Graphical Models 136 (2024) 101234

Contents lists available at ScienceDirect

Graphical Models

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

Building semantic segmentation from large-scale point clouds via primitive recognition

Chiara Romanengo[1], Daniela Cabiddu *,[1], Simone Pittaluga, Michela Mortara

CNR-IMATI, Via De Marini 6, Genova, 16149, Liguria, Italy

ARTICLE INFO

Keywords:
Point clouds
Semantic segmentation
Fitting primitives
Feature recognition
Urban digital twins

ABSTRACT

Modelling objects at a large resolution or scale brings challenges in the storage and processing of data and requires efficient structures. In the context of modelling urban environments, we face both issues: 3D data from acquisition extends at geographic scale, and digitization of buildings of historical value can be particularly dense. Therefore, it is crucial to exploit the point cloud derived from acquisition as much as possible, before (or alongside) deriving other representations (e.g., surface or volume meshes) for further needs (e.g., visualization, simulation). In this paper, we present our work in processing 3D data of urban areas towards the generation of a semantic model for a city digital twin. Specifically, we focus on the recognition of shape primitives (e.g., planes, cylinders, spheres) in point clouds representing urban scenes, with the main application being the semantic segmentation into walls, roofs, streets, domes, vaults, arches, and so on.

Here, we extend the conference contribution in Romanengo et al. (2023a), where we presented our preliminary results on single buildings. In this extended version, we generalize the approach to manage whole cities by preliminarily splitting the point cloud building-wise and streamlining the pipeline. We added a thorough experimentation with a benchmark dataset from the city of Tallinn (47,000 buildings), a portion of Vaihingen (170 building) and our case studies in Catania and Matera, Italy (4 high-resolution buildings). Results show that our approach successfully deals with point clouds of considerable size, either surveyed at high resolution or covering wide areas. In both cases, it proves robust to input noise and outliers but sensitive to uneven sampling density.

1. Introduction

Urbanization is rapidly increasing and our cities are facing challenges like increasing load of traffic, air pollution, energy consumption, limited green spaces beside climate change.

Interdisciplinary research and development efforts are seeking the Digital Twins of cities [1–5] to represent, monitor, simulate and predict the complex processes that take place in urban environments.

This work positions itself within several initiatives on urban digital twins [6,7], specifically on the geometric modelling of the urban space, its characterization and annotation. The analysis of the morphology of the built environment is crucial to understand urban processes: for example, the built structures determine how air flows in the city, what surfaces receive sunlight or are shaded by nearby buildings, and where architectural barriers hinder accessibility. Through the mechanism of annotation, contextual knowledge from different sources can be linked to the geometry, and heterogeneous information related to the same region or urban element can be directly accessed: as an example, the volume of a specific building, its energy demand and the number of residents. The salient urban features, in the above example buildings, need to be identified within the whole geometric representation, and annotated with diverse knowledge, e.g., from GIS layers, from georeferenced sensors, by manual user annotation of the 3D model through a Graphical User Interface, etc. Automatic or at least semi-automatic recognition of salient urban features would be a great help in the creation of the semantic 3D city model. An "as-automatic-as-possible" annotation would need to solve two main problems: (i) locate the portion of geometry that represents an urban feature, and (ii) extract geometric attributes that provide additional knowledge automatically.

Many methods address the extraction of features in indoor scenes [8]. In outdoor environments, most urban segmentation methods have focused on detecting large-scale elements such as buildings, vegetation, and roads. Methods that operate at a smaller, sub-building scale, such

* Corresponding author.
E-mail addresses: chiara.romanengo@cnr.it (C. Romanengo), daniela.cabiddu@cnr.it (D. Cabiddu), simone.pittaluga@cnr.it (S. Pittaluga), michela.mortara@cnr.it (M. Mortara).

[1] Joint first authors.

https://doi.org/10.1016/j.gmod.2024.101234
Received 6 April 2024; Received in revised form 9 August 2024; Accepted 1 October 2024
Available online 10 October 2024

as roof extraction, primarily aim at reconstruction and concentrate on identifying the planes that constitute the roof itself [9,10]. In this paper, we focus on the constituent parts of buildings: walls, roofs, and pavement. We propose a recognition method based on fitting primitives to identify these features and extract geometric attributes like orientation and size as additional information. Our method can distinguish individual façades and roof faces and is not limited to identifying planar features, unlike the majority of approaches in the literature. In addition, we are able to manage point clouds representing a single building, differently from the methods in the literature, whose input point cloud typically represents an entire or a part of a city.

Our approach works on point clouds acquired either by laser scanning technologies, aerial and/or terrestrial, or photogrammetry; the input point clouds may present diverse characteristics (e.g., resolution, accuracy, coverage) and additional attributes (e.g., colour, classification) depending on the acquisition mode, technology and survey campaign. Each acquisition methodology has drawbacks (e.g., occlusions, outliers and noise) and the survey often happens in an uncontrolled environment, with occluding elements such as parked or passing vehicles and pedestrians. Therefore, the method must be robust to missing data, noise and outliers. Modelling and performing geometric analysis at geographic scale is also challenging in terms of data storage and processing efficiency.

Like other approaches, we exploit the Hough transform, which prevents over-segmentation and is robust to input noise and outliers, but it is not efficient; therefore, we employ a two-stage preliminary space partitioning approach, first dividing the input cloud by buildings and then further partitioning each building into sub-parts to enhance computational efficiency. By combining RANSAC with the Hough Transform (HT), we improve HT computation time through initial segmentation and mitigate the oversegmentation issue typically associated with RANSAC alone. Our approach can segment large point clouds and is inherently suitable for parallel processing.

Our semantic segmentation recognizes instances of planes, cylinders, and spheres, providing the parametric form of each feature, and partitions the input point cloud into separate files of points belonging to roofs, walls, floor, domes, arches, vaults of each specific building. The cloud is not required to have additional information but the 3D coordinates of each point. For each feature, additional attributes that characterize the shape are automatically computed: orientation and pitch of roofs, height and width of walls, radius of domes, arches and vaults (and length for the latter two). The characterization of urban features at sub-building scale represents a huge support to the heavy work of annotating a whole 3D city, and more features will be managed in future developments. The semantic city model is anticipated to enhance urban management; for example, attributes such as the area, orientation, and inclination of roof faces can offer informed estimates of photovoltaic energy potential.

This work extends the conference contribution in [11], where we presented our preliminary results on single high-resolution buildings. In this extended version, we generalize the approach to manage whole cities, by preliminary splitting the point cloud building-wise, and streamlining the pipeline. We run a thorough experimentation with a benchmark dataset from the city of Tallinn (about 47 K buildings) [12] and two case studies in Italy, namely the city of Catania (3 buildings of interest at high resolution acquired by terrestrial laser scanning and the city centre area acquired by aerial photogrammetry) and Matera (1 building at high resolution acquired partially by terrestrial laser scanning and partially by aerial LIDAR technology). Finally, we compare our results with related works using a portion of the Vaihingen dataset.

2. Related works

Representing an object through a set of geometric components with a semantic meaning is a long-standing problem in different domains, such as Computer Vision, Computer Graphics, Computer-Assisted Design (CAD). Recently, the semantic segmentation of high-resolution point clouds representing urban contexts has gained much attention [13].

Generally, the objective of 3D point cloud segmentation is to subdivide points into separate homogeneous regions, ensuring that points within the same region exhibit similar characteristics or meaning. Challenges in point cloud segmentation include the quality of input data, which frequently includes high redundancy, uneven sampling density, missing data, noise, and outliers, as well as the absence of a clear structure within the data. In the urban setting, point clouds of large areas typically come from aerial laser scanning and may suffer of low or uneven density or missing data due to the orientation and inclination of external surfaces with respect to the acquisition trajectory.

The majority of methods for 3D semantic segmentation are based on one of these approaches: global energy optimization, feature clustering, region growing or model fitting. Methods within the first category transform the spatial segmentation problem into an energy optimization problem (e.g., [14] and references therein). These methods can achieve global optimization, but most of them require an initial segmentation, and the minimization of complex functions is challenging on discrete and unordered input data. Feature clustering classifies points according to some characterization (e.g., normals) and aggregate points with similar attributes with k-means or similar clustering approaches (e.g., [15]). These methods are sensitive to noisy data and depend on the definition of a proper neighbourhood size for points. Region growing approaches iteratively expand an initial point or region to adjacent areas until some growing criteria are satisfied, e.g., [9]. These methods are sensitive to the selection of the seeds and performance depends on the selection of the growth criteria. Model fitting-based approaches can estimate robust primitive parameters from the points with high noise and outliers [16–19].

Among these approaches, we find stochastic methods based on the RANSAC (RANdom SAmple Consensus) method [20] and its various optimizations, and parameter space techniques that rely on Hough-like voting and parameter space clustering [21]. Li et al. [22] aims at reconstructing scenes from point clouds assuming a regularity of the distribution of buildings, that is the Manhattan world assumption [23]. The reconstruction of buildings model is proposed also in [24] exploiting the RANSAC algorithm to segment the planar patches that constitute rooftops.

The Hough transform is exploited by [25], where a recognition method able to segment the input point cloud into geometric primitives of different types (e.g., planes, cylinders, spheres, cones and tori) is proposed. However, this approach is focused on CAD objects. Finally, the study provided in [26] aims to extract building roof planes from airborne LIDAR data applying an extended Randomized Hough Transform, without incorporating semantic information.

HT is time-consuming, sensitive to the parameter values, and may find spurious planes [27]. RANSAC is relatively less time-consuming, but cannot handle the problem of spurious planes [28].

We propose a strategy focused on a combination of the RANSAC approach and a recognition method based on the Hough transform (HT). The RANSAC is able to provide an initial segmentation of the input point cloud in a quite efficient way, associating to each segment the type of primitive it corresponds to. The subsequent use of the HT for the recognition of surfaces allows us recognizing different types of primitives, not just planes, associating the parameters that uniquely identify each segment as geometric descriptors, as shown in [29], with an approach that has proven robust to noise in the input data. Through the analysis of the geometric descriptors, it is possible to group different segments that belong to the same primitive thus avoiding the oversegmentation typical of the RANSAC. Combining the two approaches, we are able to manage large scale point clouds and to extract salient information on the urban features.

Concerning the type of segmentation outcome, we can identify a large group of methods tackling the identification of urban features at a higher scale: such approaches classify points as building, vegetation, road and other general types, similar to the LiDAR classification, e.g., [30,31]. Other tackle a mixed selection of features, e.g., facades, ground, cars, motorcycles, traffic signs, pedestrians, vegetation in [32]. Therefore, it is not straightforward to compare our segmentation results with previous work; the same problem arises when looking for datasets with ground truth to test our results (e.g., [33,34]). Some methods tackle the identification of detail features, such as roofs, but their aim is the reconstruction of buildings rather than an accurate segmentation [35,36]. Li et al. [9],Wang and Ji [14] also tackle the identification of roof planes and optimization of roof edges for further reconstruction, but achieve an intermediate point cloud segmentation into roof faces as we do. We provide a qualitative comparison with these work on the Vaihingen dataset (see Fig. 13). However, their methods seek only planar features. Some methods, particularly supervised learning approaches (e.g., [37,38]), rely on additional information that may be embedded within the point cloud, such as intensity or RGB colour. However, the availability of this data largely depends on the acquisition technology and may not always be present.

