that encompasses large land areas, allowing for detailed and updated data on natural environments. PNOA provides high-resolution aerial images at the national level. EEOSDA provides open access to Earth observation data at the European level, including climate information and atmospheric measurements. The Copernicus program, using a constellation of satellites, provides satellite information, including optical and radar images.

Additionally, these entities provide Digital Elevation Model (DEM) with an altimetry precision of approximately 1 meter in urban areas and up to 2 meters in rural areas. These digital models are generated from LiDAR (Light Detection and Ranging) systems, which provide elevation measurements over large surfaces. The accuracy of DEMs may be affected by the quality of LiDAR data and the atmospheric conditions during data acquisition. Therefore, it is important to note that the resolution of these models is sometimes not sufficient to accurately model specific areas.

In this work, we propose a methodology for modeling complex and detailed 3D rural environments by merging multiscale data. Our proposal combines information acquired through UAVs with high-resolution aerial images provided by PNOA. The proposed methodology is divided into three stages taking as input data multiscale information: (1) semantic segmentation of PNOA images identifying features such as vegetation, ground, and human-made structures, (2) increasing the

* Corresponding author.
E-mail address: drodrigu@ujaen.es (D. Jurado-Rodríguez).

https://doi.org/10.1016/j.cag.2024.103982
Received 11 April 2024; Received in revised form 10 June 2024; Accepted 18 June 2024
Available online 22 June 2024

resolution of the Digital Terrain Model using point cloud densification algorithms, and (3) data fusion and 3D modeling of the rural environment.

The main contribution of this work is to propose a novel methodology for modeling rural environments in 3D through the fusion of multiscale data from different sources. Our method provides an effective tool for understanding and managing rural environments and contributing to the development of sustainable ecosystems. We also added a validation process to certify that high-resolution imagery can be sufficient to obtain a detailed 3D model of rural landscapes while providing advanced knowledge of the environment.

The rest of this document is structured as follows: Section 2 presents the current state-of-the-art, reviewing relevant technologies and methodologies related to the digitalization of rural environments using multiscale information. Then, Section 3 outlines the proposed methodology, whereas the obtained results and the experiments conducted to validate our proposal are presented in Section 4. Finally, the main contributions of this work are summarized in Section 5, including insights toward future work that aid in further enhancing the proposed methodology.

2. Previous work

The digitization of rural areas has aroused great interest in the scientific community for the realistic simulation of real-world scenarios, allowing users to explore and analyze these environments in virtual space. Remote sensing techniques make possible the data acquisition of natural environments from a distance using different sensors mounted on aircrafts, or UAVs (Unmanned Aerial Vehicles) [1]. These sensors collect a set of images (RGB, multispectral, hyperspectral, thermal) or point clouds (Light Detection and Ranging, LiDAR) [2].

Each specific sensor provides data of a different nature, spectral range (visible or non-visible), resolution (high or low), or periodicity (single or periodic). Then, depending on the final objective or type of data analysis, one of these data sources is selected. From all of them, RGB images are probably the most versatile and available data source. A collection of 2D images can be used for the 3D reconstruction of natural environments by using the Structure-from-Motion (SfM) technique [3]. The resulting models can be used for land cover classification, the study of vegetation density and distribution.

The quality of the final 3D model mainly depends on the device's sensor resolution. Drone-attached sensors provide higher-density point clouds or quality images over smaller areas, while satellite information is of lower resolution but covers larger areas of land without the need for human intervention in the acquisition process. An important advance in terms of remote sensing acquisition is the possibility of merging multi-source and multiscale data [4–6]. This benefits from the automatic procurement mechanism and the periodicity of satellite information [4,5]. To achieve this fusion of data from multiple sources and scales, several issues must be addressed, most notably the densification of areas captured with lower resolution sensors [7] or to fill in holes in the terrain [8]. Low resolution of DEM is solved by different techniques such as densifying the associated TIN (Triangulated Irregular Network) [9–11].

However, achieving high accuracy in digitizing these environments is not the only goal when scene understanding is required. Image segmentation is a previous step to classification in order to select the elements in the scene and pre-training datasets [12]. In rural environments, segmentation of individual species [13–16] and its classification [17] are often necessary processes that precede further analysis. Discarding human-made objects [18] and the rest of the surrounding elements (ground, roads, buildings, etc.) is also important in precision agriculture to focus only on vegetation data acquisition [19], and avoiding distortions associated with the spectral ranges. Many of these methods are addressed through machine learning approaches, particularly neural networks due to their ability to learn complex patterns from the point cloud data [18]. Convolutional Neural Networks (CNNs) are commonly used for the segmentation and classification of vegetation species [20–22].

The result of the 3D reconstruction of natural environments should be as faithful as possible to the original natural environment. Any of the several phases developed during this process may introduce noise. The remote sensing capture, the fusion of multiscale data, the segmentation and classification methods as well as the additional techniques for 3D modeling are susceptible to error accumulation. To demonstrate the level of accuracy of the resulting 3D models regarding reality, it is important to provide validation tests using Root Mean Squared Error (RMSE) of singular parameters as high or diameter of canopy or displacement error when fusing multi-source input data [23,24].

This paper faces the challenge of obtaining a reliable model of a rural environment by merging information from different sources and scales, in particular, PNOA images and a DEM. The resulting model is validated with a point cloud generated from high-resolution UAVs images. To our knowledge, this approach has not been previously explored in literature. The results show remarkable accuracy compared to the actual data.

