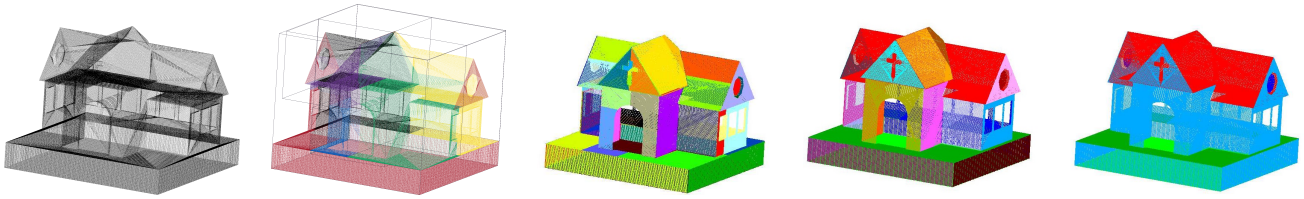
STAG: Smart Tools and Applications in Graphics (2023)
K. Lupinetti, F. Banterle, N. Capece, and U. Erra (Editors)

Semantic segmentation of high-resolution point clouds representing urban contexts

C. Romanengo[1] and D. Cabiddu[1†] and S. Pittaluga[1] and M. Mortara[1]

[1]CNR-IMATI Genova, Italy



(a) (b) (c) (d) (e)

Figure 1: *Segmentation Pipeline. (a) Input high-resolution 3D point cloud (thanks to[ZJ16]). (b) Binary space partitioning into chunks that can be processed. (c) RANSAC segmentation in simple geometric primitives. (d) Parameter estimation through the Hough transform and aggregation of segments belonging to the same primitives. (e) Semantic segmentation.*

Abstract

Point clouds are becoming an increasingly common digital representation of real-world objects, and they are particularly efficient when dealing with large-scale objects and/or when extremely high-resolution is required. The focus of our work is on the analysis, 3D feature extraction and semantic annotation of point clouds representing urban scenes, coming from various acquisition technologies, e.g., terrestrial (fixed or mobile) or aerial laser scanning or photogrammetry; the task is challenging, due to data dimensionality and noise.

In particular, we present a pipeline to segment high-resolution point clouds representing urban environments into geometric primitives; we focus on planes, cylinders and spheres, which are the main features of buildings (walls, roofs, arches, ...) and ground surfaces (streets, pavements, platforms), and identify the unique parameters of each instance. This paper focuses on the semantic segmentation of buildings, but the approach is currently being generalised to manage extended urban areas. Given a dense point cloud representing a specific building, we firstly apply a binary space partitioning method to obtain small enough sub-clouds that can be processed. Then, a combination of the well-known RANSAC algorithm and a recognition method based on the Hough transform (HT) is applied to each sub-cloud to obtain a semantic segmentation into salient elements, like façades, walls and roofs. The parameters of primitive instances are saved as metadata to document the structural element of buildings for further thematic analyses, e.g., energy efficiency.

We present a case study on the city of Catania, Italy, where two buildings of historical and artistic value have been digitized at very high resolution. Our approach is able to semantically segment these huge point clouds and it proves robust to uneven sampling density, input noise and outliers.

CCS Concepts

• ***Computing methodologies*** → *Point-based models; Shape analysis;* • ***Applied computing*** → *Architecture (buildings);*

1. Introduction

The continuous growth of urban areas urgently requires to address fundamental issues, such as healthy life and well-being, sustainability, resilience to climate change, as reported in the NextGenerationEU recovery plan [Nex]. Exploiting the new enabling technologies (i.e., modelling and simulations, sensing, Internet of Things, Artificial Intelligence) for the correct representation, understanding, and integration of urban phenomena, led to the development of urban Digital Twins (DTs) for the monitoring, prediction and