We experimented our method on the Tallin dataset from a recent benchmark called Building3D [12]; it provides the largest urban-scale dataset meant for aerial LiDAR point cloud modelling of building roofs. We also compared our results quantitatively and qualitatively with [9,14] on the Vaihingen dataset achieving very good outcomes.

To summarize, our approach focuses on the segmentation of buildings into their constituting salient parts, i.e., roof, façades and pavement, and it is able to manage also high-resolution point clouds representing a single building, differently from most of urban segmentation methods that identify features at larger scale. In addition, it is not limited to planar patches but recognizes curved roof surfaces as well (e.g., domes and vaults). Our methods use geometry alone and does not rely on additional information that may depend from the acquisition device (e.g., colour). We apply a combination of RANSAC and HT to improve efficiency, achieve robustness to noise and avoid over-segmentation.

3. Our approach

As anticipated, the pipeline for urban feature recognition and documentation works directly with the point cloud to limit memory allocation. By taking into account that for large-scale objects such as cities, data size can be critical and that the recognition process is quite time-consuming, we aim to split the input in order to manage storage and processing efficiently.

Therefore, we partition the whole initial cloud at two levels: firstly, at the semantic level, we segment the urban area building-wise. Secondly, at the geometric level, we apply binary space partitioning to each building to obtain small chunks that can be efficiently processed. Then, a combination of the well-known RANSAC algorithm and a recognition method based on the Hough transform is applied chunk by chunk to obtain a semantic segmentation of buildings into their main features such as façades, walls, and roofs. Finally, the recognition is performed on the remaining cloud(s) representing the open spaces.

The combination of the RANSAC algorithm and the HT-based recognition method is advantageous, since the first one provides the classification of points primitives, while the HT associates to each primitive the geometric descriptors that uniquely identify it. Specifically, these descriptors are necessary for segmenting the point cloud in a semantic way, since they provide important information regarding the location and orientation of the primitives. On the other hand, the traditional HT requires in input a point cloud representing a single primitive and, if possible, the type of primitive associated to it, in order to reduce the computational cost, so the use of the RANSAC algorithm is a good way to achieve this end.

Our method documents each feature with its salient quantitative attributes as metadata, such as inclination and orientation for roofs.

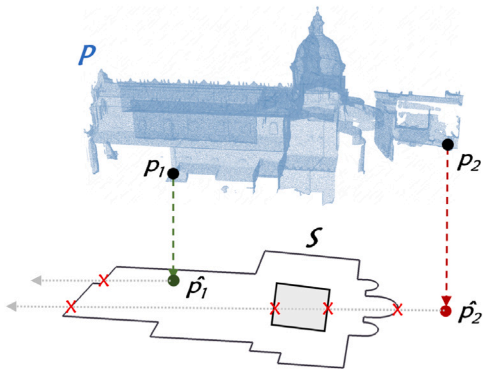

Fig. 1. Point-in-Polygon. The input includes both the point cloud $\mathcal{P}$ and a polygonal representation $\mathcal{S}$ of the footprint of the building of interest. Each point p in $\mathcal{P}$ is projected on the plane of $\mathcal{S}$ and the Jordan Curve Theorem is applied to assess if the projected point $\hat{p}$ lies within the footprint.

3.1. Building-wise partitioning

We assume prior knowledge about the (2D) footprints of the city buildings is available, with no loss of generality, in the form of standard ESRI Shapefiles [39]. Many municipalities do already offer this kind of information as OpenData, but if this is not the case, online repositories may also be used (e.g., OpenStreetMap [40]).

The partitioning is based on the point-in-polygon test, a basic geometry operation widely adopted in GIS applications. For a detailed discussion on the point-in-polygon problem and different approaches, see [41].

Our implementation extends the point-in-polygon method introduced by W. Randolph Franklin [42] to handle also polygons with multiple boundaries, including holes; this is crucial for urban environments, as many buildings exhibit inner courtyards. The algorithm relies on the Jordan Curve Theorem [43], which asserts that a point $\hat{p}$ resides inside a polygon if the number of crossings of a half-line starting at $\hat{p}$ in any arbitrary direction is odd (see Fig. 1). If the shapefile contains multiple disjoint polygons (e.g., for a subset of buildings), we apply the same approach (possibly in parallel) to the entire set of polygons in the shapefile. Similarly, we can restrict the domain to blocks, areas, or city districts whose boundary shapefile is known and run the semantic segmentation in parallel, thus improving efficiency.

To prevent numerical precision issues and account for situations where parts of the building may extend beyond the footprint (e.g., balconies or sloping roofs), our implementation allows for the option to perform the point-in-polygon check by offsetting the polygon by a specified distance.

3.2. Binary space partitioning

After dividing the input cloud into buildings, we process each unit separately.

However, even point clouds containing a single building might be huge (see Table 1). If this is the case, we partition the point set again, this time using a space-based approach, namely the out-of-core partitioning approach described in [44]. The algorithm segments the input cloud into chunks with a maximum cardinality, an input user-defined parameter. The cardinality should be tuned according to the performance of the machine doing the processing: lower cardinality means more chunks, that is faster processing of a single chunk but slower combination of results when a feature spans multiple chunks.

Fig. 2 shows how the partitioning algorithm works. After computing the cloud bounding box, denoted as $\mathcal{B}(\mathcal{P})$, (Fig. 2b), points are downsampled to a representative set $\mathcal{S}(\mathcal{P})$ by randomly selecting one vertex every 1000 in $\mathcal{P}$. Starting from $\mathcal{B}(\mathcal{P})$, the in-core binary space partition

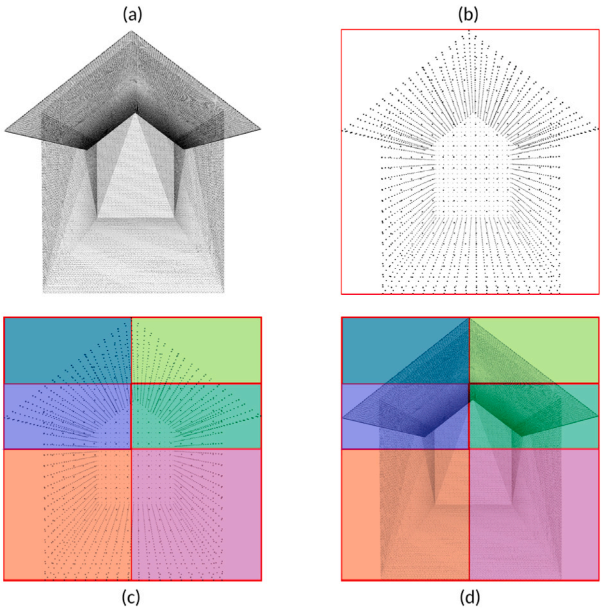

Fig. 2. Binary Space Partitioning. (a) Input point cloud. (b) Bounding Box and downsampling of the input cloud. (c) BSP computed on the downsampled set. (d) Final BSP.

(BSP) is constructed by iteratively subdividing the bounding box cell containing the largest number of points of $\mathcal{S}(\mathcal{P})$. Each cell is split along its longest side. The root of the binary space partition corresponds to the entire downsampled cloud $\mathcal{S}(\mathcal{P})$. During each subdivision, the points in the parent cell are assigned to one of the two offspring cells based on their position. If a vertex lands precisely on the dividing plane, it is assigned to the cell with the lowest lexicographical barycentre. The process continues until the number of points within each BSP cell is below to a predefined threshold (Fig. 2c). Once the BSP structure is established as described above, the remaining points in $\mathcal{P}$ must be assigned to their respective BSP cells, based on their spatial positioning (Fig. 2d); since the cardinality of the input cloud is much higher, this segmentation is done out-of-core.

The process saves the output chunks into N separate files, namely *cell_i.xyz*, where $i = 0, \dots, N-1$.

3.3. Point classification

After the (optional) partitioning phases (building-wise and/or BSP), the recognition of geometric primitives is performed using a fitting approach. We point out that both the previous partitioning enable a parallel execution of the fitting procedure over the partitions. However, the algorithm can proceed sequentially analysing a chunk at a time, benefiting of the cardinality reduction nonetheless. In the following, we describe the sequential approach on each sub-cloud $\mathcal{P}_i$ returned by the binary space partitioning.

Firstly, we apply a RANSAC classification [20], that is an automatic algorithm to detect basic shapes in unorganized point clouds. This method requires in input the minimum number of points constituting a segment and the type of primitive to look for. In our implementation, we set the first parameter as a percentage of the input point cloud (i.e. 0.5% of the cardinality of each cell) and we select three types of primitives that are more likely to be found in an urban environment: planes, cylinders and spheres. The result is a collection of subsets of points belonging to the same primitive, saved as a *.txt* file, whose name identifies the type of primitive and a sequential identifier (e.g., *sphere_1.txt*). An example of this result is shown in Fig. 5b, where the RANSAC algorithm is applied to each sub-cloud.

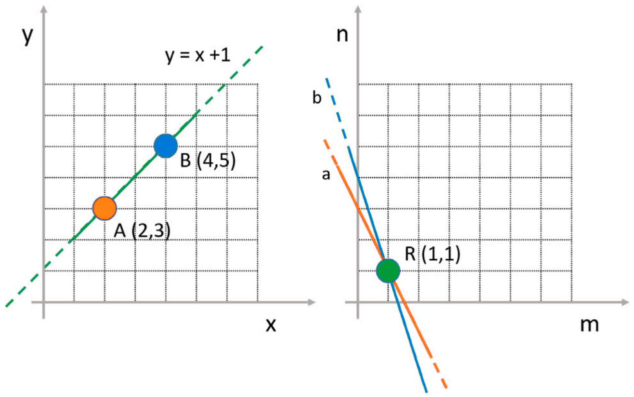

Fig. 3. The HT is based on the *point-line duality*: points A and B lie on a straight line. These lines correspond to lines in the parameter space that intersect in a single point R. This point uniquely identifies the coefficients in the equation of the original straight line.

Note that, the combination of the BSP partitioning and the tendency of the RANSAC algorithm to oversegment the point clouds (see, for example, [18]) is likely to generate different segments of points that actually belong to the same primitive; for this reason, we complement the RANSAC partitioning with a recognition step, based on an extension of the Hough Transform (HT).

3.4. Primitive recognition and characterization

The original definition of the Hough transform (HT) is based on the *point-line duality* as follows: points on a straight line, defined by the equation $y = mx + n$, correspond to lines in the parameter space that intersect in a single point. This point uniquely identifies the coefficients in the equation of the original straight line (see Fig. 3). This concept can be naturally extended to a generic family $\mathcal{F} = \{\mathcal{S}_{\mathbf{a}}\}$ of curves or surfaces that depend on a set of parameters $\mathbf{a} = (a_1, \dots, a_n)$ [45]. More in details, given a family $\mathcal{F} = \{\mathcal{S}_{\mathbf{a}}\}$ depending on a set of parameters $\mathbf{a} = (a_1, \dots, a_n) \in U' \subset \mathbb{R}^n$, where U' is an open set of $\mathbb{R}^n$, a general point P in the space corresponds to a locus, $\Gamma_P(\mathcal{F})$, in the parameter space U'. As P varies on a given curve $C_{\mathbf{a}}$ from $\mathcal{F}$, a set of curves Γ_P is generated. If the set of curves Γ_P meets in one and only one point $\bar{\mathbf{a}} \in U'$, the family of curves $\mathcal{F}$ verifies the so-called regularity condition and the intersection point defines the parameters of the best fitting curve $C_{\bar{\mathbf{a}}}$. The duality concept is fundamental for the HT based recognition algorithm, since it translates the recognition problem into detecting which value of the parameters that determine the family $\mathcal{F}$ corresponds to the curve or the surface best fitting a given set of points (such a value may be non unique). The common strategy to identify the solution (or a solution) is based on the so-called *accumulator function*; it consists in a voting system whereby each point in a point cloud $\mathcal{P}$ votes a n-uple $\mathbf{a} = (a_1, \dots, a_n)$; the most voted n-uple corresponds to the most representative curve or surface for the profile.