3. Methodology

In this section, we present the workflow carried out for generating 3D virtual models of rural environments. As shown in Fig. 1, our work is divided into three steps taking as input data multiscale information acquired through UAVs and PNOA. The first step is based on the semantic segmentation of the PNOA images, in which features such as vegetation, ground, and human-made structures are detected. In this step, on the one hand, computer vision algorithms based on image contour detection were developed to detect vegetation. On the other hand, the Segment Anything Model (SAM) [12] was employed as a semi-automatic tool to identify human-made structures. The second step relies on increasing the resolution of the Digital Terrain Model by applying 3D point cloud densification algorithms. During the third step, the 3D modeling process is carried out, from which a dense 3D point cloud is generated covering the entire rural ecosystem. Finally, a validation phase is presented, focusing on verifying the accuracy and reliability of the generated 3D models compared with the 3D point cloud obtained from the UAVs. This dataset is rasterized to serve as ground-truth data. During the rasterization process, the point cloud data is converted into a raster or grid format, where each cell in the grid represents a specific area on the ground. This allows for easier comparison and analysis of the 3D model against the actual terrain features captured by the UAVs.

The software we use to merge data from various sources is a custom solution we have developed over time for different projects. It is built in C++ and employs specialized libraries like PCL (Point Cloud Library) and OpenCV. The only part of our process done separately is the segmentation of structures, for which we use Meta's Segment Anything model in Python.

3.1. Data acquisition

The growing use of high-resolution aerial images provided by entities such as the National Aerial Orthophotography Plan (PNOA), the European Earth Observation Open Science Data Hub (EEOSDA), and the Copernicus program has contributed to the digitization of both rural and urban environments. In this work, we have used images provided by PNOA. These images are captured from planes equipped with high-resolution cameras and subsequently processed to correct distortions and ensure high precision in representing the Earth's surface. This high resolution allows for detailed identification of key features of the rural environment, such as water bodies, wooded areas, and crops, which is essential for the planning and sustainable management of natural resources. It is important to note that by using georeferenced PNOA

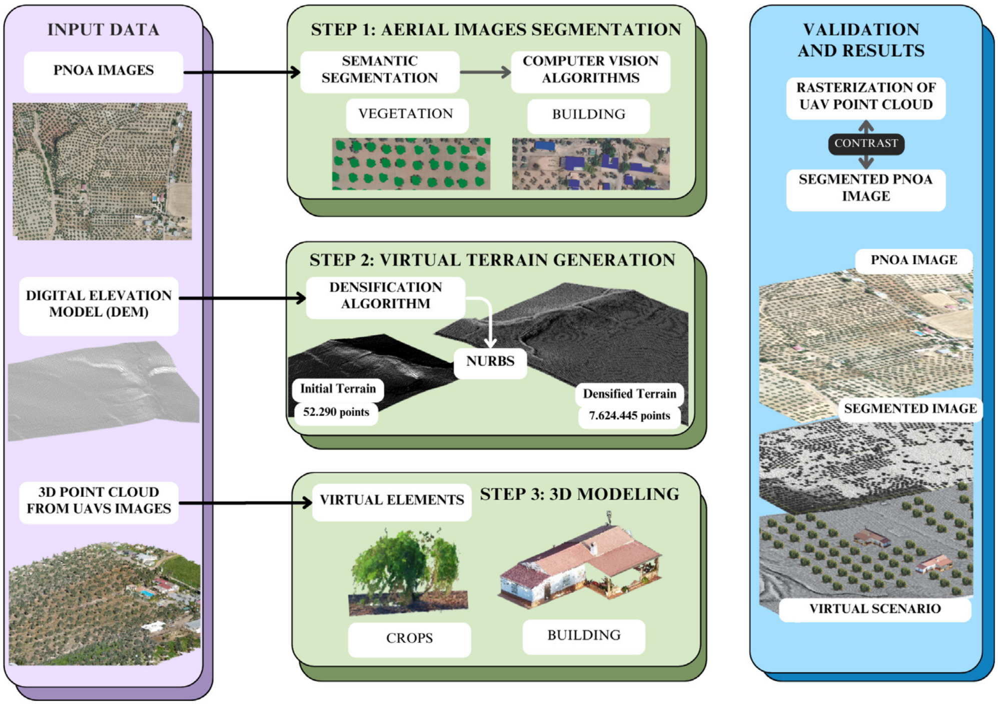

Fig. 1. An overview of the proposed workflow that summarizes the main steps of the proposed solution combines an AI-based technique for fast-forward image labeling and the generation of 3D rural scenarios.

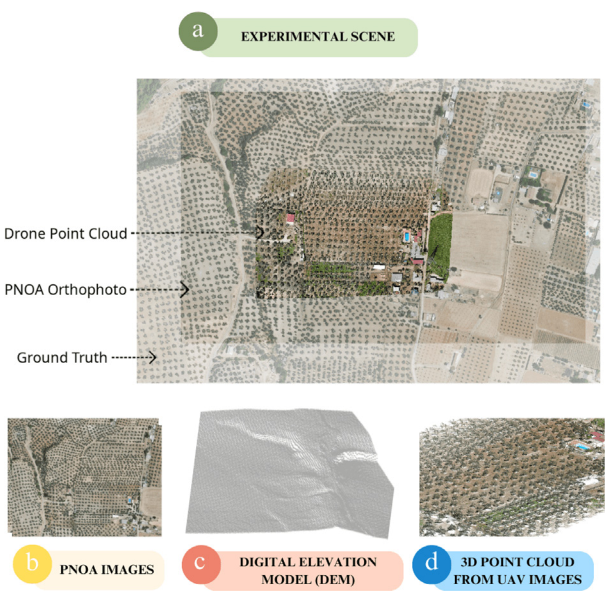

Fig. 2. Input data of our proposed methodology to generate a 3D model of rural environments using multiscale information.

images for the generation of the virtual environment, we ensure that the generation of 3D models are georeferenced, which facilitates the integration and analysis of the data in Geographic Information Systems (GIS) platforms.

On the other hand, in this work, we have included the use of Digital Elevation Models (DEMs) to provide detailed information on terrain elevation with an altimetric precision of approximately 1 meter in urban areas and up to 2 meters in rural environments. The precision of these models is based on the emission of laser pulses from planes or satellites and the measurement of the time they take to bounce off the Earth's surface.