† Corresponding Author

DOI: 10.2312/stag.20231296

simulation of the state of the city in specific scenarios [CCD*19] [CCC*22]. This paradigm is becoming a widespread approach for the day-by-day administration and for the definition of long-term policies for the future of sustainable cities [DWL*20; Hel; SH20; Ber; Sin]. The DT of the city comprehends several components, which can be briefly summarised as follows: a 3D geometric model of the urban space, representing the morphology of the built and natural environment; a set of informative layers (e.g., road network, street types, position of traffic lights, weather forecast), representing the knowledge of the city involved in the processes that the DT should investigate (e.g., traffic conditions), and of course the algorithmic modules that, integrating information from different layers, monitor actual conditions, predict future states, simulate scenarios and visualise results.

3D models are typically used to communicate visual aspects, for example enabling remote viewing of places of interest. If this is the case, the main requirement is to have good-looking models whose geometric resolution represents a good trade-off between level of detail and rendering/interaction speed. In the context of urban DTs, the 3D model can provide direct access to information related to specific entities present in a scene. Indeed, many informative layers refer to precise spatial locations, so that portions of the 3D model (points, lines, areas, objects) can be identified as specific urban features (e.g., streets, parks, buildings) and be associated with thematic knowledge (park opening hour, building height, pavement material, and so on). In this view, the 3D model can integrate knowledge from the diverse thematic layers as far as such knowledge is location-based [SCMS22]. Linking information to geometric entities is called *part-based annotation*, and builds on a former *segmentation* step, where the whole geometric model is firstly decomposed into salient elements either manually or automatically. The identification of specific homogeneous parts is also known as feature recognition. The set of features of interest obviously depends on the context and objectives of the project.

In this paper, we present a segmentation and recognition pipeline for the urban context, aiming in particular to the semantic segmentation and documentation of buildings to support further analysis. For solar radiation and heat dispersion considerations, for instance, it is crucial to identify morphological attributes of buildings, such as exposed surface area and inclination and orientation of roofs. Therefore, we target the recognition of primitive shapes that best fit parts of buildings: planes in the first place, but also cylinders (e.g., arches, columns) and spheres (e.g., domes). Our method is able to identify geometry portions corresponding to instances of the primitives, providing the parametric form of each feature. Even if this contribution focus on single buildings, our research is wider and tackles extended urban areas; the method described here is currently being generalised to manage the level of a city district.

The challenges posed by the urban context are twofold: firstly, acquisition at geographic scale produces a large amount of data, which typically requires out-of-core segmentation and recognition algorithms. Secondly, different acquisition technologies might be deployed, resulting in varying resolution, precision and accuracy of data; thirdly, each acquisition methodology has drawbacks (e.g., occlusions, outliers and noise) and the acquisition of urban scenes necessarily 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.

We apply our pipeline to scenes represented as point clouds. On the one hand, both laser scanning and photogrammetry provide this data type; furthermore, it is suited to limit the memory allocation, which, for large scale models, might be critical. We firstly apply a binary space partitioning method to obtain small enough subclouds 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 to each segment to obtain a semantic segmentation into its main elements.

We present a case study on the city of Catania, Italy, where a large urban area has been surveyed in different modalities for the creation of the city DT, in the framework of a national project under development [UIS]. As far as this paper is concerned, we report on experiments on two historic buildings of the city centre, the "Palazzo degli Elefanti" and the "Palazzo dei Chierici", given as point clouds of 541 and 757 millions points, respectively. Our approach is able to segment huge point clouds and, thanks to the properties of the Hough transform, it prevents over-segmentation and is robust to input noise and outliers.

The rest of the paper is organised as follows. Section 2 overviews previous work in point cloud segmentation and fitting geometric primitives, with specific focus on the application to the analysis of urban scenarios, while Section 3 summarizes some notions to facilitate the comprehension of the approach. Section 4 describes our method for the building-wise semantic segmentation of high-resolution point clouds, performed by fitting primitives and computing their parameters. An ongoing generalisation of the method to manage wider urban areas rather than single buildings is presented in Section 5. Section 6 provides experimental results of our case study and demonstrates the effectiveness of our strategy. Final remarks (Section 7) and next steps conclude the paper.

