Mem. S.A.It. Vol. 96, 153

Memorie della

ArbörIA

An open-data-based urban tree inventory tree solution improved by AI optimization

M. Angelinelli[1], C. Barrera[2], B. García-Pascual[3,4], and G. Pedrazzi[1]

[1] HPC Department - Cineca, Via Magnanelli, 6/3, 40033 Casalecchio di Reno, Bologna, Italy e-mail: m.angelinelli@cineca.it
[2] Föra, C Alonso de Velázquez, 2, portal 2, 42003, Soria, España
[3] Luke - Natural Resources Institute of Finland, Yliopistokatu 6, 80100, Joensuu, Finland
[4] UEF - University of Eastern Finland, Yliopistokatu 7, School of Forest Sciences, 80100 Joensuu, Finland

Received: 30-10-2024; Accepted: 13-11-2024

Abstract. Understanding the exact location and species of urban trees and forests is crucial for climate change mitigation and urban ecosystem management. To address this need, we propose an automated method to detect and locate trees in urban areas using LiDAR technology together with aerial RGB and infrared imagery. In addition, we have developed a semi-automatic approach to create a tree species classification dataset using Google Street View imagery. In future versions of this project, the integration of these methods will enable the geolocation and identification of urban trees with minimal human intervention.

Key words. Urban Forestry, LiDAR, geo-positioning, computer vision, clustering

1. Introduction

In recent years, two major trends have driven innovation in large-scale, low-cost urban tree inventories: the use of Convolutional Neural Networks (CNNs, Albawi et al. 2017) for abstract feature and object extraction in imagery, and the increasing availability of detailed, low-cost street-level imagery. Simultaneously, remote sensing technologies such as Light Detection and Ranging (LiDAR, Calvert 1990), aerial photography, and multispectral and hyperspectral imagery have become widely utilized in Earth Observation (EO) and large-scale analysis (e.g, Rapinel et al. 2015). The integration of these innovative remote sensing technologies with advanced computer vision algorithms enables the semiautomated identification of urban trees and the automatic extraction of their main metrics, providing a more time-efficient and cost-effective alternative to field inventories.

Among the data sources that have recently gained attention are street-level images from Google Street View (GSV, Anguelov et al. 2010), a geospatial platform with global coverage that offers geocoded, standardized street-level images in various formats and resolutions at a relatively low cost. In the context of smart urban forestry, GSV imagery combined with computer vision techniques is applied in three main areas: estimating the shade provided by trees, quantifying perceived urban canopy cover, and mapping the location of urban trees. These metrics serve as key indicators for assessing the sustainability of cities.

The proposed experiment builds on existing results, as the first version of the solution, named *ArbörIA*, a digital twin for urban trees that combines computer vision and machine learning to automate the inventory of street trees. The project was proposed by *Föra - Forest Technologies*[1], funded by the European project *EUHubs4Data*[2] and developed together with *CeADAR's strategic research and development service*[3] and *Cineca HPC data analytics group*[4]. By leveraging and integrating data from airborne LiDAR, aerial orthoimages, and GSV, *ArbörIA* develops advanced algorithms to provide high-resolution, reliable, and up-to-date geospatial information on urban tree assets.

This paper presents the methods and results developed within the *ArbörIA* project, an extension of the work presented by Rodríguez-Puerta et al. (2022), and it is structured as follows: in Sect. 2 we introduce the approaches and the technologies developed in the project; we present and discuss the main results in Sect. 3, while in Sect. 4 we summarise the project and possible future scenarios.

2. Methods

In this section, we will go through the two different tasks in which *ArbörIA* project is organised, explaining the technologies and methods, as well as the steps that we followed to obtain the results that we will discuss in the next section. The combination of these methods allows the automatic geolocalisation and classification of public green spaces in specific areas. Although we discuss the tasks separately and they could be used independently, it is useful to stress the fact that the combination of them represents the power and the main goal of the project.

2.1. Urban trees geo-positioning

The objective of this task is to detect and position trees within an urban environment. To achieve this, various data sources were utilised, each with specific characteristics and limitations tied to the quality of the remote sensors used. The primary sources of data included:

- Aerial orthophotographs, which are geo-referenced photographs, available through the National Plan of Aerial Orthophotographs (PNOA, Arozarena et al. 2005). The images have a resolution ranging from 25 to 50 cm, depending on the area, and are updated every 2 to 3 years. This data provides a detailed visual representation of the urban landscape.
- LiDAR technology is used to determine distances from a sensor to objects or surfaces using a pulsed laser beam. In this study, LiDAR devices mounted on drones or airplanes were employed to generate a point cloud consisting of X, Y, and Z coordinates for each point on the terrain. The point density varies from 0.5 to 4 points per square meter, with some areas having even higher densities. The data is freely accessible via the CNIG Download Center[5] in digital files covering 2x2 km areas.
- Images at street level, obtained through sources such as Google Street View, OpenStreets[6], and Bing Maps[7], were used. These images, in addition to being georeferenced, also provide information about orientation and inclination, which is crucial for accurate analysis. In this study, images from Google Street View, accessed via its API, were used.
- Multispectral images, particularly those within the infrared spectrum, were also employed. The PNOA captures infrared images that are instrumental in classifying

[1] Föra website
[2] EUH4D website
[3] CeADAR website
[4] Cineca HPC website
[5] CNIG website
[6] OpenStreets website
[7] Bing Maps website

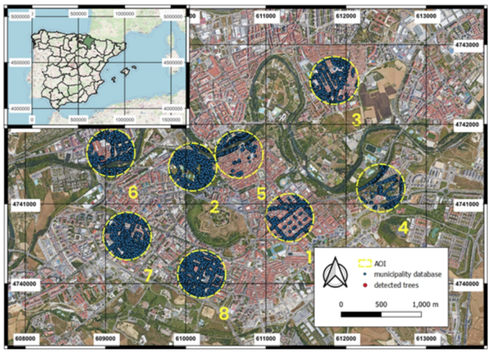

Fig. 1. Validation regions sampled within the city of Pamplona.

LiDAR cloud maps. However, access to this data comes with an associated cost.

The methodology was applied to the city of Pamplona, selected for its unique availability of high-density LiDAR data (14 points per square meter), as well as comprehensive Google Street View coverage and multiple years of PNOA orthoimagery. To validate the tree detection and positioning results, the municipal urban tree inventory of Pamplona was used. This inventory, a publicly accessible database, provides approximate locations of trees, collected at the street level by local authorities[8].

For validation purposes, eight circular regions with a radius of 300 meters each were defined across different urban topologies in Pamplona (see Fig. 1). These regions served as representative samples for comparing the results obtained from the methodology.

To establish a baseline for the tree positioning project, the PyCrown library (Zörner et al. 2018), a Python package for identifying tree top positions in a canopy height model (CHM) and delineating individual tree crowns, was utilized with a range of parameters. The goal was to find the parameter combination that minimized false positives while maintaining high F1 scores and accuracy. The results of these various parameter settings were analysed, and the best-performing configurations were highlighted. Although a specific combination of a smooth filter (SF) and window size (WS) of 2 **pixels** did not always yield the highest F1 score or accuracy, it provided a low false positive rate, which was prioritized in this study. Consequently, these parameters were selected as the project's baseline.

The detection of urban trees was carried out using a multi-step process that began with LiDAR point cloud data and employed the PyCrown library for detecting tree crowns and tops. The methodology was designed to maximize the number of geolocated tree points, even at the expense of higher false positive rates. Subsequent filtering techniques were applied to reduce false positives and improve overall results compared to the optimal PyCrown configuration.

The process involved several key steps:

1. Initially, the LiDAR point cloud data was processed to generate height models, such as the Vegetation Height Model (MDAV), Terrain Height Model (TDM), and Surface Height Model (MDS) (see Ginzler 2021, for application of similar models in another regional context). The PyCrown algorithm was then applied to these models to detect and extract tree tops and crowns in a GIS format (Tomlinson et al. 1976).
2. The detection of tree crowns and tops relied on the identification of zonal **maxima** and the analysis of point cloud slopes. The algorithm's configuration was set to maximise the number of detected points, accepting the likelihood of more false positives.
3. A cadastral layer, obtained from local council records, was used to filter out points located within building boundaries, as these were unlikely to represent trees.
4. Trees detected within 4 meters of each other were grouped together using a specified radius value, allowing the algorithm to identify and cluster neighbouring trees based on their proximity.
5. To further reduce false positives, an object detection model, which automatically identifies trees in RGB images by drawing bounding boxes around them, was ap-

[8] Tree inventory of Pamplona

plied to the aerial orthograph images. Any points from the previous steps that did not fall within the bounding boxes drawn by the model were filtered out.
6. The Normalized Difference Vegetation Index (NDVI, see Huang et al. 2020, for a recent review) was applied to the LiDAR data to identify and retain only those points associated with vegetation. This technique improved accuracy by filtering out non-vegetated surfaces, such as buildings and roads.
7. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise, Ester et al. 1996) algorithm was used to further cluster and analyse the spatial data, particularly for identifying and grouping tree points within the LiDAR data.