Regarding the use of drones, it is important to highlight their ability to capture multiview and multisensory images, providing a new perspective on land distribution and use, as well as ecosystem health. For validation purposes, we have obtained a dense 3D point cloud generated after a photogrammetry process reflecting the current state of a plot of land of approximately 2 hectares. For this purpose, it was necessary to plan a flight of around 30 min at a height of 25 m, and a speed of 4 m/s, capturing a total of 900 RGB images with a resolution of 5472x3648. RGB imagery was acquired using a DJI Phantom 4 Pro quadcopter (DJI TECHNOLOGY CO., LTD, Shenzhen, China), which is paired with a three-axis electronic gimbal for camera stabilization. This system is equipped with an onboard 1-inch CMOS camera (model FC6310S) with a resolution of 20 MP.

The combined use of these technologies contributes significantly to more efficient and sustainable management of rural ecosystems. However, there are still challenges to address regarding the fusion of all this information from various sources (drones, aerial images, satellite images). For this reason, this work proposes a novel methodology for multiscale data fusion, from which large rural areas are modeled at a high level of detail. In Fig. 2, a schematic visualization of the information used by our method to generate virtual models of rural environments is presented.

3.2. Aerial images segmentation

To achieve a 3D model that faithfully represents the reality of the environment, it is crucial to categorize the elements of the scene. To do so, we have implemented multiple semantic image segmentation algorithms that categorize each pixel into vegetation, terrain or building. In this particular study, we focus on the detection of vegetation and human-built structures. For the identification of vegetation in PNOA images, we have developed a computer vision algorithm based on contour extraction. In addition, to identify human-built structures, we have employed the Semi-Automatic Model (SAM) [12], based on Artificial Intelligence (AI).

In this section, we describe the techniques and processes used in each of the steps of the algorithms, as well as the results obtained in the detection and classification of vegetation and human-made structures.

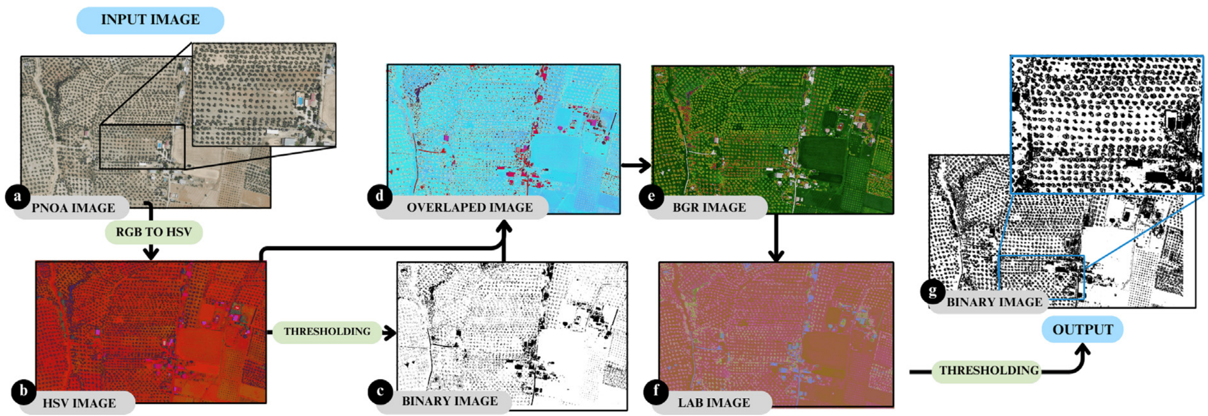

Fig. 3. The main steps of our automatic method to identify vegetation in PNOA images. As a result, the algorithm provides a binary image in which the image pixels belonging to vegetation are identified.

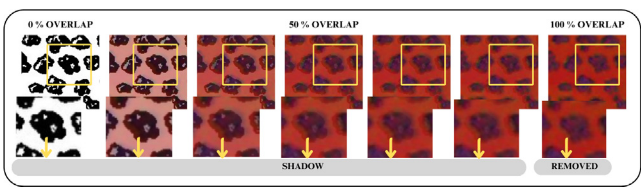

Fig. 4. Demonstration of the effectiveness of shadow removal. Images show varying percentages of overlap between binary image and HSV image.

3.2.1. Vegetation identification

To carry out the vegetation identification, we have implemented a computer vision algorithm based on detection, filtering, and contour extraction in images. The algorithm focuses on identifying each shaped olive grove crop in the image by extracting its centroids at the pixel level.

Fig. 3 displays the main steps of the implemented algorithm. It begins by applying a Gaussian filter to the PNOA image (Fig. 3(a)) using a 5x5 kernel size. This step reduces the noise inherent in the image, which improves the quality of detection. Subsequently, the image is converted to the HSV (Hue Saturation Value) color space (Fig. 3(b)). In the HSV color space, the vegetation tends to better show several shades of green with common intensities and brightness levels, making it distinguishable from other elements.

The next step is to apply image thresholding and obtain a binary image, as in Fig. 3(c), upon which morphological operations of dilation and erosion are applied. These operations facilitate the extraction of contours related to vegetation in the image. Subsequently, the binary image is overlapped over the HSV input image to identify and remove the shadow cast by vegetation (Fig. 3(d)). This process improves the accuracy of vegetation detection by reducing the impact of shadows (see Fig. 4 to visualize how our method reduces the impact of shadows), resulting in a clearer and more accurate representation of vegetation zones (see Fig. 3(e)). Finally, an additional color space conversion is required to differentiate vegetation from terrain. This process involves converting the image to the LAB color space as Fig. 3(f) shows. In the LAB color space, each pixel is characterized by three components: luminance (L) provides information on the brightness levels of vegetation, while the chromaticity components (A and B) contribute to the understanding of variations in shades of green present in vegetation. Leveraging these components, our method robustly separates vegetation from terrain. This process results in a binary image in which vegetation is identified in contrast to any other element in the image.