2. Previous Works

3D point cloud segmentation aims at splitting points into distinct homogeneous regions, so that points in the same region share the same properties. Fitting primitives methods segment point clouds based on the homogeneous property of belonging to a certain primitive type having the same parameters. The challenges of point cloud segmentation are related to the input data quality, which often suffer of high redundancy, non-uniform sampling density, missing data, noise and outliers, plus the lack of an explicit structure in the data.

Over the past few decades, there has been a significant proliferation of algorithms developed for the purpose of breaking down digitalized point clouds or meshes into regions that can be approximated by primitive shapes from predefined sets [XCW*20]. According to a study by Kaiser in 2019 [KYB19], these approaches can be categorized into four main groups: stochastic methods, parameter space techniques, clustering approaches, and learning-based methods. The first group encompasses techniques like the RANSAC method [SWK07] and its various optimizations. The second family includes methods that rely on Hough-like voting and

parameter space clustering, as demonstrated in works such as Limberger et al.'s research [LO15]. The third category encompasses a wide range of clustering techniques. It can be further subdivided into three primary types: primitive-driven region growing, as seen in the work by Attene et al. [AP10]; automatic clustering and algorithms based on Lloyd's method, as demonstrated in Yan et al.'s work [YWLY12]; and primitive-oblivious segmentation, as explored by Le et al [LD17]. More recently, with the increasing popularity of deep learning techniques, many supervised fitting methods have been proposed for detecting multi-class primitives [LSD*19] [SLM*20], [YYM*21] [BPAT20] [LH19] [BBN*20] [MJB21].

Focusing on recent methods for analysing and classifying point clouds representing urban scenes, [PGE21] proposes a learning-based method that addresses the problem of segmenting and labeling point clouds in scenes with densely populated of Mechanical/Electrical/Plumbing (MEP) building systems. In order to deal with the increasing availability of acquired data, [PMSK22] introduces a region-growing-based system for the segmentation of large point clouds in planar regions. The method proposed in [WXW*23] reconstructs semantic building models by improving façades-level semantic 3D segmentation, exploiting both point cloud and images.

In our context, we deal with high-resolution point clouds representing building of different types subjected to different types of noise given by errors due to acquisition instruments. In addition, we are interested in providing semantic information about parts that compose a building. With this final goal in mind, we propose a strategy focused on a combination of the RANSAC approach and an HT-based recognition method. A analytic comparison of two algorithms for automatic roof planes detection from Lidar data is provided in [TLG07]. The RANSAC is able to provide a segmentation of the input point cloud in an efficient way, associating to each segment the type of primitive which corresponds to it. However, it is prone to over-segmentation. The Hough transform used for the recognition of surfaces as shown in [RRFB22] is able to recognise different types of primitives, associating the parameters (geometric descriptors) that uniquely identify each segment. Reasoning on the geometric descriptors, it is possible to group different segments that belong to the same primitive and to find semantic information about the parts that compose a building. In conclusion, combining the two approaches we are able to manage large scale point clouds and to extract important information on the buildings.

Other previous studies in urban contexts have employed either the RANSAC algorithm or the Hough transform. As a matter of example, both [LWN16] and [CZMH14] propose new methodologies based on RANSAC, but they are different from the proposed method. More in details, the goal of [LWN16] is to reconstruct scenes from point clouds assuming a regularity of the distribution of buildings, that is the Manhattan world assumption [CY00]. The reconstruction of buildings model is proposed also in [CZMH14] exploiting the RANSAC algorithm to segment the planar patches that constitute rooftops. Regarding the exploitation of the Hough transformation, [RRBF23] proposes a recognition method able to segment the input point cloud into geometric primitives of differente types (e.g., planes, cylinders, spheres, cones and tori), but it is focused for CAD objects. Finally, the study provided in [MI16] aims to extract building roof planes from airborne LIDAR data applying an extended Randomized Hough Transform, without incorporating semantic information.