We apply a generalization of the HT to families of surfaces (planes, spheres, cylinders, cones and tori) devised for the CAD context [29]; in urban scenes, the main structural elements can be identified by planes, cylinders and spheres, so we restrict the recognition to these primitives (see Fig. 4).

To optimize the recognition of instances of these primitives, we use the approach in [29], which exploits primitive canonical forms to reduce the dimensionality of the parameter space. Intuitively, a sphere centred in the origin can be uniquely determined by the value of its radius; similarly, a cylinder with its rotational axis aligned to z-axis; a plane passing through the origin is defined by its normal versor. The preliminary classification of (chunks of) point clouds (given by RANSAC in our case) allows to choose the family of primitives to use

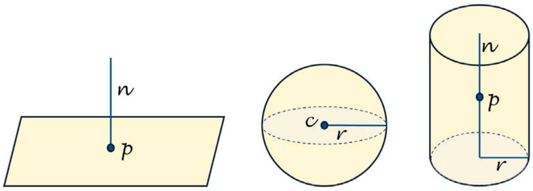

Fig. 4. Plane, sphere and cylinder primitives respectively, along with their attributes (geometric descriptors).

by the HT-recognition algorithm. Indeed, the expected primitive type is provided as the name of the *.txt* file containing each segment. Once the family is select, it is possible to move segments to their canonical forms, find their parameters using the HT, move back the points to the original position and finally determine the complete set of geometric attributes that specify the instance:

- *Planes.* The geometric descriptor of a plane is represented by the vector [**n**, **p**], where **n** is the normal to the plane and **p** is a point lying on it.
- *Spheres.* The geometric descriptor of a sphere is identified by the vector [**c**, *r*], where **c** contains the coordinates of its centre and *r* is the length of the radius.
- *Cylinders.* The geometric descriptor of a cylinder is represented by the vector [**n**, **p**, *r*], where **n** is the rotational axis **n**, **p** is a point lying on it and *r* is the length of its radius.

However, in some cases, especially in presence of significant noise, the RANSAC algorithm fails in associating the correct primitive type. Following the strategy described in [29], we enrich the recognition procedure with the computation of an approximation error to evaluate the RANSAC classification. If the approximation error is higher than a fixed threshold, we iteratively test the other types of primitives and select the one with lowest approximation error, correcting the classification and exploiting the robustness to noise typical of the HT.

3.5. Primitive aggregation

The geometric attributes described in Section 3.4 are used to aggregate segments belonging to the same primitive, as shown in Fig. 5c. In this step, we exploit the complete linkage, that is a hierarchical clustering approach useful to compare clusters and build a dendrogram [29].

First, the aggregation assigns every single segment to a cluster and then it iteratively merges clusters that are closest with respect to the following map

$$D(C_h, C_j) := \max_{\boldsymbol{\alpha}_k \in C_h, \boldsymbol{\alpha}_l \in C_j} d(\boldsymbol{\alpha}_k, \boldsymbol{\alpha}_l),$$

where (C_h, C_j) is a given pair of clusters and d is a measure of distance or dissimilarity. Following the notation introduced in Section 3.4, the distances considered in this work differ with respect to the type of primitive:

- in case of planes, $d(\alpha_1, \alpha_2) = \|\mathbf{n}_1 \times \mathbf{n}_2\|_2 + |\mathbf{n}_1 \cdot (\mathbf{p}_1 - \mathbf{p}_2)|$;
- for spheres, $d(\alpha_1, \alpha_2) = |r_1 - r_2| + |\mathbf{c}_1 - \mathbf{c}_2\|_2$;
- in case of cylinders, $d(\alpha_1, \alpha_2) = |r_1 - r_2| + \|\mathbf{n}_1 \times \mathbf{n}_2\|_2 + \|\mathbf{n}_1 \times (\mathbf{p}_1 - \mathbf{p}_2)\|_2$.

Note that, if $d(\alpha_1, \alpha_2) = 0$ (or less than a threshold in our implementation), then the primitives α_1 and α_2 are equal with respect to the selected criterion.

3.6. Semantic segmentation

So far, we have performed geometric analyses and derived a segmentation into instances of primitive types, avoiding over-segmentation. Finally, the contextual knowledge provides rules to recognize features of the urban environment and their components. Typically, as shown in Figs. 19, 20 and 21, spherical parts identify domes, cylinders arches, and planes represent parts of buildings, such as roofs, walls and pavements. The geometric attributes described in Section 3.4 can be used to refine the classification. So far, we did not elaborate further on spheres and cylinders, because our experimental datasets provide too few examples; however, the radius, height and principal axis orientation will likely discriminate arches, vaults and columns. We are investigating this point in current developments. Instead, we focus on planes at first, whose normals and relative position effectively distinguish them into façades, walls, roofs, and pavement, as shown in Fig. 5d.

The components of the normal **n** of each plane determine whether it is vertical, horizontal or oblique. According to this:

- vertical planes are annotated as façades;
- oblique planes are labelled as roofs;
- horizontal planes are classified as pavement or roofs based on their elevation.

For each semantic feature, further attributes will be needed according to the application scenario. Currently, we are interested in describing buildings from the energetic point of view (energy demand and consumption). In particular, the orientation and pitch are crucial parameters to determine the photovoltaic potential of roofs, and they are trivially determined from the roof normal. For the Digital Twin of Catania [46] for instance, we are going to couple these findings with the overall sunlight received by each roof during a year [47] and the roof surface to determine the amount of solar panels, their photovoltaic potential and the expected saving [48].

4. Experiments and results

To test and evaluate our method, we set up a few experiments. The first exploits the publicly available dataset representing the city of Tallinn, included in the Building3D framework [12]. The second experiment focuses on the Vaihingen dataset, which is part of the Urban Modelling and Semantic Labelling Benchmark by the International Society for Photogrammetry and Remote Sensing (ISPRS). This dataset has been provided as a reference for a challenge on urban classification, 3D reconstruction, and 3D labelling. Specifically, we refer to the 3D labelling, which includes the“roof” class. The last experiments come from datasets acquired in the framework of two Italian projects: the UISH project [46], and the CTEM project [49], aiming at developing Digital Twins of Catania and Matera in Italy. The three experiments are described in the subsections below. Section 4.4 provides a theoretical analysis of the computational complexity of the pipeline, step by step.

Experiments were conducted on a Windows 11 desktop workstation equipped with an i9 18-core CPU and 128 GB of RAM.

4.1. Building3D benchmark dataset

Building3D [12] offers an extensive urban-scale dataset designed for building roof modelling using aerial LiDAR point clouds. This dataset encompasses over 160 thousand buildings across 16 Estonian cities, spanning approximately 998 square kilometers. It comprises building point clouds, roof point clouds, mesh models, and wireframe models.

Tallinn, the largest city within this dataset, contains around 47,000 building point clouds, each stored in XYZ format with detailed information including coordinates, RGB colour, near-infrared data, intensity, and reflectance. Most importantly, the dataset comprehends the results

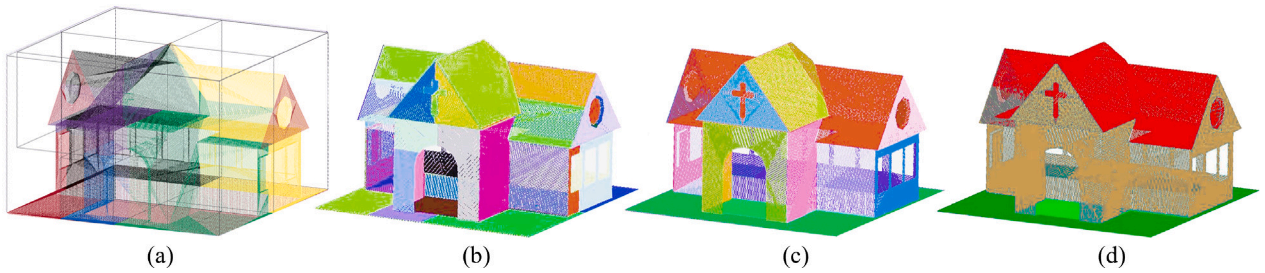

Fig. 5. Segmentation and primitive fitting. (a) The set of chunks $\mathcal{P}_i$, with $i = 0, \ldots, 5$ returned by the BSP. (b) Result of the RANSAC segmentation applied to each $\mathcal{P}_i$. (c) Result of the aggregation of segments belonging to the same planes after the recognition step. (d) The semantic segmentation including roofs, walls and pavements.

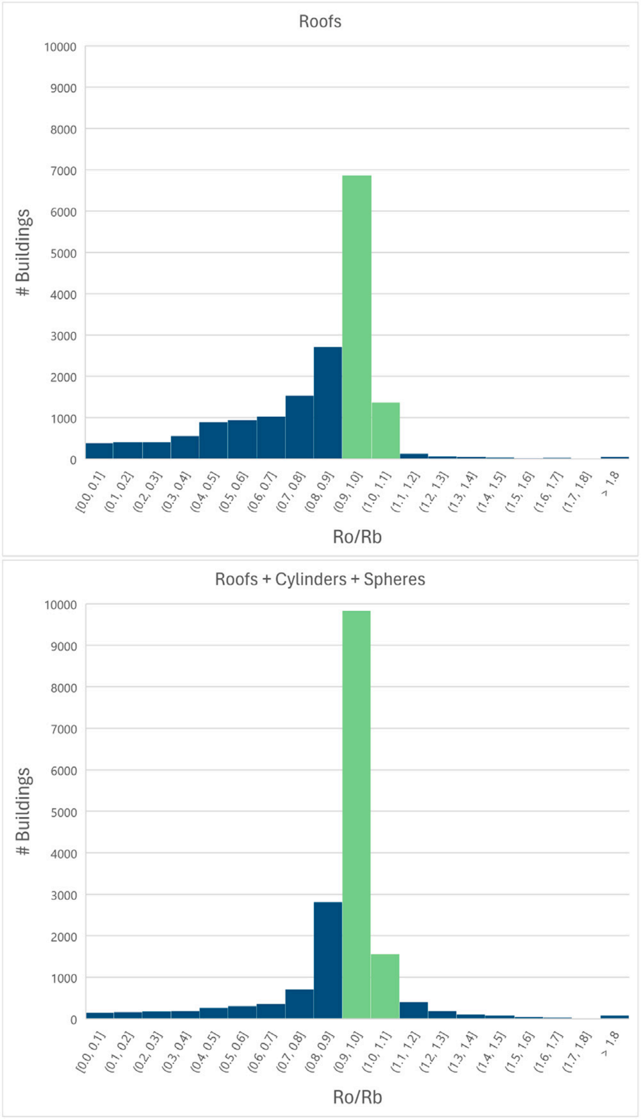

Fig. 6. Histogram of the ratio between the point count of each resulting R_o and its corresponding R_b in the benchmark with respect to the number of buildings. In the top image, R_0 considers only points on planes classified as roofs, while in the bottom picture, the value of R_o represent the sum of points classified as roofs, cylinders and spheres. Green bins highlight the buildings for which the same number of points of the benchmark is almost reached, that is the ratio between the point count of each resulting R_o and its corresponding R_b is approximately equal to 1.