The final critical step in the vegetation identification process involves applying the Canny algorithm to extract contours from the binary image above generated (Fig. 3(g)). This step helps to extract image contours by identifying sharp changes in the gradient intensity of each pixel as shown in Fig. 5(b). Once the contours have been extracted, it is imperative to apply a contour closure algorithm. This operation involves filling and closing contours previously identified by connecting neighboring contour segments (Fig. 5(c)). Furthermore, in addition to contour closure, a filtering process is carried out to select only the contours corresponding to the vegetation (see Fig. 5(d)). This filtering is achieved by applying selection criteria based on contour area and shape. Consequently, contours not related to vegetation, such as those representing buildings, roads or other landscape elements, are discarded.

As a result, our algorithm obtains the central points of each contour. These centroids are crucial for accurately identifying and characterizing the location and shape of vegetation in the image. Besides providing information on vegetation's spatial distribution, centroids are utilized to calculate additional metrics, such as vegetation density and distribution in the study area.

3.2.2. Structures identification

Unlike the fully automated process for identifying vegetation in images, the identification of architectural features in rural environments presents significant challenges due to the diversity and complexity of the structures involved, such as houses, towers, water pools, parking lots, and other buildings. This heterogeneity complicates the application of a fully automated approach.

Therefore, our approach is based on a semi-automatic method that requires manual intervention by the user to initially mark the areas where architectural elements are believed to exist. This human intervention provides crucial information and allows for greater accuracy in identifying structures, taking into account the complexity of the rural environment and the variability of the elements present.

Fig. 6(a) displays the full PNOA image upon which we begin the segmentation process. Fig. 6 (a_1) demonstrates the functionality of the AI model used (Segment Anything Model). This model takes as input a pixel from the image and produces as output common points, generating the entirety of the structure as seen in Fig. 6 (a_2). Finally, Fig. 6(b) shows the final result of the PNOA image segmentation using SAM.

Once the user has manually marked these areas of interest, our method leverages the capabilities of SAM (Spatially Aware Machines), a pre-trained AI model that can accurately identify and isolate distinct objects or areas of interest in an image. SAM stands out in performing meticulous analysis of visual features in images through pattern recognition and classification techniques based on geometric and textural features, using advanced machine learning algorithms and convolutional neural networks. SAM can effectively detect various human structures, such as buildings, agricultural facilities, and roads, in PNOA images.

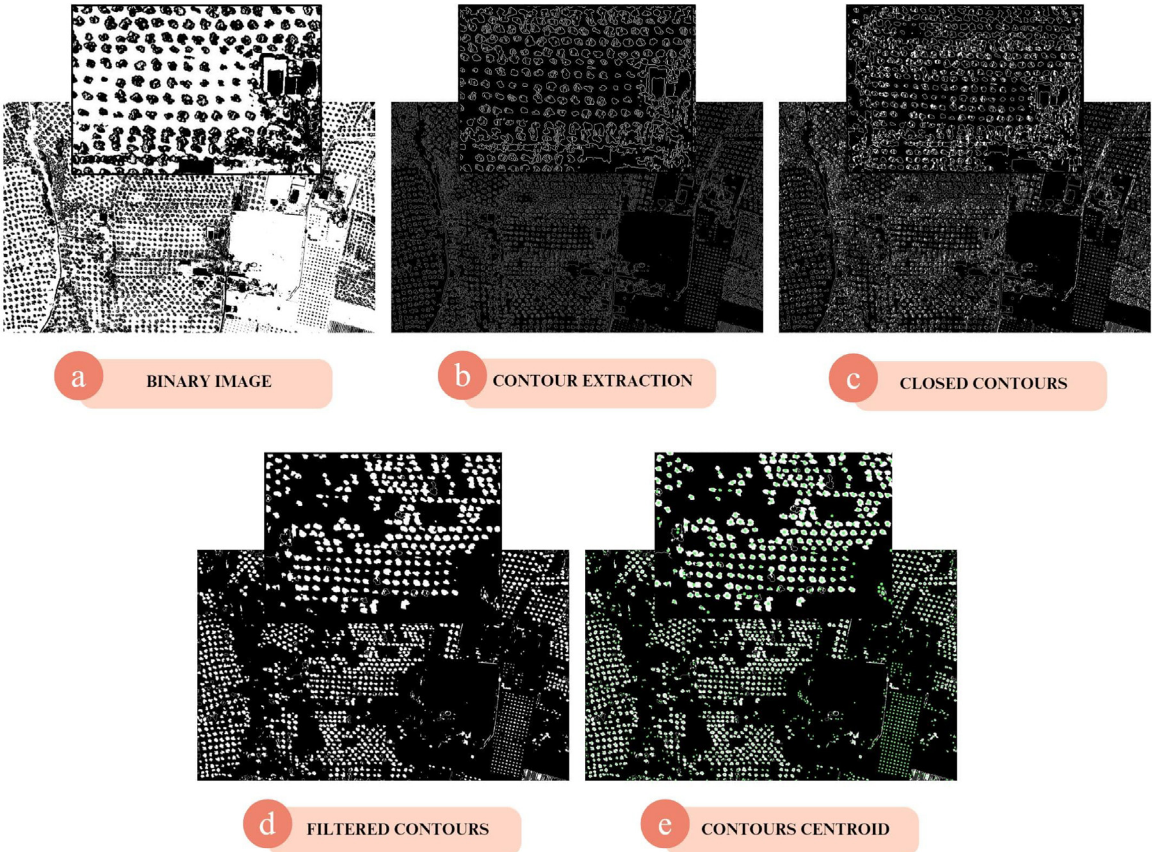

Fig. 5. Development flow of the procedure to obtain the centroid of olive trees in segmented vegetation images.