To enhance the overall feasibility of running this approach on a standard personal computer independently of the size of the input dataset, we also apply a binary space partitioning beforehand, to lower the cardinality of the input point cloud.

3. Basic concepts

In this section, we provide some theoretical background on the Hough transform, along with an exploration of the specific geometric primitives chosen to facilitate our pursuit of the intended objective.

Hough transform. 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 Figure 2). This concept can be naturally extended to generic family $\mathcal{F} = \{\mathcal{S}_{\mathbf{a}}\}$ of curves or surfaces that depend on a set of parameters $\mathbf{a} = (a_1, \ldots, a_n)$ (see [BR12]). 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 $\mathbf{a} = (a_1, \ldots, a_n)$; the most voted n-uple corresponds to the most representative curve or surface for the profile.

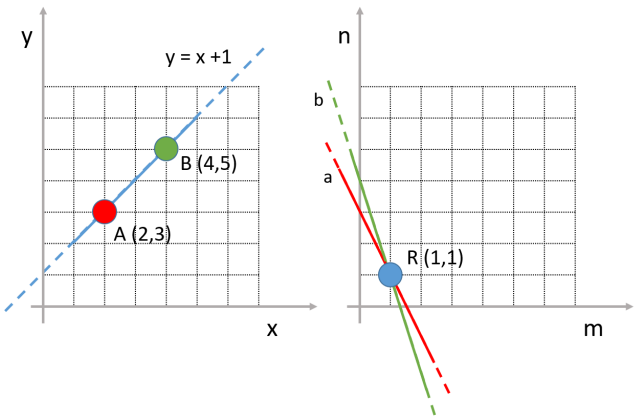


Figure 2: *The HT is based on the point-line duality: points A and B lie on a straight line. These lines correspond to lines in the parameter space that intersect in a single point R. This point uniquely identifies the coefficients in the equation of the original straight line.*

Dictionary of geometric primitives. We select a specific dictionary of geometric primitives suitable for our scope. Similarly to other methods specific for urban scenes, we select the families of

planes, cylinders and spheres (see Figure 3), since the main structural elements in the city can be identified by such types of surfaces. Exploiting the technique in [RRFB22], we reduce the number of parameters involved in the HT computation. Specifically, for the primitives considered in this work, the number of parameters is reduced to three in case of family of planes and one in case of cylinders and spheres. Starting from these parameters, it is possible to extract some geometric parameters that uniquely identify each primitive and that allow to make the primitives comparable each other.

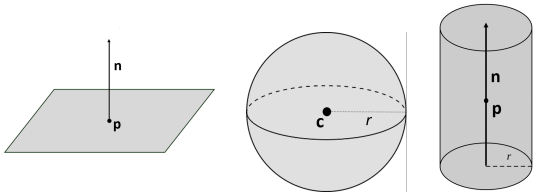


Figure 3: *Plane, sphere and cylinder primitives respectively, along with their attributes (geometric descriptors).*

In the following, we provide the geometric descriptors that we consider in the comparison step.

- *Planes.* A plane can be uniquely represented by its normal vector **n** and a point **p** lying on it; then, the geometric descriptor is represented by the vector $[\mathbf{n}, \mathbf{p}]$.
- *Spheres.* A sphere is uniquely identified by the coordinates of its center **c** and the value of the radius r. For this reason, the geometric descriptor for the sphere is the vector $[\mathbf{c}, r]$.
- *Cylinders.* A cylinder can be uniquely represented by its rotational axis **n**, a point **p** lying on it and the value of its radius r; then, the geometric descriptor is the vector $[\mathbf{n}, \mathbf{p}, r]$.