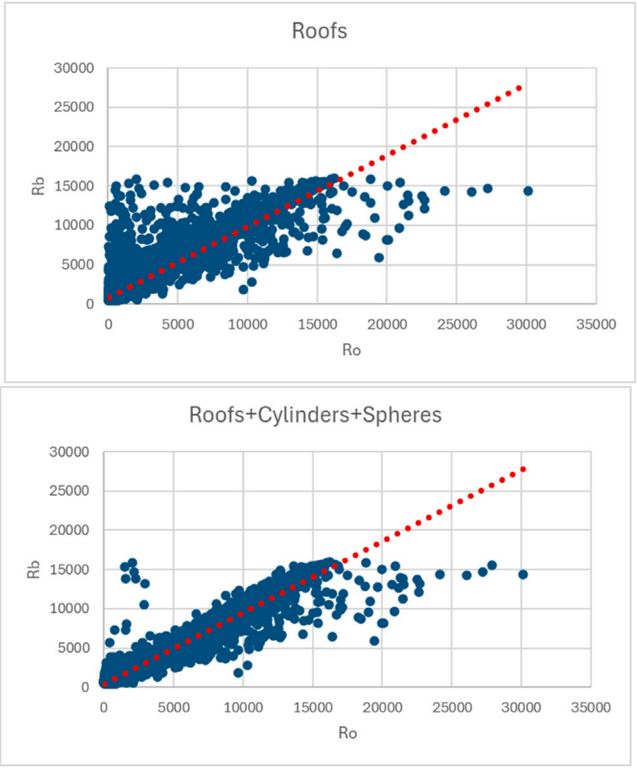

Fig. 7. The ratio between the point count of each resulting R_o and its corresponding R_b in the benchmark. In the top image R_0 considers only the points classified as roofs, while in the bottom, the value of R_o represent the sum of points classified as roofs, cylinders and spheres.

of roof partitioning, which were generated using a commercial software, Terrasolid,[2] with manual editing, to create building mesh models from aerial LiDAR point clouds and building footprints. Subsequently, mesh faces parallel to the XY plane, assumed to represent facades, were removed. A point is designated as part of a roof if its distance to the roof mesh model falls within a specified threshold. Every roof is represented as a sub-point cloud of its building of origin.

Therefore, in our experiment, we focus on roof identification to compare our results with the Building3D dataset. We point out that this dataset cannot be used as a proper ground truth, as it has not been apparently validated. Indeed, not all the buildings in the dataset do have a recognized roof. It is however the largest freely available dataset and better resource for comparison.

As each point cloud in the benchmark represents a single building and is manageable in size (the largest consisting of approximately 1

[2] https://terrasolid.com/

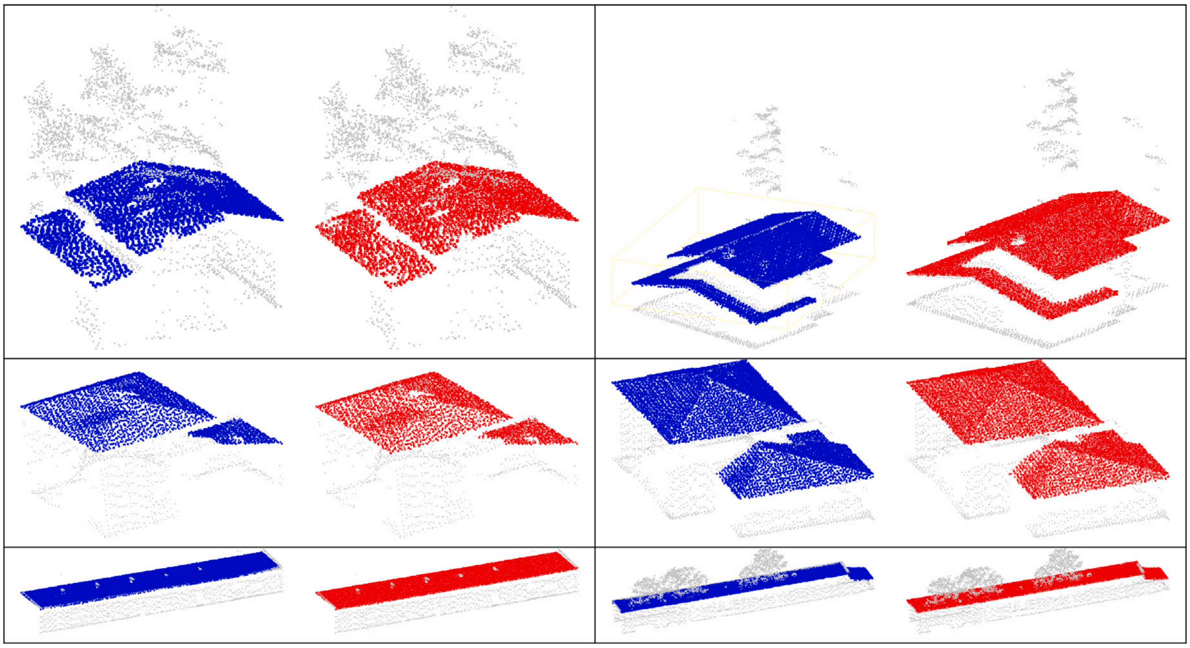

Fig. 8. Example of results on Building3D datasets. For each couple, blue points are classified as roof in the Building3D dataset, while red points show our roof classification.

million points), there was no need for preprocessing the input dataset to partition it building-wise or to conduct binary space partitioning; therefore, the process can start directly with point classification. On each building, the RANSAC technique was applied with the minimum requisite number of points necessary to identify a geometric feature, set at 0.2% of the number of the input points. This specific value was chosen through empirical means and demonstrated effective within this particular case study. Various segments were produced for each building, encompassing planes, cylinders, and spheres. Finally, the Hough transform was employed to aggregate these segments, leading to the semantic segmentation.

To assess the performance of our approach, we randomly sampled 30,000 buildings from the Building3D dataset, 17,567 of which have a reference roof available for comparison.

To measure the accuracy of our results, we compare each roof given by our approach (R_o) with the corresponding roof in the benchmark (R_b), and analyse:

- the ratio between the number of points of R_o and R_b;
- the Hausdorff distance $d_H(R_o, R_b)$ between R_o and R_b, i.e.

  $$d_H(R_o, R_b) = max\{\max_{i\in R_o}\min_{j\in R_b} d(i,j), \max_{j\in R_b}\min_{i\in R_o} d(j,i)\}$$

  where d is the Euclidean distance. Since coordinates of this dataset are metric, the distance value is expressed in meters.

The ratio between the number of points of each R_o and its corresponding R_b with respect to the number of building is shown in Fig. 6 through histograms. The upper image considers as R_o the number of points classified as roofs, that is, belonging to planes not strictly vertical. As you can see, in the first case, we were able to mark as roof nearly the same number of points of the Building3D benchmark for about 8200 buildings of the tested cases.

Fig. 8 displays some results of roofs with the most similar number of points, which are also visually close in terms of shape.

We investigated the extreme cases, where our method labels much more or much less points. Where R_o is much greater than R_b, our method generally performs better, as it is able to identify whole roofs while the benchmark misses a considerable number of points (see Fig. 9). Conversely, where R_o is much lower than R_b, we miss-classified roofs, even if we correctly segmented them. For very small relative inclination of adjacent roof pieces, RANSAC classifies the whole set as a huge cylinder rather than two planes of similar inclination, or simply mis-classifies planes into other primitives. In our method, we assumed that roofs might only be planar, but this is actually not the case: indeed, other failure examples exhibit cylindrical roofs. Therefore, we tested the results including also cylinders and spheres as roofs (see Fig. 10). The bottom histogram in Fig. 6 shows the distribution considering as R_o the number of points classified as roofs, cylinders and spheres. The performance apparently improves, since the same number of points of the benchmark is approximately reached for about 11400 buildings.

This result is confirmed by the graph in Fig. 7 in which the trend of the ratio between the number of points of each R_o and its corresponding R_b tends to be linear. Specifically, by comparing the two graphs, it is evident that including cylindrical and spherical primitives improves th result.

Finally, Fig. 10 show same samples of the dataset in which part of roofs are classified as cylinders and spheres by our method. In conclusion of this analysis, the classification of roofs composed of only planar segments can be improved by considering also segments classified as cylinders or spheres. This assumption holds for this dataset, because it is derived from aerial LIDAR acquisition and so, roof points are well represented while façades are nearly absent. In any case, we plan to automatically classify cylinders and spheres as roofs (and discriminate them from other features, like columns) by studying the rotational axis (cylinders) and the position of the centre (centre). Note that some parts of roofs in Fig. 10 are misclassified by RANSAC despite the fact that they are planar segments and the recognition method approximated them with degenerate cylinders and spheres with a satisfactory approximation error.

Fig. 11 provides histograms of the Hausdorff distance with respect to the number of buildings. This analysis shows that the distance to their counterparts in the benchmark is lower than 1 meter for nearly 6,000 buildings out of the 17,567 tested. This number increases in the second case, in which R_o represent the sum of points classified as roofs, cylinders and spheres: more than 8,000 buildings have distance less than one meter, and more than 14,000 are below five meter distance.

Unsuccessful cases are due to noisy points wrongly classified as roofs, as shown in Fig. 12. In particular, the ten cases with highest Hausdorff distance comprehend trees next to the buildings, which our method misclassifies as roof in some cases.

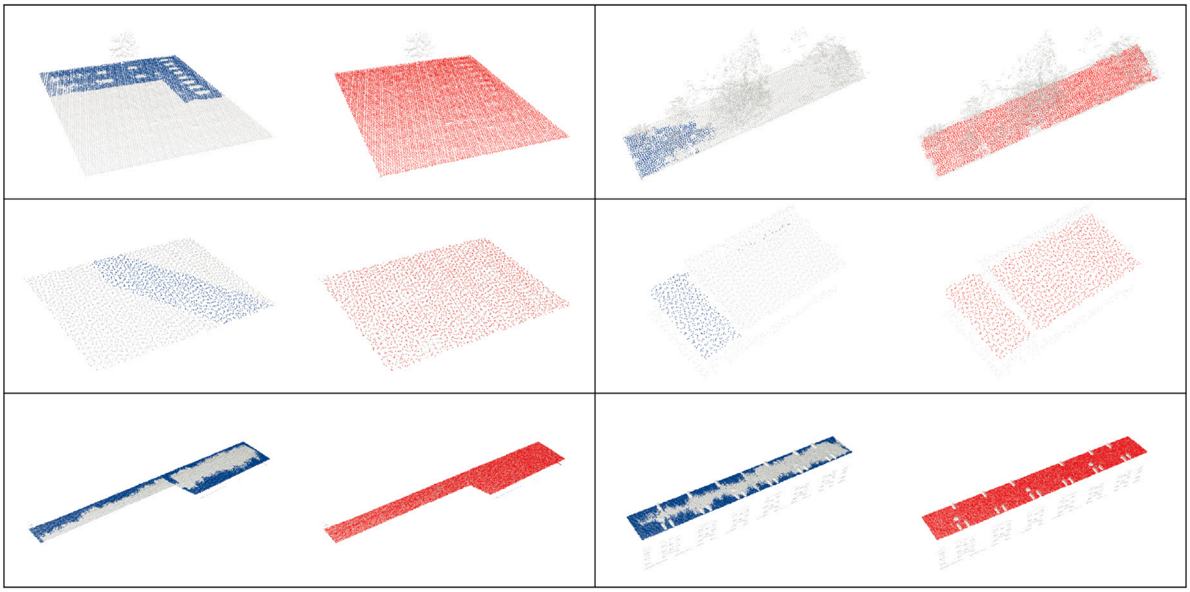

Fig. 9. Results on the Building3D datasets where our method performs better in the roof identification. For each couple, blue points are classified as roof in the Building3D dataset, while red points show our roof classification. The examples in this figure have the highest number of points ratio between R_o and its corresponding R_b in the Building3D dataset.