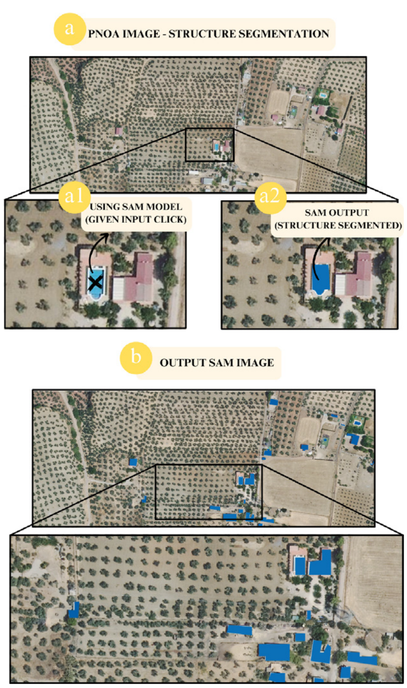

Fig. 6. Segmentation method to identify architectural elements in PNOA images. The input PNOA image (a) and the final segmented results (b).

This hybrid approach, which combines human expertise in the initial identification of areas of interest with the power of SAM's machine learning algorithms, offers an effective and robust solution for detecting architectural elements in PNOA images. The result is precise segmentation of structures in rural environments, essential for various applications such as urban planning, land management, and natural hazard assessment.

3.3. Terrain generation

Regarding terrain modeling, our method uses as input data the Digital Elevation Model (DEM) provided by PNOA. The DEM offers essential information about the region's topography, allowing for a general understanding of the terrain elevation in terms of spatial coordinates. However, it is crucial to note that when modeling small land areas (less than 5 hectares), the accuracy of the DEM is not sufficient to model the terrain elevation accurately. Therefore, it is necessary to perform densification of the DEM to obtain a more detailed and accurate representation, which guarantees the generation of an environment close to reality. In this section we describe the process of densifying the DEM using advanced interpolation techniques, such as Non-Uniform Rational B-Spline (NURBS). This approach improves the quality and resolution of the resulting terrain model, which is crucial for applications that require an accurate representation of the geographical environment.

In order to enhance terrain modeling from DEMs, the initial step involves reading and processing the DEM information using the GDAL library. This operation allows us to extract the normalized heights for each pixel in terms of X and Y coordinates, considering a resolution of 2 meters per pixel. Once these heights are obtained, representative points are instantiated for each pixel coordinate, resulting in an initial 3D point cloud providing a general overview of the region's topography. However, as the DEM may not be sufficiently detailed, advanced interpolation techniques such as Non-Uniform Rational B-Spline (NURBS) are employed to obtain a densified 3D point cloud.

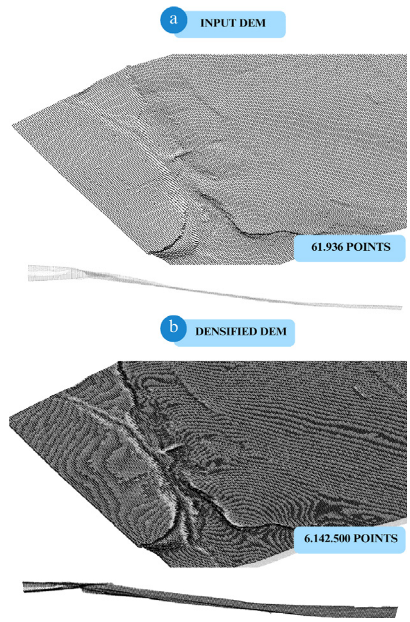

Fig. 7. The implemented densification method by applying the Non-Uniform Rational B-Splines algorithm (NURBS) to initial DEM.

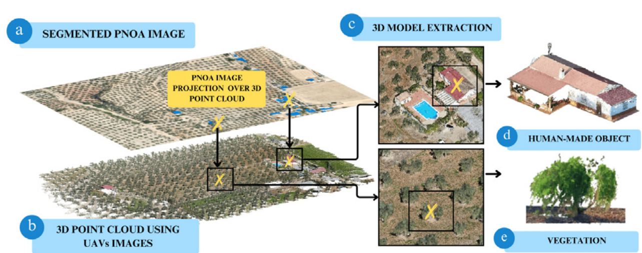

Fig. 8. The proposed methodology to extract 3D models from point clouds using the PNOA image segmented.

These methods enable the addition of extra points between existing ones, filling the gaps without compromising the essential features of the original cloud. Fig. 7(b) shows the resulting 3D point cloud contains a total of 6.142.500 points, compared to the initial cloud using the DEM with 61.936 points (Fig. 7(a)). This significant increase in point density enhances the accuracy and fidelity of the terrain model.

3.4. 3D modeling

The 3D modeling stage is the last step of the proposed method. In this section, the integration of multiscale data into a virtual environment is performed. This task is based on the extraction of 3D models for each segmented entity in PNOA image. To achieve this goal, it is necessary to project the segmented PNOA image over the reconstructed 3D point cloud generated by high-definition UAV images. This process allows us to extract partial 3D point clouds corresponding to the segmented entities. Fig. 8 displays the implemented method to extract the models used to represent real 3D rural environments.

The ability to extract partial point clouds from the segmented entities in PNOA images enables us to generate a library of precise and compact entities (see Fig. 9). Once the models have been extracted, the next step is to estimate their position on the previously densified Digital Terrain Model (DEM) in the previous section. To do this, it is necessary to project the segmented PNOA image onto the terrain. By knowing the centroids of each identified class in the image, we only need to project those points and estimate their 3D position on the terrain.

Fig. 9. The 3D models extracted and used to 3D rural environments modeling.