To ensure the accuracy of the tree detection methodology, the urban tree inventory of Pamplona was used as a reference. Given that the inventory's spatial accuracy was approximate, ground truth points were manually revised, adding missing trees and removing non-tree points. The validation process involved matching the closest control tree to each detected tree, checking for reciprocity, and calculating the distance between points. If the distance was within 3.5 meters, the detection was classified as a true positive; otherwise, it was considered a false positive. Ground truth points without associated detections were classified as false negatives. This thorough validation process ensured a reliable assessment of the tree detection methodology.

2.2. Image database for urban tree species classification

Hereafter we present a methodology for the automatic generation of a dataset containing RGB images of various tree species captured at ground level.

These images are retrieved from GSV based on geo-located points, and several algorithms are applied to categorise the images by tree species. The resulting dataset is designed for training a classification model capable of automatically identifying tree species in operational settings.

Our primary goal is to create a dataset of 20,000 images, focusing on species such as *Acer negundo, Catalpa bignonioides, Celtis australis, Gleditsia triacanthos, Ligustrum japonicum, Melia azedarach, Platanus x hybrida, Robinia pseudoacacia, Sophora japonica, Ulmus pumila.* This approach will serve as a framework for analysing additional tree species. Images are automatically downloaded by directing the GSV camera towards trees based on their geo-locations, with each image labelled according to the known species at that location. Two different classifiers and an anomaly detection algorithm were trained on these images, and the classification into specific species of images was only considered to be corrected when all the developed models gave consistent results.

The dataset was created using the publicly available Madrid's urban tree inventory[9], which provides geo-located points that specify the location, species, and height of the trees. The information was stored in a GeoPackage ("gpkg") file. Using the coordinates from this inventory, we extracted RGB images of trees via the Google Street View service. This service allows for the retrieval of the closest GSV image to a given point and supports camera adjustments through parameters such as coordinates (latitude and longitude), heading (horizontal orientation), pitch (vertical orientation), and field of view. For our methodology, we set the pitch to 0° and the field of view to 110°.

Due to potential inaccuracies in geolocation and the possibility of multiple trees appearing in a single image (e.g Fig. 2), we employed the Faster R-CNN object detection model (Ren et al. 2015), with Inception-ResNet v2 (Szegedy et al. 2016) as the backbone, pre-trained on the Open Images Dataset (Krasin et al. 2016). This model was implemented using TensorFlow (Abadi et al. 2015) to detect and isolate trees within the images. The most centred tree in each image was selected as corresponding to the database point. Detailed views were then obtained by adjusting

[9] Open data portal of Madrid

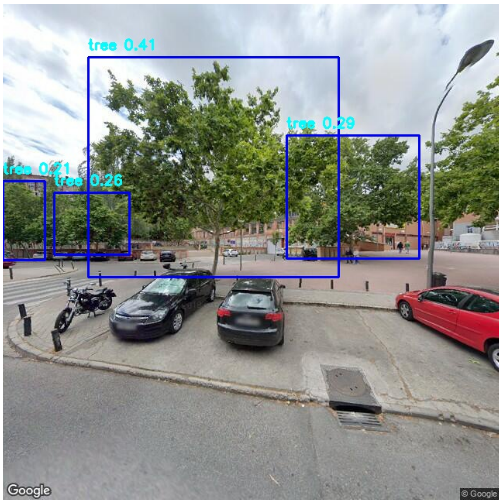

Fig. 2. Multiple trees detected in the same field of view of an image downloaded from Google Stree View.

the zoom parameter. The resulting images were stored in folders categorised by tree species.

Given potential misclassification due to geolocation errors, we performed a manual curation step, resulting in a dataset of 16,854 images. To address species with limited data, additional images were sourced from the GBIF dataset[10]. These images were filtered using a ResNet101 model (He et al. 2016) to extract features, followed by dimensionality reduction with UMAP (McInnes et al. 2018) and clustering with HDBSCAN (Campello et al. 2013), adding 5,749 high-quality images to the dataset. The final manually curated dataset consisted of 22,603 images distributed across several tree species.

The curated dataset was split into training and validation sets, with 80% allocated for training and 20% for validation. A ResNet152 image classifier (He et al. 2016), pre-trained on ImageNet (Deng et al. 2009), was fine-tuned to automatically identify tree species.

Additionally, a separate dataset of 44,751 images, automatically acquired was used to train the second classifier. This dataset was split into a 95% training set and a 5% validation set. A VGG16 classifier (Simonyan 2014) was trained to differentiate between each species, complemented by ROC curve analysis to optimise classification thresholds. Moreover, a one-class Support Vector Machine (OCSVM, Schölkopf et al. 1999) was trained on this dataset to score images based on their likelihood of belonging to a species class.