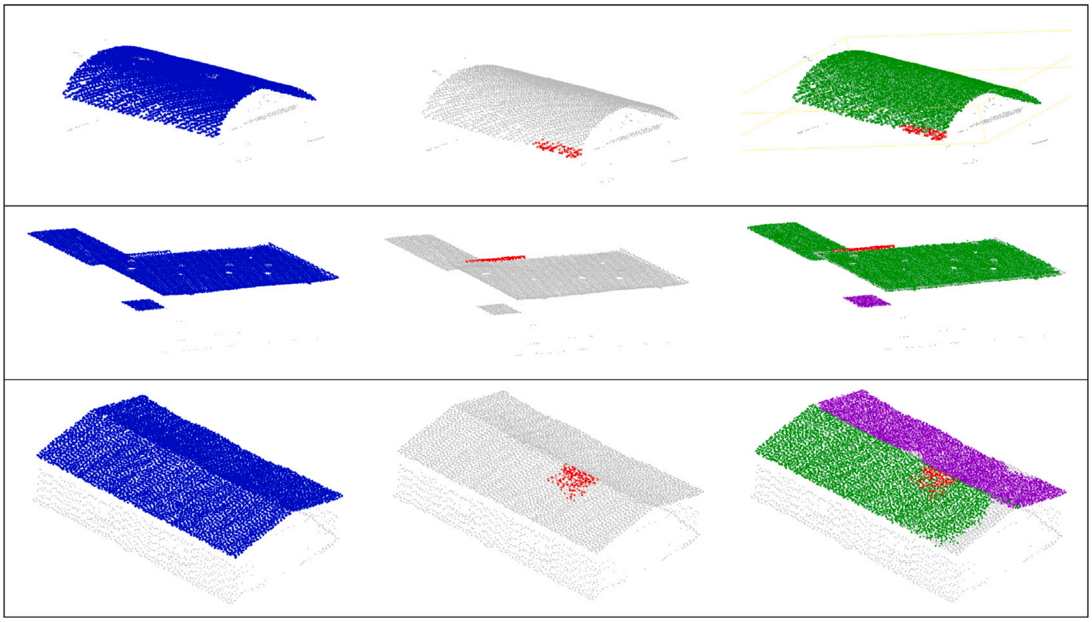

Fig. 10. Results on the Building3D datasets. In the first column, blue points are classified as roof in the Building3D dataset. In the second column, red points represent a planar area and are classified as roof by our method. In the third column, the combination of planar segments and cylindrical (in green) and spherical (in purple) segments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Vaihingen dataset

The Vaihingen dataset is a subset of the data used for the test of digital aerial cameras carried out by the German Association of Photogrammetry and Remote Sensing (DGPF) [50]. In particular, it consists of areas of the city characterized by different kind of built structures. This dataset has been used by several works, especially based on supervised learning, because has been the subject of some challenges and as such provides a ground truth (e.g., [36]). The challenge presented in [51] was focused on urban object detection and 3D building reconstruction and involved high-scale features (building, road, tree, low vegetation/grass, and artificial ground). The challenge was extended to 3D semantic labelling,[3] providing a training set and a test set. Each point in the dataset is labelled according to 9 classes, including façades and roofs [52].

To evaluate our approach, we focused on the test set shown in Fig. 13 and we computed the *precision* and F1 score to compare it with the other challenge participants, besides the accuracy measures proposed in Section 4.1. Regarding roofs, we reach a precision of

[3] https://www.isprs.org/education/benchmarks/UrbanSemLab/3d-semantic-labeling.aspx

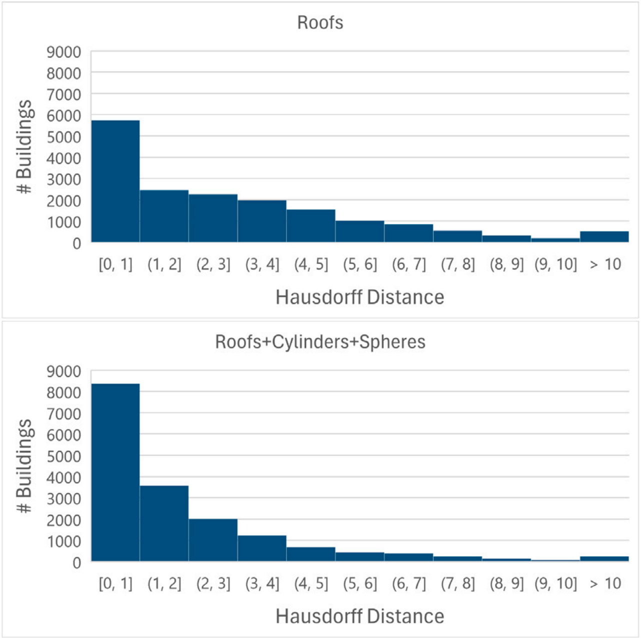

Fig. 11. The histogram of the Hausdorff distance with respect to the number of buildings. In the upper image R_0 considers only the points classified as roofs, while in the lower R_o includes points classified as cylinders and spheres.

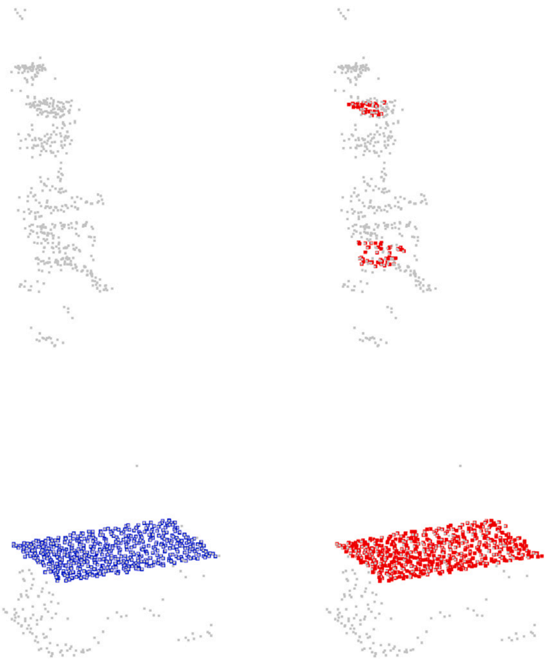

Fig. 12. Example of failure result, where part of the vegetation is classified as roof. This example corresponds to the highest Hausdorff distance value.

0.94, while for façades we gain a precision of 0.62. The value obtained for façades is lower because in the building-wise partitioning step (Section 3.1) some points of façades are often missed and the RANSAC suffers from the different resolution between façades and roofs typical of aerial LiDAR acquisition. At the same time, the use of the footprint allows us to extract even roof folds that are not labelled in the reference. We provide Fig. 14 for a qualitative evaluation of the result.

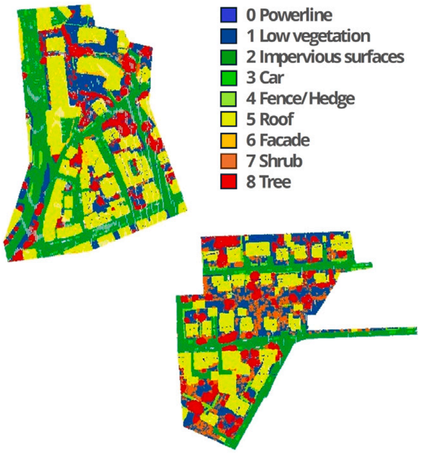

Fig. 13. Labelled points extracted from the Vaihingen dataset (test set).

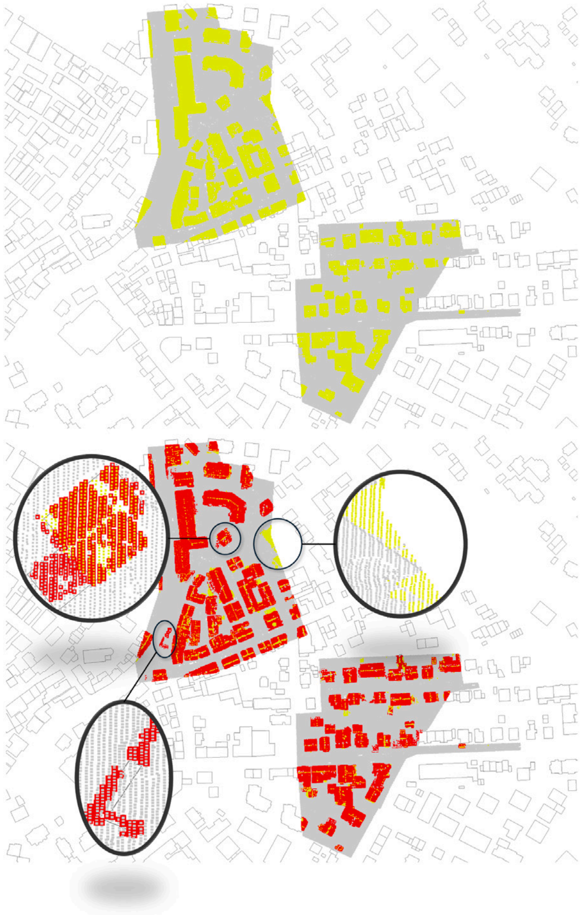

Fig. 14. Visual comparison. (Top) The original Vaihingen dataset with points classified as roof highlighted in yellow and building footprints shown in the background. (Bottom) Our results with points classified as roof highlighted in red, projected onto the original dataset. We zoomed in on specific details to demonstrate that the quality of the result depends on the footprints used for the building-wise partitioning. The building-wise partition was achieved by offsetting the footprints by 1 meter.

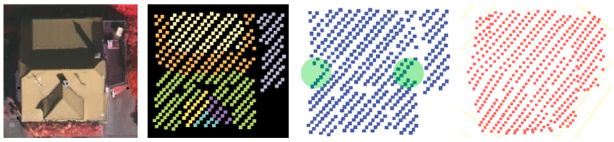

Fig. 15. Visual comparison with [14]. From left to right: reference image; results by Wang and Ji [14]; results without fold distinction, and in green two areas of false negatives; our result. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

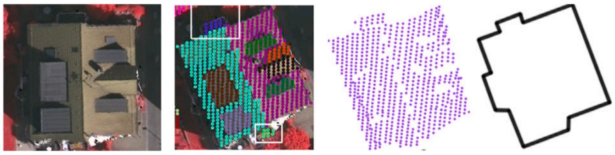

Fig. 16. Visual comparison with [9]. From left to right: reference image; results by Li et al. [9], and in the white boxes two dubious areas; our result; the building footprint as an additional reference. Given the footprint, it seems points in the upper box are correctly identified as roof by both methods, while those in the lower box seems to be false positives.