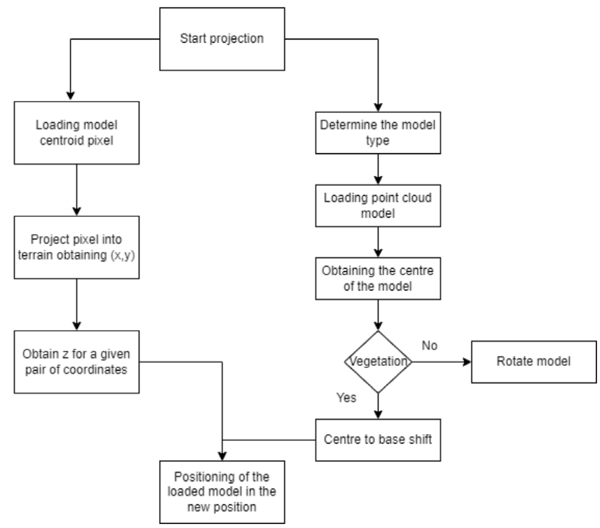

Fig. 10. Workflow for the projection of an element on the ground. The left side represents the obtaining of the central positioning coordinate, while the right side represents the transformations the model undergoes before being instantiated.

Considering (C_x, C_y) as the centroid of a segmented entity, $(width, height)$ the dimensions of the PNOA image, and (Max_x, Max_y) as the surface terrain, Eq. (1) shows the operation to transform a pixel from the PNOA image into normalized world coordinates (C'_x, C'_y). The last step of the projection is to obtain the height or z-coordinate. To obtain it we must access the height of the surface in the pair of coordinates (C'_x, C'_y) previously defined. This information is accessible thanks to the DEM that we processed in previous steps. This process (see the left side in caption of Fig. 10) ensures accurate projection of segmented entities onto the terrain, allowing for a faithful representation of the geographic environment in the resulting 3D model.

$$C'_x = \frac{C_x \cdot Max_x}{width} \qquad C'_y = \frac{C_y \cdot Max_y}{height} \tag{1}$$

Once the location on the terrain where the 3D model will be instantiated has been identified, the last step of our method consists of moving the previously loaded and processed model to this point. The preparation of the model (see the right side in caption of Fig. 10) consists of a series of processes that start by selecting from among all the models extracted from the point cloud the one that best fits.

It is essential to take into account the semantic entity (vegetation or artificial structures) assigned during the segmentation process. Subsequently, the most appropriate model is chosen from among all extracted models (large structures, olive trees, small houses and pools) to best fit both in shape and size to the entity identified in the image. By carefully selecting the models for each segmented entity, we ensure an accurate and realistic representation of our virtual environment, significantly improving user immersion and the quality of their experience. In addition, to achieve greater fidelity to reality, we also take into account the rotation of the model, ensuring that it aligns correctly with the orientation of the entity in the real world.

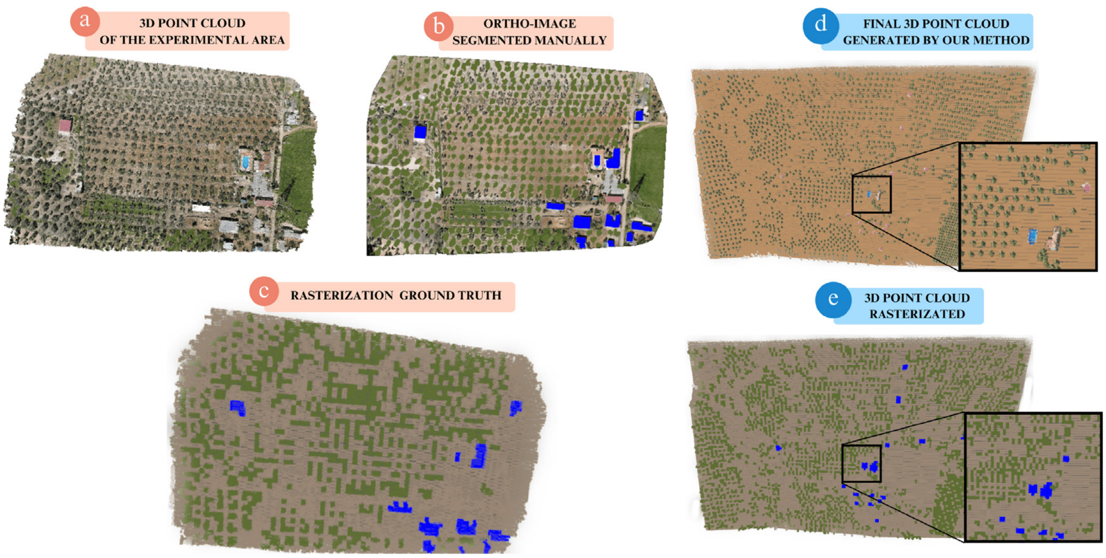

Fig. 11. The validation of the final 3D point cloud. (a) The UAV point cloud, (b) a semantic segmentation of the aerial image of the zone. On the top right (d) the procedurally generated cloud. Finally, below (c)(e) the regular classified rasterizations where standard colors are assigned by class (vegetation - green, terrain - brown, structures - blue).

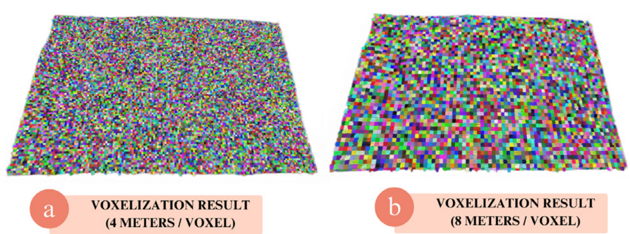

Fig. 12. 3D point cloud voxelization taking into account different voxel sizes.

Table 1
Validation of the generated procedural terrain by pixel accuracy.

Class	Ground truth	Our method	Accuracy
Vegetation	965	806	83.52%
Terrain	2386	2156	90.36%
Structure	70	66	94,28%

4. Results and validation

In this section, we describe the experiments performed to validate our proposed methodology for 3D rural scenario modeling. The accuracy, performance, and robustness of our method were tested over a rural area of about 2 hectares. These results were obtained using a CPU (Intel® CoreTM i7-10510U 2.30 GHz) with 8 GB RAM and Ubuntu 20.04.1 as the operating system.

The validation process is carried out by comparing two rasterized and classified images. One corresponds to the 3D point cloud generated using UAV imagery (Fig. 11(c)), and the other to the resulting point cloud using our method proposed methodology (Fig. 11(e)). It is important to note that, for validation, a specific subarea of the entire global region covered by PNOA has been selected.