The final procedure to classify the images is given by the combination of the results of ResNet152, VGG16, and OCSVM classifiers. Images were considered valid only if they were classified as the same species by both ResNet152 and VGG16 and were not flagged as anomalies by OCSVM. This approach was applied to the automatically retrieved dataset of 27988 images, resulting in a final clean dataset discussed in Sect. 3.2. Precision was used as the primary performance metric, given the ability to repeatedly access Google Street View data for additional images. Due to the scarcity of certain species, the GBIF image retrieval method was employed to ensure a minimum of 2,000 images per species.

To estimate the performance of the dataset generation process, a sample of 200 images per tree species was randomly selected and manually validated. This sampling approach was chosen due to time constraints, under the assumption that it represents the overall quality of the dataset.

3. Results and discussion

In this section, we will go through the results of the two different tasks that are part of the *ArbörIA* project. The results demonstrate the effectiveness of the methods presented above in achieving significant improvements, highlighting the potential of sophisticated algorithms in urban tree geolocation and identification.

3.1. Urban trees geo-positioning

The implemented object detection filter effectively reduced the overall number of false positives across all eight zones by 7.24%, surpassing the initial target of a 5% reduction. Although this improvement is less than

[10] GBIF website

	Improvement in %						
Zone	TP	FN	FP	Precision	Recall	F1	Accuracy
Zone 1	2.65	-0.45	-61.11	13.7	0.65	6.12	7.83
Zone 2	21.65	-28.98	44.44	-3.82	12.46	4.79	5.23
Zone 3	6.08	-7.69	-43.64	14.47	3.25	8.79	9.39
Zone 4	23.99	-37.79	22.76	0.22	13.46	2.57	2.06
Zone 5	15.23	-19.81	15.42	-0.05	8.52	3.54	3.45
Zone 6	22.06	-35.23	21.30	0.13	13.54	6.91	8.02
Zone 7	14.90	-28.41	-24.53	5.1	9.56	7.79	10.48
Zone 8	-0.67	1.86	-7.04	0.87	-0.52	0.05	0.07
Mean	13.43	-21.88	7.24	3.83	7.62	5.07	5.82

Table 1. True Positive (TP), False Negative (FN), False Positive (FP), Precision, Recall, F1 and Accuracy percentage improvements reported for each zone and averaged over them.

the 14% reduction observed at the mid-term stage, the integration of new algorithms has significantly enhanced the performance metrics, including Precision, Recall, F1 score, and Accuracy.

The results (see Tab. 1) demonstrate a marked improvement in detection accuracy. For instance, in Zone 1, Precision increased from 75.19% to 88.89%, and Accuracy from 56.05% to 63.88%. Similar enhancements were observed across other zones, with an average increase of 3.83% in Precision and 5.82% in Accuracy. This improvement is attributed to the strategic application of advanced techniques, including NDVI calculation and DBSCAN clustering. The NDVI filter, sensitive to vegetation, effectively distinguished vegetation points, while the DBSCAN method identified and preserved dense clusters of true positives, reducing noise and false positives.

In conclusion, the geolocation methodology used for identifying trees in urban environments proved to be highly effective, achieving a 7.24% reduction in false positives and significantly enhancing overall precision. The combination of mid-term and final-term techniques has not only met but exceeded the objectives, underscoring the efficacy of advanced filtering and clustering methods in improving the accuracy and reliability of urban tree detection. Nevertheless, new techniques and better data are being sought to compensate for the weaknesses of the current method, while preserving or even improving its strengths, and also to reduce the differences we find between the different areas under consideration.

3.2. Image database for urban tree species classification

The application of our approach to a final dataset of 27988 images yielded a finally automatically labelled and curated dataset of 26,519 images, surpassing the target of 20,000 images across ten tree species, with each species represented by at least 2,000 images (see Tab. 2). The precision of our method varied depending on the species, likely due to morphological and anatomical differences. For example, Platanus x hybrida and Celtis australis had the highest precision rates at 97.5%, as their distinctive features were more easily identifiable. Conversely, Gleditsia triacanthos proved challenging to distinguish, possibly due to anatomical similarities with other species or lower-quality training data, resulting in a lower precision of 89.5%. Despite these variations, the overall precision was satisfactory across all species, with the number of correctly labelled images compensating for the few misclassified samples. This outcome highlights the robustness of our methodology in generating a reliable dataset, which is essential for training accurate tree species classifiers.