Considering the evaluation of the results of about 25 methods provided by the authors of the challenge,[4] we note that our approach achieves the highest precision with respect to the other methods analysed, except for few of them. Those methods actually use not only geometric information as we do, but consider additional data, i.e., intensity and RGB attributes for the approach in [38] and the intensity in [37]. Concerning the F1 score, we reach a value of about 0.91, which is among the highest for methods based on geometry alone. We also note that our method suffers of a few false negatives due to the building-wise segmentation based on footprints: extruding sections of roofs and part of façades are not analysed at all. For this reason, we have introduced a small buffer area around the footprints, but we have to improve this aspect in future developments.

We end the analysis of the results quantitatively considering the accuracy measures introduced in Section 4.1. The method obtained a value of ratio of 0.89 for roofs and 0.72 for façades, that confirm the presence of few false positive mentioned before. Finally, the values of Hausdorff distance 13.3 and 17.8 for roofs and façades, respectively. These values represent the distance between the buildings in the groundtruth that are not identified by our method (see Fig. 14, bottom, zoom on the right side).

From a qualitative point of view, we visually compare our results with [9,14] on two buildings of Vaihingen. With reference to Fig. 15, our approach recognizes the two lateral triangular folds, but [14] provides more pleasant roof borders thanks to an optimization step. In Fig. 16 our method misses the central ridge points but avoids the false positives at the bottom. Overall, we note that comparison with other methods is hindered by too scarce and not validated ground truth.

4.3. Matera and Catania digital twin case studies

The digital representation of the morphology of the urban environment is a core component of the system and the work reported in this paper represents a building block for the reconstruction of a semantic 3D model for two Italian cities, Matera [49] and Catania [46], where two digital twins are being developed on part of their urban area. In both cases, various datasets were acquired, using different acquisition techniques. In particular, an extended area of the city centre of Catania (approximatively 2.5 squared km) has been digitized by aerial photogrammetry, giving a point cloud of nearly 100G points (average ground sampling distance 3–5 cm per pixel); terrestrial long and mid range laser scanning systems, reaching a resolution smaller than 12 mm has been used to capture at high resolution buildings of high artistic and historical value. These are important buildings in the main square of Catania (Piazza del Duomo), namely Palazzo degli Elefanti (Elephants' Palace), the St. Agatha's Cathedral (also called "Duomo") and Palazzo dei Chierici (Table 1). Acquiring stations have been placed in 110 locations overall, at ground and higher floors, in the nearby alleys and on opposite buildings; the setting did not allow a perfectly even covering and density.

Table 1
Size of the input datasets of Catania (millions of points).

Input	# Points
Elephants' Palace	≈ 541
St.Agatha's Cathedral	≈ 850
Palazzo dei Chierici	≈ 757

Concerning the acquisition of the city centre of Matera, LIDAR aerial acquisition technologies has been adopted for surveying a large area of the municipality, while a portable, lightweight laser scanner has been used for terrestrial acquisition of limited sites of interest, as shown in [53].

Referring to the three historical buildings of Catania, the input point clouds have been divided into sub-clouds, each containing a maximum of 10 million points. In contrast, this partitioning step was not required for processing the Matera dataset, as the size of each building was small enough for direct processing. However, in the case of Matera, the building-wise partitioning approach was necessary to isolate single buildings.

Fig. 17 shows the result of the pipeline run on the "Palazzo degli Elefanti" dataset. The binary space partitioning returned 103 sub-clouds, that are then processed by the RANSAC obtaining 306 segments (see Fig. 17a). The last step, that is the aggregation of segments belonging to the same primitives (see Fig. 17b). Finally, Fig. 17c shows the resulting semantic segmentation of the palace, distinguishing the planes among façades (in light blue), roofs (in red) and pavement (in green). As you can see in Fig. 21, this dataset also presents cylinders and spheres within the building, on the arcades of the cloister.

Fig. 18 presents the result of our method on the "Palazzo dei Chierici" dataset, a more complex structure than the previous one. The result of the binary space partitioning step is a set of 162 sub-clouds. This point clouds are processed by the RANSAC producing 435 segments (see Fig. 18a) that are then aggregated (see Fig. 18b). Finally, in Fig. 18c the resulting semantic segmentation of the building is shown, highlighting in light blue the planes belonging to façades, in red the planes belonging to roofs and in green the planes belonging to the pavement.

Fig. 19 provides the result of our algorithm on the "St. Agatha Cathedral" dataset. As you can see, this dataset has heterogeneous density: the upper parts including the roof and the decorations in the high portion of the cathedral exhibit a lower resolution. This is due to the constraints in the acquisition phase, as the terrestrial laser stations could only be placed on accessible terraces of adjacent buildings and necessarily could not cover at best the highest parts. Unfortunately, the segmentation given by RANSAC suffers from this double resolution and consequently loses some elements of the original point cloud. The binary space partitioning step produces in this case a set of 189 sub-clouds. This point clouds are processed by the RANSAC producing 3698 segments (see Fig. 19a) that are then aggregated into 172 primitives (see Fig. 19b). Fig. 19c shows the resulting semantic segmentation of the building, highlighting in light blue the planes belonging to façades, in red the planes belonging to roofs, in green the planes belonging to the pavement and in pink the parts corresponding to cylinders and spheres.

[4] https://www.isprs.org/education/benchmarks/UrbanSemLab/results/vaihingen-3d-semantic-labeling.aspx

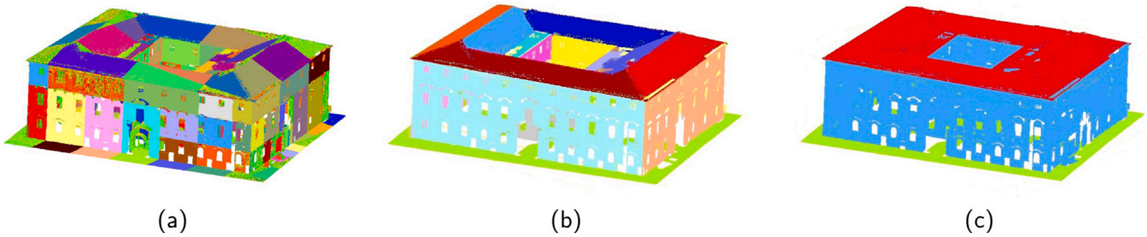

Fig. 17. Palazzo degli Elefanti: in (a) the resulting preliminary segmentation, where different colours correspond to different segments; in (b) the segments that belong to the same plane are grouped; in (c) the planes classified as façade, roof and floor are grouped. Courtesy of Romanengo et al. [11].

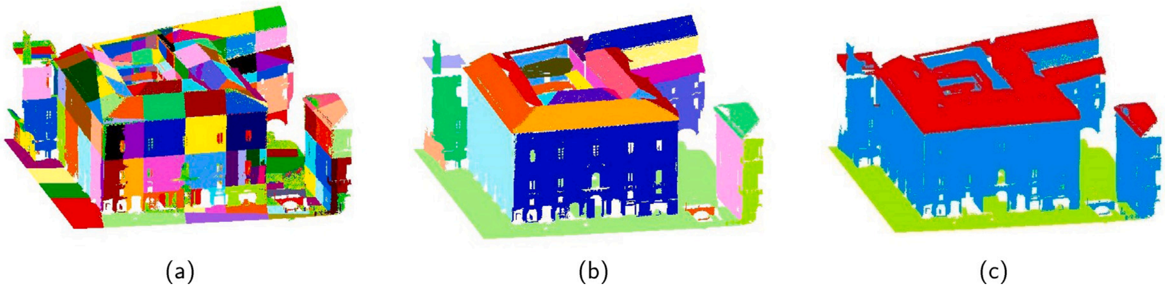

Fig. 18. Palazzo dei Chierici: in (a) the resulting preliminary segmentation, where different colours correspond to different segments; in (b) the segments that belong to the same plane are grouped; in (c) the planes classified as façade, roof and floor are grouped. Courtesy of Romanengo et al. [11].

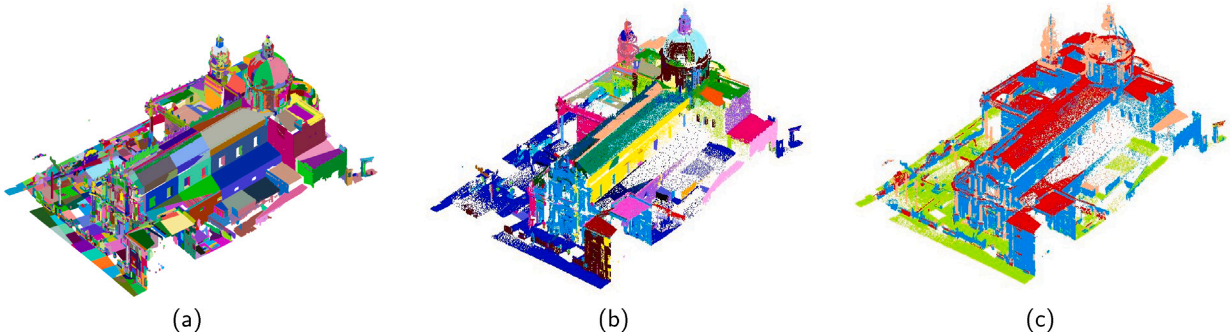

Fig. 19. St. Agatha's Cathedral: in (a) the resulting preliminary segmentation, where different colours correspond to different segments; in (b) the segments that belong to the same plane are grouped; in (c) the planes classified as façade, roof and floor are grouped.

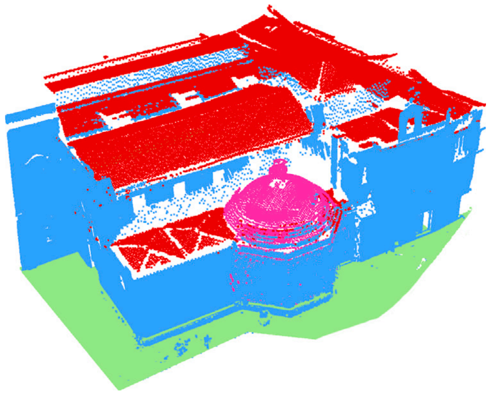

Fig. 20. Semantic segmentation of the Church of Saint Dominic in Matera (5.8 millions of points).