To carry out the rasterization process, it is crucial to voxelize the point cloud and set a dominant color per voxel. Fig. 12 shows different voxel size configurations. In this case, we have selected a voxel size of 4 meters to ensure that each olive tree is represented by a single voxel, rather than multiple voxels. This approach simplifies the process, allowing for one dominant color per voxel, which aids in the identification of vegetation entities and artificial structures.

In order to obtain the ground truth, a manual classification of the 3D point cloud generated by UAV images is performed using the same segmentation algorithm employed in Section 3.2.2. In this way we obtain an accurate and robust image with which to validate our method.

Table 2
Validation of the generated procedural terrain by instance accuracy.

Class	Ground truth	Our method	Accuracy
Vegetation	481	441	93.14%
Structure	11	11	100%

Once both rasterizations are classified, we proceed with a comparison by pairs of pixels coinciding in coordinates. The results obtained after validating using this method can be seen in Table 1. In this table we can see the coincidence of pixels according to classes and for the total, since having the classified pixels we can determine for each class how many pixels we have in each cloud and in which positions both pixels are marked under the same class.

To consider other aspects of the validation, a comparison of the number of instances of each type that we have in the generated terrain and in the manually classified cloud has been made. As we can see in Table 2, in the case of the structures there is a complete accuracy in the segmentation of these since the process is the same while by pixels there was failure. This change is due to the standardization of the vegetation 3D model representing all with the same size which causes failures in the correspondence of total number of pixels for each instance of vegetation.

Regarding the vegetation, it can be observed how the automatic segmentation achieves a proportion of very accurate instances with 93.14% higher than the validation at pixel level. This may be since when working at pixel level as an integer there may be a slight displacement in the positioning of the olive tree which produces a small inaccuracy in the number of matching pixels.

The data presented in Table 3 illustrates the computational cost associated with each phase of our algorithm. Notably, the structure segmentation phase stands out with a time of 60 s, which is significantly higher than the other processes. This could be attributed to the complexity and the high-resolution details required for accurate structure identification. On the other hand, the terrain generation processes, including DEM generation and NURBS, are relatively efficient, with times of 3.27 s and 1.37 s, respectively.

The entity projection phase shows a marked difference in processing times between vegetation (15.7 s) and structures (0.19 s), highlighting that vegetation projection is more computationally intensive, likely due to the higher variability and complexity of vegetation data. The

Table 3
Computational cost of each process of the generated algorithm.

Process	Time (s)
Aerial segmentation	
Vegetation	1.99
Structures	60
Terrain generation	
Generate DEM	3.27
NURBS	1.37
Entity projection	
Vegetation	15.7
Structures	0.19
Validation process	
Rasterization	0.35
Comparison	13.88

validation process is split into rasterization and comparison, with the latter being the more time-consuming task at 13.88 s. This indicates that while rasterizing the data is quick, the detailed comparison of each pixel takes a significant amount of computational effort, which is essential for ensuring accuracy in the validation step. Overall, these insights into computational costs provide a clear understanding of where optimizations could be targeted in future iterations of the algorithm.

5. Conclusions and future works

In conclusion, this work introduces a novel methodology for modeling complex and detailed 3D rural environments by integrating multiscale data from various sources. By combining information obtained from UAVs with high-resolution aerial images provided by entities like PNOA, this methodology addresses the challenges of limited data availability and coverage in extensive and remote areas.

The proposed methodology consists of three stages: semantic segmentation of aerial images to identify key features, increasing the resolution of Digital Terrain Models using point cloud densification algorithms, data fusion and 3D modeling of the rural environment, and finally, the validation of the proposal method and the 3D model obtained. This approach offers an effective tool for analyzing and managing rural landscapes, thereby contributing to the promotion of sustainable ecosystems and informed decision-making in environmental management.

Through experiments and validation, the effectiveness of the proposed methodology has been demonstrated, highlighting its potential for creating detailed and accurate 3D representations of rural environments. However, it is important to acknowledge the limitations, such as the resolution constraints of Digital Terrain Models and potential inaccuracies in data acquisition.

Looking forward, we propose the implementation of new methodologies to enhance the resolution of PNOA images. Moreover, to evaluate the robustness of the method, we plan to extend the application of this methodology to a variety of geographic regions characterized by diverse vegetation types, seasonal variations, and a range of lighting conditions. Overall, this work lays a foundation for leveraging multiscale data integration in the digitization and 3D modeling of rural environments, contributing to broader efforts in sustainable ecosystem management and conservation.

CRediT authorship contribution statement

David Jurado-Rodríguez: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Pablo Latorre-Hortelano:** Software, Resources, Investigation, Funding acquisition, Data curation. **Luís René-Dominguez:** Software, Methodology, Investigation, Formal analysis, Data curation. **Lidia M. Ortega:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This project has been funded under the research projects with references PID2022-137938OA-I00, PID2021-126339OB-I00 and TED2021-132120B-I00. These projects are co-financed by the Junta de Andalucía, Ministerio de Ciencia e Innovación (Spain), and the European Union's ERDF funds.

References

[1] Wegmann M, Leutner B, Dech S, Wegmann M, Leutner B, Dech S. Remote sensing and GIS for ecologists: Using open source software. Exeter, UK.: Pelagic Publishing; 2016.

[2] Sabins F, Ellis J. Remote sensing: Principles, interpretation, and applications. fourth ed. Waveland Press; 2020, URL: https://books.google.es/books?id=rAnaDwAAQBAJ.

[3] Iglhaut J, Cabo C, Puliti S, Piermattei L, O'Connor J, Rosette J. Structure from motion photogrammetry in forestry: a review. Curr For Rep 2019;5(3):155–68. http://dx.doi.org/10.1007/s40725-019-00094-3, Type: Review.