Our results demonstrate that while the method's effectiveness is influenced by species-specific characteristics and data quality, the precision achieved is adequate for practical applications. The approach shows

Specie	Images	Precision (%)
Acer negundo	2088	95.5
Catalpa bignoides	2234	90.5
Celtis australis	2066	97.5
Gleditsia triacanthos	3188	89.5
Liigustrum japonicum	2191	96.0
Melia azedarach	3826	95.5
Platanus x hybrida	4083	97.5
Robinia pseudoacacia	2016	95.0
Sophora japonica	2115	95.5
Ulmus pumila	2712	95.5

Table 2. Resulting number of images after applying our ensemble model to filter mislabeled images and precision of the model.

significant potential for automated tree species identification and can be further improved by incorporating more advanced computer vision techniques and larger training datasets. This would enhance the reliability and accuracy of tree species classification, particularly when dealing with a broader range of species.

4. Conclusions

This study presented two pivotal activities within the *ArbörIA* project, each contributing to the project's advancement in distinct yet complementary ways. This project is an extension of the work presented by Rodríguez-Puerta et al. (2022). Different methods for detecting and geolocating trees in urban environments were developed, using several open data sources to generate automatically curated datasets of urban public green space. The proposed approach would provide municipalities with much more accurate data on the amount of vegetation in a city and the benefits it provides.

Firstly, the geolocation methodology for identifying trees in urban settings demonstrated a notable 7.24% reduction in false positives, surpassing the initial goal of 5%. This achievement was facilitated by the integration of LiDAR point cloud data, advanced filtering techniques, and NDVI calculations, which collectively enhanced the precision of the tree geolocation algorithm. The results consistently outperformed the baseline, emphasising the role of advanced algorithms in optimising geolocation accuracy and reducing false positives.

Secondly, an automated methodology for generating an RGB image dataset of urban tree species was developed using Google Street View, Madrid's urban tree inventory, and GBIF data. By leveraging deep learning, clustering, and dimensionality reduction algorithms, we successfully curated a dataset of the ten most common tree species with satisfactory precision, despite challenges related to mislabeling and visually similar species. This approach demonstrates the potential for future applications in automatic species identification, though improvements in training data and algorithmic sophistication are needed for more reliable classification across diverse species.

Future efforts will focus on expanding both the image database and refining the geolocation methodology. Enhancements in the dataset, through external databases and data augmentation, alongside further optimisation of tree detection algorithms, are expected to improve species classification and geolocation accuracy, pushing the project's impact on urban tree management forward.

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org

Albawi, S., Mohammed, T. A., & Al-Zawi, S. 2017, in 2017 International Conference on

Engineering and Technology (ICET), 1–6
Anguelov, D., Dulong, C., Filip, D., et al. 2010, IEEE Computer, 43, 32
Arozarena, A., Villa, G., & Valcarcel, N. 2005, in
Calvert, J. G. 1990, Pure and Applied Chemistry, 62, 2167
Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in Advances in Knowledge Discovery and Data Mining, ed. J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Berlin, Heidelberg: Springer Berlin Heidelberg), 160–172
Deng, J., Dong, W., Socher, R., et al. 2009, in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96 (AAAI Press), 226–231
Ginzler, C. 2021, Vegetation Height Model NFI
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, 1
Huang, S., Tang, L., Hupy, J., Wang, Y., & Shao, G. 2020, Journal of Forestry Research, 32
Krasin, I., Duerig, T., Alldrin, N., et al. 2016, Dataset available from https://github.com/openimages
McInnes, L., Healy, J., & Melville, J. 2018, arXiv e-prints, arXiv:1802.03426
Rapinel, S., Hubert-Moy, L., & Clément, B. 2015, International Journal of Applied Earth Observation and Geoinformation, 37, 56, special Issue on Earth observation for habitat mapping and biodiversity monitoring
Ren, S., He, K., Girshick, R., & Sun, J. 2015, arXiv e-prints, arXiv:1506.01497
Rodríguez-Puerta, F., Barrera, C., García, B., Pérez-Rodríguez, F., & García-Pedrero, A. M. 2022, Sensors, 22, 3269
Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., & Platt, J. 1999, in Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS'99 (Cambridge, MA, USA: MIT Press), 582–588
Simonyan, K. 2014, arXiv preprint arXiv:1409.1556
Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. 2016, arXiv e-prints, arXiv:1602.07261
Tomlinson, R., Calkins, H., & Marble, D. 1976, Computer Handling of Geographic Data: An Examination of Selected Geographic Information Systems (The UNESCO)
Zörner, J., Dymond, J. R., Shepherd, J. D., Wiser, S. K., & Jolly, B. 2018, Forests, 9