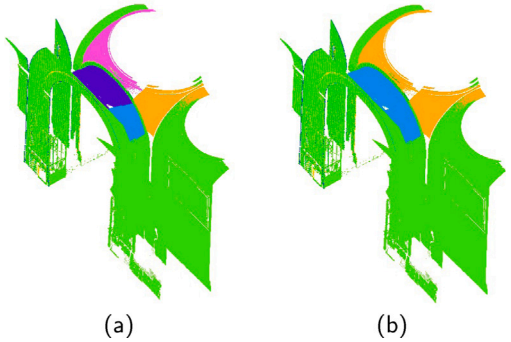

Fig. 21. Focus on an arcade in the cloister of Palazzo degli Elefanti dataset, in which parts of cylinders and spheres are detected. In (a) four segments classified as cylinders (in purple and light blue) and spheres (in yellow and magenta); in (b) the aggregation of segments belonging to the same cylinder (in light blue) and sphere (in yellow). Courtesy of Romanengo et al. [11]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the experimental configuration described in this section, we set a constant value for the input parameter in the RANSAC method, which corresponds to 0.5% of the size of every sub-cloud produced through binary space partitioning. It is important to note that the proper parameter value is linked to the density of the point cloud: for point clouds with heterogeneous density this can provoke an undersegmentation, as in the case of St. Agatha Cathedral where only a partial detection of the roof occurs. Indeed, roof and facades were acquired using photogrammetry and terrestrial laser scanning respectively, resulting in sensibly different resolutions.

As for the Matera case study, we used the footprints downloaded from OpenStreetMap [40] to segment individual point clouds. In Fig. 20 we show a result over the fusion of the aforementioned segmented cloud from the aerial survey, plus points acquired with a mobile scanner from the ground level. There is a significant different in data density, with the ground acquisition providing much higher resolution. Despite this, roofs, walls, dome and also the pavement are nicely captured.

4.4. Computational complexity

In this section, we discuss and analyse the computational cost of our proposed pipeline, which is a smart combination of various techniques. We provide a detailed breakdown of the computational complexity for each component of the pipeline, demonstrating how our approach achieves a balance between computational efficiency and segmentation quality.

The building-wise partitioning explained in Section 3.1 uses the ray intersection method. For each polygon's edges, an intersection analysis is conducted. The time complexity for each intersection analysis is $O(1)$. Since the number of edges is equal to the number of nodes, the time complexity of point-in-polygon determination is linear with respect to the number of input points [54].

The binary space partitioning in Section 3.2 follows the methodology outlined in [44], simplified to consider an input point cloud without topology. According to the original approach, the time complexity is $O(N)$, linear with respect to the number of input points.

The classification step is based on the implementation provided by Schnabel et al. [20].

The recognition step based on the HT is based on a voting procedure, the complexity is dominated by the size of the accumulator function, discretized as a matrix: denoting M the number of entries of this matrix, the computational cost of the HT recognition on a segment is $O(MS)$, where S represents the number of points of the segment. Note that, the number of parameters involved in the HT computation directly influences the size of M. Following the strategy described in [29], the number of parameters is 3 for planes and 1 for spheres and cylinders.

The aggregation of segments that belong to the same primitive described in Section 3.5 consists of a complete-linkage clustering, one can consider more efficient implementations, such as the one proposed in [55], which costs $O(N_{\text{seg}}^2)$ where N_{seg} denotes the number of segments in the output segmentation. Although the dissimilarity-matrix assembly costs $O(N_{\text{seg}}^2)$, one may note that each entry is computed independently; note that the task is embarrassingly parallel.

5. Discussion and conclusions

In this paper we have proposed a new method for segmenting large scale point clouds representing urban 3D scenarios into geometric primitives to detect some urban features, such as building façades, roofs and arcades. Our approach is able to segment huge point clouds thanks to the application of an out-of-core partitioning, it avoids oversegmentation and is robust to input noise and outliers. We showed results on the semantic segmentation of buildings of high artistic and historical value, acquired at very high resolution, and our method was able to handle such large data sets. The method is not limited to point clouds representing a single building: the approach is general, and we extended it to handle whole districts or cities, by partitioning the cloud building-wise first.

Limitations. As expected, the HT makes the overall approach robust to noise, outliers and missing data; however, the approach can fail in case of misclassification of the RANSAC segmentation that the HT is unable to correct. We have already identified the search for cylindrical and spherical parts of roofs as one of the priorities to improve results.

In addition, our method is not insensible to uneven sampling density. Indeed, the RANSAC algorithm tends to loose some elements of the original point cloud in case it present parts with different resolutions (see Fig. 19).

A major issue for the approach is still the computation time. Currently, the BSP and RANSAC run in parallel; however, the HT computation is still too slow for a single process. Indeed it depends both on the number of points processed and on the dimension of the parameter space as explained in Raffo et al. [29]. To give a general idea of the computational cost, the time required for the experiments in Section 4.1 is in the order of days, for the results shown in Section 4.2 is in the order of minutes, while for the high-resolution point cloud in Section 4.3 is in the order of hours. For this reason, it definitely requires a parallel implementation as well, which should be straightforward to obtain given the nature of the approach. This will be tackled in the very next steps.

The pipeline, while fully automatic, depends on the selection of a few parameters used as thresholds that need to be adjusted on the specific dataset.

Future Works. Further improvements are needed to reach the goal of a semantic 3D city model. One concerns the building-wise partitioning, now based on the building footprints. Since parts of the building may extend beyond the footprint, such as sloping roofs, in our implementation we added the option to set a buffer to the polygon. In our experiments, we set the same buffer for all buildings, but a special care must be paid to points belonging to the street floors and platforms, which would be split in different clouds. Besides, the distinction between horizontal roofs and ground is done by elevation comparison, building-wise. This works flawlessly for flat cities, but we expect it to fail in case of very steep morphology, or with complex housing units that can be adjacent and partially overlapping. An extreme case can be seen in the historical centre of the city of Matera, where it is common to ground a building partially on top of another. For these reasons, we have to work on the choice of the buffer, according to the features of the urban environment.

Aerial acquisition is typically characterized by missing data on the façades aligned with the flight direction: buildings cast "shadows" to the structures that remain behind and are hidden to the sensor. To overcome this intrinsic issue, we are developing a refinement procedure in which we combine information form the building footprint and the aerial point cloud to refine the covered façades by sampling the parametric representations of the corresponding planes.

Future efforts will regard the recognition of further urban elements. Small, short or thin objects require a high sampling resolution for being represented in the point cloud with a proper density to allow automatic recognition: this is the case of platforms, ramps, stairs, street lamps, traffic lights, and more. At a certain extent, "semantic" rules can help the identification even if the coverage in the input data is poor: e.g., we look for platform alongside the street floor. For urban furniture, sometimes the punctual location of items is available from on site surveys, and this could be a strong a-priori to drive the recognition of nearby points belonging to the element.

On the long term, our research will define a hierarchy of urban elements and arrange the city model accordingly, the ultimate goal being the definition of a semantic Level-of-Detail (LOD) for city representation, where urban elements can be represented at full resolution, or with a simplified geometric mock-up, or simply by a placeholder, or not being represented at all according to the user query to the semantic 3D model.

CRediT authorship contribution statement

Chiara Romanengo: Writing – review & editing, Writing – original draft, Validation, Software, Formal analysis, Data curation, Conceptualization. **Daniela Cabiddu:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Simone Pittaluga:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation. **Michela Mortara:** Writing – review & editing, Writing – original draft, Methodology, 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.

Acknowledgements

Funded by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project "RAISE - Robotics and AI for Socio-economic Empowerment" (ECS00000035).

C. Romanengo is part of RAISE Innovation Ecosystem.

We wish to thank the city and municipality of Matera and Catania, direct beneficiary of outcomes of the CTEM and UISH projects respectively, for their support during the 3D survey. The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation[5] (DGPF) [50]. The authors also thank their colleagues at CNR-IMATI Marianna Miola, Silvia Biasotti and Bianca Falcidieno for their support and fruitful discussions.

References

[1] F. Dembski, U. Wössner, M. Letzgus, M. Ruddat, C. Yamu, Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany, Sustainability 12 (6) (2020) http://dx.doi.org/10.3390/su12062307.

[2] Helsinki, The Kalasatama Digital Twins project, https://www.hel.fi/helsinki/en/administration/information/general/3d/3d.

[3] G. Schrotter, C. Hurzeler, The digital twin of the city of Zurich for urban planning, PFG 88 (2020) 99–112, http://dx.doi.org/10.1007/s41064-020-00092-2.

[4] Berlin, Berlin 3D - Download Portal, 2019, https://www.businesslocationcenter.de/en/economic-atlas/download-portal.

[5] Singapore, Virtual Singapore, https://www.nrf.gov.sg/programmes/virtual-singapore.

[6] G. Castelli, A. Cesta, M. Diez, M. Padula, P. Ravazzani, G. Rinaldi, S. Savazzi, M. Spagnuolo, L. Strambini, G. Tognola, et al., Urban intelligence: a modular, fully integrated, and evolving model for cities digital twinning, in: 2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life using ICT & IoT and AI, HONET-ICT, IEEE, 2019, pp. 033–037, http://dx.doi.org/10.1109/HONET.2019.8907962.

[7] G. Castelli, A. Cesta, M. Ciampi, R. De Benedictis, G. De Pietro, M. Diez, G. Felici, R. Malvezzi, B. Masini, R. Pellegrini, A. Scalas, G. Stecca, L. Strambini, G. Tognola, P. Ravazzani, E.F. Campana, Urban intelligence: Toward the digital twin of matera and catania, in: 2022 Workshop on Blockchain for Renewables Integration, BLORIN, 2022, pp. 132–137, http://dx.doi.org/10.1109/BLORIN54731.2022.10028437.

[8] F. Poux, C. Mattes, Z. Selman, L. Kobbelt, Automatic region-growing system for the segmentation of large point clouds, Autom. Constr. 138 (2022) 104250, http://dx.doi.org/10.1016/j.autcon.2022.104250.

[9] L. Li, J. Yao, J. Tu, X. Liu, Y. Li, L. Guo, Roof plane segmentation from airborne LiDAR data using hierarchical clustering and boundary relabeling, Remote Sens. 12 (9) (2020) http://dx.doi.org/10.3390/rs12091363.

[10] B. Xu, W. Jiang, J. Shan, J. Zhang, L. Li, Investigation on the weighted RANSAC approaches for building roof plane segmentation from LiDAR point clouds, Remote Sens. 8 (1) (2016) http://dx.doi.org/10.3390/rs8010005.

[11] C. Romanengo, D. Cabiddu, S. Pittaluga, M. Mortara, Semantic Segmentation of High-resolution Point Clouds Representing Urban Contexts, in: F. Banterle, G. Caggianese, N. Capece, U. Erra, K. Lupinetti, G. Manfredi (Eds.), Smart Tools and Applications in Graphics - Eurographics Italian Chapter Conference, The Eurographics Association, 2023, http://dx.doi.org/10.2312/stag.20231296.

[12] R. Wang, H.Y. Shangfeng Huang, Building3D: An urban-scale dataset and benchmarks for learning roof structures from point clouds, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023.

[13] P. Tang, D. Huber, B. Akinci, R. Lipman, A. Lytle, Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques, Automation in Construction 19 (7) (2010) 829–843, http://dx.doi.org/10.1016/j.autcon.2010.06.007.

[14] X. Wang, S. Ji, Roof plane segmentation from LiDAR point cloud Data Using Region expansion based L0 gradient minimization and graph cut, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 10101–10116, http://dx.doi.org/10.1109/JSTARS.2021.3113083.

[15] D. Kong, L. Xu, X. Li, S. Li, K-plane-based classification of airborne LiDAR data for accurate building roof measurement, IEEE Trans. Instrum. Meas. 63 (5) (2014) 1200–1214, http://dx.doi.org/10.1109/TIM.2013.2292310.