[4] Bhatnagar S, Gill L, Regan S, Waldren S, Ghosh B. A nested drone-satellite approach to monitoring the ecological conditions of wetlands. ISPRS J Photogramm Remote Sens 2021;174:151–65. http://dx.doi.org/10.1016/j.isprsjprs.2021.01.012, Type: ARTICLE.

[5] Murugan D, Garg A, Ahmed T, Singh D. Fusion of drone and satellite data for precision agriculture monitoring. In: 11th international conference on industrial and information systems, ICIIS 2016 - conference proceedings, vol. 2018-January. 2016, p. 910–4. http://dx.doi.org/10.1109/ICIINFS.2016.8263068, Type: Conference paper.

[6] Argudo O, Comino M, Chica A, Andújar C, Lumbreras F. Segmentation of aerial images for plausible detail synthesis. Comput Graph 2018.

[7] Sajedizadeh S, Maghsoudi Y. PS-Insar point cloud densification using Sentinel-1 and TerraSAR-x data. Int J Remote Sens 2023;44(20):6375–98. http://dx.doi.org/10.1080/01431161.2023.2266121, Type: ARTICLE.

[8] Zhang C, Zhou H, Ju X, Duan J. A point cloud hole spiral-filling method based on 2D and 3D data fusion. Measurement: J Int Meas Confed 2023;223. http://dx.doi.org/10.1016/j.measurement.2023.113788, Type: ARTICLE.

[9] Ma H, Zhou W, Zhang L. DEM refinement by low vegetation removal based on the combination of full waveform data and progressive TIN densification. ISPRS J Photogramm Remote Sens 2018;146:260–71. http://dx.doi.org/10.1016/j.isprsjprs.2018.09.009, Type: ARTICLE.

[10] Deng S, Shi W. Integration of different filter algorithms for improving the ground surface extraction from airborne lidar data. In: International archives of the photogrammetry, remote sensing and spatial information sciences - ISPRS archives, vol. 40, 2013, p. 105–10, Issue: 2W1 Type: Conference paper.

[11] Tomková M, Potůčková M, Lysák J, Jančovič M, Holman L, Vilímek V. Improvements to airborne laser scanning data filtering in sandstone landscapes. Geomorphology 2022;414. http://dx.doi.org/10.1016/j.geomorph.2022.108377, Type: ARTICLE.

[12] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. 2023, _eprint: 2304.02643.

[13] Jurado JM, Cárdenas JL, Ogayar CJ, Ortega L, Feito FR. Semantic segmentation of natural materials on a point cloud using spatial and multispectral features. Sensors (Switzerland) 2020;20(8). http://dx.doi.org/10.3390/s20082244, Type: ARTICLE.

[14] Cardama FJ, Heras DB, Argüello F. Consensus techniques for unsupervised binary change detection using multi-scale segmentation detectors for land cover vegetation images. Remote Sens 2023;15(11). http://dx.doi.org/10.3390/rs15112889, URL: https://www.mdpi.com/2072-4292/15/11/2889.

[15] Lindberg E, Holmgren J. Individual tree crown methods for 3D data from remote sensing. Curr For Rep 2017;3(1):19–31. http://dx.doi.org/10.1007/s40725-017-0051-6, Type: Review.

[16] Argudo O, Chica A, Andujar C. Terrain Super-resolution through Aerial Imagery and Fully Convolutional Networks. Comput Graph Forum 2018;37(2):101–10. http://dx.doi.org/10.1111/cgf.13345, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13345, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13345.

[17] Xu K, Tian Q, Yang Y, Yue J, Tang S. How up-scaling of remote-sensing images affects land-cover classification by comparison with multiscale satellite images. Int J Remote Sens 2019;40(7):2784–810. http://dx.doi.org/10.1080/01431161.2018.1533656, Type: ARTICLE.

[18] Chehreh B, Moutinho A, Viegas C. Latest trends on tree classification and segmentation using UAV data—A review of agroforestry applications. Remote Sens 2023;15(9). http://dx.doi.org/10.3390/rs15092263, Type: Review.

[19] López Ruiz A. Prototipo de control avanzado de grandes plantaciones mediante teledetección. Universidad de Jaén; 2019, URL: http://crea.ujaen.es/jspui/handle/10953.1/14302, Accepted: 2021-06-15T09:52:45Z Publisher: Jaén: Universidad de Jaén.

[20] Deng L, Fu B, Wu Y, He H, Sun W, Jia M, et al. Comparison of 2D and 3D vegetation species mapping in three natural scenarios using UAV-LiDAR point clouds and improved deep learning methods. Int J Appl Earth Obs Geoinf 2023;125. http://dx.doi.org/10.1016/j.jag.2023.103588, Type: ARTICLE.

[21] Jayakumari R, Nidamanuri RR, Ramiya AM. Object-level classification of vegetable crops in 3D LiDAR point cloud using deep learning convolutional neural networks. Precis Agric 2021;22(5):1617–33. http://dx.doi.org/10.1007/s11119-021-09803-0, Type: ARTICLE.

[22] Hell M, Brandmeier M, Briechle S, Krzystek P. Classification of tree species and standing dead trees with lidar point clouds using two deep neural networks: PointCNN and 3DmFV-Net. PFG - J Photogramm Remote Sens Geoinf Sci 2022;90(2):103–21. http://dx.doi.org/10.1007/s41064-022-00200-4, Type: ARTICLE.

[23] Du L, Pang Y, Ni W, Liang X, Li Z, Suarez J, et al. Forest terrain and canopy height estimation using stereo images and spaceborne LiDAR data from GF-7 satellite. Geo-Spatial Inf Sci 2023. http://dx.doi.org/10.1080/10095020.2023.2249037, Type: ARTICLE.

[24] Murcia HF, Tilaguy S, Ouazaa S. Development of a low-cost system for 3D orchard mapping integrating UGV and LiDAR. Plants (Basel, Switzerland) 2021;10(12). http://dx.doi.org/10.3390/plants10122804.