4. Pipeline Overview

The input of our semantic segmentation pipeline is a dense high-resolution point cloud representing the 3D morphology of a building (see Section 5 for a possible generalization of the approach to wider areas). With no loss of generality, we assume the input point cloud to be available either as a single or a collection of LAS files, the standard file format designed for the interchange and archiving of LiDAR point cloud data. However, the point cloud might come or not (or not only) from a LiDAR campaign, and even so, we don't assume points to be classified.

The approach works as follows (see Figure 1). First, the input point cloud is partitioned into smaller chunks by performing an out-of-core binary space partitioning [CA15]. Then, a combination of the RANSAC approach [SWK07] and the Hough transform is used to semantically segment each chunk into planes which describe the main elements of buildings, such as façades, walls and roofs, and pavement.

The following section describe the two steps in details.

4.1. Binary space partitioning

Depending on the origin of the input dataset (i.e. acquisition technique), the resolution of the input point clouds can be too high to be efficiently processed. Thus, a binary space partitioning (BSP) method is applied to split it into smaller chunks, each of them small enough to be processed by an ordinary machine. Our implementation is based on the out-of-core partitioning approach described in [CA15] and guarantees the possibility to split arbitrary large clouds into smaller chunks. The desired size of the generated chunks (in terms of maximum number of points) is user-defined and provided as an input parameter to the binary space partitioning algorithm, based on the performance of the machine doing the processing.

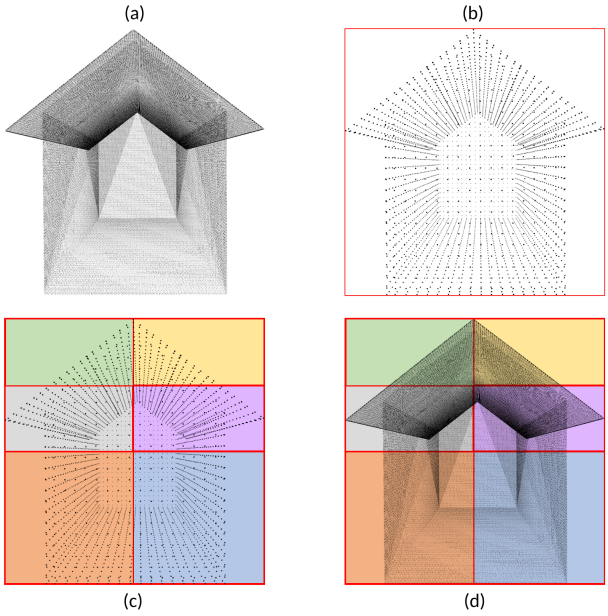


Figure 4: *Binary Space Partitioning. (a) Input point cloud. (b) Bounding Box and downsampling of the input cloud. (c) BSP computed on the downsampled set. (d) Final BSP.*

Figure 4 shows how the partitioning algorithm works. To begin, the point cloud's bounding box, denoted as $\mathcal{B}(\mathcal{P})$, is determined by scanning through all the point coordinates in $\mathcal{P}$ (Figure 4b). Concurrently, a representative vertex down-sampling $\mathcal{DS}(\mathcal{P})$ is generated by randomly selecting one vertex every 1000 in $\mathcal{P}$. Starting from $\mathcal{B}(\mathcal{P})$, an in-core binary space partition is constructed through an iterative process of subdividing the cell containing the largest number of points of $\mathcal{DS}(\mathcal{P})$. Each cell is split along its longest side. The root of the binary space partition corresponds to the entire downsampled cloud $\mathcal{DS}(\mathcal{P})$. During each subdivision, each vertex within the parent cell is assigned to one of the two offspring cells based on its spatial position. If a vertex lands precisely on the dividing plane, it is assigned to the cell with the lowest lexicographical barycenter. This process continues until the number of vertices allocated to each BSP cell is at most equal to a predefined threshold (Figure 4c). Once the BSP structure is established as described above, the task of assigning all points in $\mathcal{P}$ to their respective BSP cells begins. Points are processed individually and assigned based

on their spatial positioning (Figure 4d) and since the cardinality of the input cloud is much higher, this segmentation is done out-of-core.