[16] M. Attene, G. Patanè, Hierarchical structure recovery of point-sampled surfaces, Comput. Graph. Forum 29 (6) (2010) 1905–1920, http://dx.doi.org/10.1111/j.1467-8659.2010.01658.x.

[17] D.-M. Yan, W. Wang, Y. Liu, Z. Yang, Variational mesh segmentation via quadric surface fitting, Comput. Aided Des. 44 (11) (2012) 1072–1082, http://dx.doi.org/10.1016/j.cad.2012.04.005.

[18] T. Le, Y. Duan, A primitive-based 3D segmentation algorithm for mechanical CAD models, Comput. Aided Geom. Design 52–53 (2017) 231–246, http://dx.doi.org/10.1016/j.cagd.2017.02.009.

[19] S. Xia, D. Chen, R. Wang, J. Li, X. Zhang, Geometric primitives in LiDAR point clouds: A review, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13 (2020) 685–707, http://dx.doi.org/10.1109/JSTARS.2020.2969119.

[20] R. Schnabel, R. Wahl, R. Klein, Efficient RANSAC for point-cloud shape detection, Comput. Graph. Forum 26 (2) (2007) 214–226, http://dx.doi.org/10.1111/j.1467-8659.2007.01016.x.

[21] F.A. Limberger, M.M. Oliveira, Real-time detection of planar regions in unorganized point clouds, Pattern Recognit. 48 (6) (2015) 2043–2053, http://dx.doi.org/10.1016/j.patcog.2014.12.020.

[22] M. Li, P. Wonka, L. Nan, Manhattan-world urban reconstruction from point clouds, in: ECCV, 2016, http://dx.doi.org/10.1007/978-3-319-46493-0_4.

[23] J.M. Coughlan, A.L. Yuille, The manhattan world assumption: Regularities in scene statistics which enable Bayesian inference, in: Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS '00, MIT Press, Cambridge, MA, USA, 2000, pp. 809–815.

[24] D. Chen, L. Zhang, P.T. Mathiopoulos, X. Huang, A methodology for automated segmentation and reconstruction of urban 3-D buildings from ALS point clouds, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (10) (2014) 4199–4217, http://dx.doi.org/10.1109/JSTARS.2014.2349003.

[25] C. Romanengo, A. Raffo, S. Biasotti, B. Falcidieno, Recognizing geometric primitives in 3D point clouds of mechanical CAD objects, Comput. Aided Des. 157 (2023) 103479, http://dx.doi.org/10.1016/j.cad.2023.103479.

[26] E. Maltezos, C. Ioannidis, Automatic extraction of building roof planes from airborne lidar data applying an extended 3D randomized hough transform, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. III-3 (2016) 209–216, http://dx.doi.org/10.5194/isprs-annals-III-3-209-2016.

[27] A. Nguyen, B. Le, 3D point cloud segmentation: A survey, in: 2013 6th IEEE Conference on Robotics, Automation and Mechatronics, RAM, 2013, pp. 225–230, http://dx.doi.org/10.1109/RAM.2013.6758588.

[28] F. Tarsha-Kurdi, T. Landes, P. Grussenmeyer, Hough-transform and extended RANSAC algorithms for automatic detection of 3D building roof planes from lidar data, in: ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, vol. XXXVI, Espoo, Finland, 2007, pp. 407–412, URL: https://shs.hal.science/halshs-00264843.

[29] A. Raffo, C. Romanengo, B. Falcidieno, S. Biasotti, Fitting and recognition of geometric primitives in segmented 3D point clouds using a localized voting procedure, Comput. Aided Geom. Design 97 (2022) 102123, http://dx.doi.org/10.1016/j.cagd.2022.102123.

[30] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, A. Markham, Randla-net: Efficient semantic segmentation of large-scale point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11108–11117.

[31] S.M. González-Collazo, N. Canedo-González, E. González, J. Balado, Semantic point cloud segmentation in urban environments with 1D convolutional neural networks, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2024) 205–211, http://dx.doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-205-2024, XLVIII-4/W9-2024.

[5] http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html

[32] T. Hackel, J.D. Wegner, K. Schindler, Fast semantic segmentation of 3D point clouds with strongly varying density, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences III-3 (2016) 177–184, http://dx.doi.org/10.5194/isprs-annals-III-3-177-2016.

[33] Q. Hu, B. Yang, S. Khalid, W. Xiao, A. Trigoni, A. Markham, Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 4975–4985, URL: https://api.semanticscholar.org/CorpusID:221516403.

[34] X. Roynard, J.-E. Deschaud, F. Goulette, Paris-lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification, Int. J. Robot. Res. 37 (6) (2018) 545–557, http://dx.doi.org/10.1177/0278364918767506.

[35] X. Sun, B. Guo, C. Li, N. Sun, Y. Wang, Y. Yao, Semantic segmentation and roof reconstruction of urban buildings based on LiDAR point clouds, ISPRS Int. J. Geo-Inf. 13 (1) (2024) http://dx.doi.org/10.3390/ijgi13010019.

[36] E.K. Dey, M. Awrangjeb, F.T. Kurdi, B. Stantic, Machine learning-based segmentation of aerial LiDAR point cloud data on building roof, European Journal of Remote Sensing 56 (1) (2023) 2210745, http://dx.doi.org/10.1080/22797254.2023.2210745.

[37] J. Niemeyer, F. Rottensteiner, U. Soergel, C. Heipke, Hierarchical higher order crf for the classification of airborne LIDAR point clouds in urban areas, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B3 (2016) 655–662, http://dx.doi.org/10.5194/isprs-archives-XLI-B3-655-2016, URL: https://isprs-archives.copernicus.org/articles/XLI-B3/655/2016/.

[38] R. Zhao, M. Pang, J. Wang, Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network, Int. J. Geogr. Inf. Sci. 32 (5) (2018) 960–979, http://dx.doi.org/10.1080/13658816.2018.1431840.

[39] ESRI, ESRI shapefile technical description – An ESRI white paper, 1998, URL: https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf.

[40] OpenStreetMap contributors, Planet dump retrieved from 2017, https://planet.osm.org, https://www.openstreetmap.org.

[41] E. Haines, Point in polygon strategies, in: Graphics Gems IV, Academic Press Professional, Inc., USA, 1994, pp. 24–46.

[42] W.R. Franklin, PNPOLY - Point inclusion in polygon test, 1970, URL: https://wrfranklin.org/Research/Short_Notes/pnpoly.html. Last update: March 2023.

[43] C. Jordan, Cours d'analyse de l'École polytechnique, vol. 1, Gauthier-Villars et fils, 1893.

[44] D. Cabiddu, M. Attene, Large mesh simplification for distributed environments, Comput. Graph. 51 (2015) 81–89, http://dx.doi.org/10.1016/j.cag.2015.05.015.

[45] M.C. Beltrametti, L. Robbiano, An algebraic approach to Hough transforms, J. Algebra 37 (2012) 669–681, http://dx.doi.org/10.1016/j.jalgebra.2012.09.012.

[46] UISH, UISH: Urban Intelligence Science Hub for City Network. Programma Operativo Complementare Città Metropolitane 2014–2020 - Ambito II - Progetti Pilota., http://www.diitet.cnr.it/pon-metro-uish/.

[47] A. Scalas, D. Cabiddu, M. Mortara, M. Spagnuolo, Potential of the geometric layer in urban digital twins, ISPRS Int. J. Geo-Inf. 11 (6) (2022) 343, http://dx.doi.org/10.3390/ijgi11060343.

[48] D. Cabiddu, M. Mortara, C. Romanengo, A. Scalas, A. Bellazzi, L. Belussi, L. Danza, M. Ghellere, 3D Feature Recognition for the Assessment of Buildings' Energy Efficiency, in: A. Serani, C. Leotardi (Eds.), BUILding a Digital Twin: Requirements, Methods, and Applications, BUILD-IT Workshop, National Research Council-Institute of Marine Engineering (CNR-INM), 2023, pp. 46–49, URL: http://inm.cnr.it/buildit2023/wp-content/uploads/2023/10/2023-BUILD-IT_book_of_abstracts.pdf.

[49] CTEM, House of emerging technologies of Matera'' (CTEMT) funded by the Ministry of Economic Development of Italy, CUP I14E2000002000, http://www.diitet.cnr.it/en/ctemt/.

[50] M. Cramer, The DGPF?Test on digital airborne camera evaluation overview and test design, Photogrammetrie ? Fernerkundung ? Geoinformation 2010 (2) (2010) 73–82, http://dx.doi.org/10.1127/1432?8364/2010/0041.

[51] F. Rottensteiner, G. Sohn, M. Gerke, J.D. Wegner, U. Breitkopf, J. Jung, Results of the ISPRS benchmark on urban object detection and 3D building reconstruction, ISPRS J. Photogramm. Remote Sens. 93 (2014) 256–271, http://dx.doi.org/10.1016/j.isprsjprs.2013.10.004.

[52] V. Spreckels, L. Syrek, A. Schlienkamp, DGPF?Project: Evaluation of digital photogrammetric camera systems stereoplotting, Photogrammetrie ? Fernerkundung ? Geoinformation 2010 (2) (2010) 117–130, http://dx.doi.org/10.1127/1432?8364/2010/0044.

[53] A. Scalas, D. Cabiddu, M. Mortara, S. Pittaluga, M. Spagnuolo, Mobile Laser Scanning of Challenging Urban Sites: a Case Study in Matera, in: F. Ponchio, R. Pintus (Eds.), Eurographics Workshop on Graphics and Cultural Heritage, The Eurographics Association, 2022, http://dx.doi.org/10.2312/gch.20221218.

[54] C.-W. Huang, T.-Y. Shih, On the complexity of point-in-polygon algorithms, Comput. Geosci. 23 (1) (1997) 109–118, http://dx.doi.org/10.1016/S0098-3004(96)00071-4.

[55] D. Defays, An efficient algorithm for a complete link method, Comput. J. 20 (4) (1977) 364–366, http://dx.doi.org/10.1093/comjnl/20.4.364.

Chiara Romanengo has a PhD in Mathematics and Applications and is currently a fixed-term researcher at IMATI-CNR Genova. Her research interests include curves and surfaces, algebraic geometry, features on 3D models, and geometric modelling. She investigates methods for the identification and recognition of characteristic parts on the surface of 3D models based on the Hough transform technique.

Daniela Cabiddu is researcher at IMATI-CNR Genova. Her current research focus is on Computer Graphics and Geometry Modelling applied to geoscience, fabrication and engineering. She has worked in several national and EU founded research projects dealing with digital representations of 3D domains, providing solutions to efficiently generate, encode and reuse high-resolution 3D models and possible embedded heterogeneous information. She gained expertise in spatial analysis, geographical information systems, and environmental monitoring methods applied to urban domains.

Simone Pittaluga is a researcher at CNR-IMATI. With a background in geology, his research interests lie in geostatistics and numerical and geometric modelling of the environment, including the planning and acquisition of topographic data using laser scanners. He has participated in various European and national projects related to modelling natural phenomena and implementing geostatistical methods. Additionally, he has also designed and developed applications that have reached commercial level.

Michela Mortara is a senior researcher at CNR-IMATI. Her research interest focusses on 3D geometric modelling and analysis for devising structural and semantic representations of objects, scenes and phenomena occurring in the 3D space. She has been involved in many international research projects dealing with shapes and semantics in a variety of application domains, including Bioinformatics, Cultural Heritage, Serious Games, Geoscience and Urban Intelligence.