The output is a set of *num_files* files, each encoding the set of points in a single chunk. For our convenience, each chunk $\mathcal{P}_i$ is saved as a *.xyz* file, *cell_i.xyz* where $i = 0, \dots, num\_files - 1$, encoding the point coordinates.

4.2. Geometric primitive fitting and parameter definition

After the BSP phase, the fitting of the selected primitive families starts. We point out that the space partitioning enables 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.

Firstly, we apply a RANSAC classification [SWK07], that is an automatic algorithm to detect basic shapes in unorganized point clouds, to each sub-cloud $\mathcal{P}_i$ returned by the binary space partitioning (saved in the *cell_i.xyz* file). 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 the 0.5% of the cardinality of $\mathcal{P}_i$ and we select three types of primitives (planes, cylinders and spheres). The result is a collection of subsets of points belonging to the same primitive, saved in a *.txt* file, whose name identifies the type of primitive.

Note that, the combination of the partitioning step in Section 4.1 and the tendency of the RANSAC algorithm to oversegment the point clouds (see, for example, [LD17]) can 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 presented in [RRFB22]. That approach is designed to recognise geometric primitives and to associate the corresponding parameters that uniquely identify each primitive in point clouds representing CAD objects; it can be naturally extended to the urban context, since the main urban elements can be identified by the same types of surfaces. The primitive type is provided in the name of the *.txt* file containing each segment, so this information is used by the HT-based recognition method to select the family of primitives to be used. In some cases, especially in presence of high noise, the RANSAC algorithm fails in associating the correct primitive type. Following the strategy described in [RRFB22], we enrich the recognition procedure with the computation of an approximation error, that allows us 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.

The parameters obtained in this step are used to define geometric descriptors (see Section 3) that uniquely characterize each segment. On the basis of these descriptors, once all chunks $\mathcal{P}_i$ are processed by the RANSAC, it is also possible to aggregate segments that belong to the same primitive or to find relationships among the recognised parts (e.g., parallel planes). An example is shown in Figure 5. To do this, we use a hierarchical clustering approach, the complete linkage, to compare clusters and build a dendrogram in a way similar to [RRFB22].

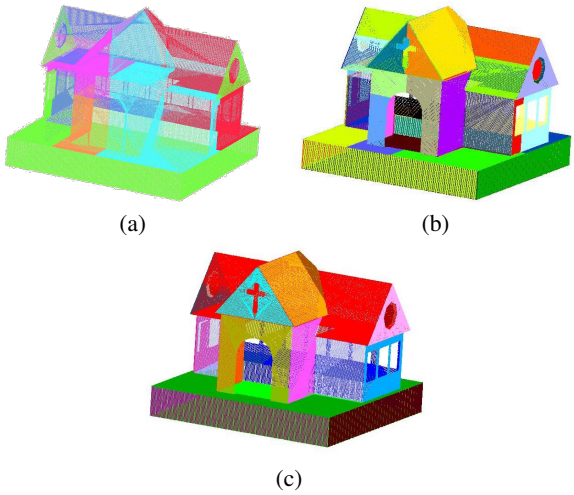


Figure 5: *Segmentation and primitive fitting. (a) The set of chunks $\mathcal{P}_i$, with $i = 0, \dots, 5$. (b) Result of the RANSAC segmentation applied to each $\mathcal{P}_i$. (c) Result of the aggregation of segments belonging to the same planes after the recognition step.*

More in details, the aggregation starts assigning every single segment to a cluster, and then iteratively merging clusters that are the closest with respect to the following map

$$D(C_h, C_j) := \max_{\alpha_k \in C_h, \alpha_l \in C_j} d(\alpha_k, \alpha_l),$$

In this map, C_h, C_j is a given pair of clusters and d is a chosen distance or dissimilarity. Table 1 shows the distances considered in this work, for each type of primitive, using the notation introduced in Section 3. Note that if $d(\alpha_1, \alpha_2) = 0$ then the primitives α_1 and α_2 are equal with respect to the selected criterion.

Table 1: *Distances considered in our work, following the notation introduced in Section 3.*

Primitive type	$d(\alpha_1, \alpha_2)$	Criterion
Planes	$\|\|\mathbf{n}_1 \times \mathbf{n}_2\|\|_2$	parallel
	$\|\mathbf{n}_1 \cdot (\mathbf{p}_1 - \mathbf{p}_2)\|$	incident
Cylinders	$\|r_1 - r_2\|$	equal radii
	$\|\|\mathbf{n}_1 \times \mathbf{n}_2\|\|_2$	parallel rotational axes
	$\|\|\mathbf{n}_1 \times (\mathbf{p}_1 - \mathbf{p}_2)\|\|_2$	incident rotational axes
Spheres	$\|r_1 - r_2\|$	equal radius
	$\|\|\mathbf{c}_1 - \mathbf{c}_2\|\|_2$	equal centers

In our context, we are interested in whether two segments lie on the same primitive, so for each type of primitive we use the sum of all distances shown in Table 1 associated to it. As an example, we check whether two segments lie on the same cylinder by using the metric $d(\alpha_1, \alpha_2) := |r_1 - r_2| + ||\mathbf{a}_1 \times \mathbf{a}_2||_2 + (\mathbf{p}_1 - \mathbf{p}_2)||_2$.

Finally, the geometric descriptors can be used to find semantic

information about the processed point cloud. Specifically, in case of planes we can distinguish them into façades, walls and roofs, and pavement, as shown in Figure 6.

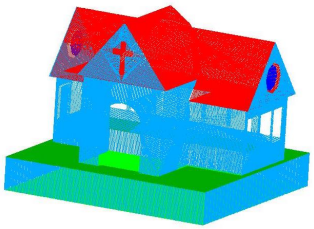


Figure 6: *Semantic segmentation of a church: the points corresponding to the façades in light blue, to the roofs in red and to the pavement in green.*

By analysing the components of the normals **n** of each plane, we can distinguish them into vertical, horizontal or oblique planes. In this way, we can grouped:

- the vertical planes, annotated as façades;
- the oblique planes, labeled as roofs; the normal also indicates the orientation and pitch of each roof;
- the horizontal planes, distinguishing them between into pavement and roofs based on the height in which they are located.

5. Pipeline generalisation

In real applications, input datasets often come from acquisitions of large urban areas, that is, input point clouds generally include more than one building. Our method works properly in the general case, but it is inefficient when the user wishes to segment one or just a few buildings of interest within a much wider area.

To optimize the execution of our method in that case, we exploit prior knowledge in the form of the 2D footprints of the desired buildings, available in online repositories (e.g., OpenStreetMap [Ope17]), to split the input point cloud building-wise. With no loss of generality, we consider 2D footprints encoded as standard ESRI Shapefiles [ESR98]. The building-wise classification 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 [Hai94].

Our implementation extends the point-in-polygon method introduced by W. Randolph Franklin [Fra]. It relies on the Jordan Curve Theorem [Jor93], which asserts that a point p resides inside a polygon if the number of crossings of an half line starting at p in any arbitrary direction is odd. To determine whether a given point p is inside a polygon $\mathcal{S}$, which is represented as an ordered list of vertices, we begin by checking if the projection $\hat{p}$ of point p onto the plane of $\mathcal{S}$ falls within the axis-aligned bounding box of $\mathcal{S}$. If it doesn't, the point lies outside $\mathcal{S}$. Otherwise, we employ a ray-casting approach, where we extend a ray from point p in arbitrary direction (e.g., horizontally to the left) and count the intersections between this ray and the edges of the polygon. If the count is zero or even, point p is outside the polygon; otherwise, it is inside (refer

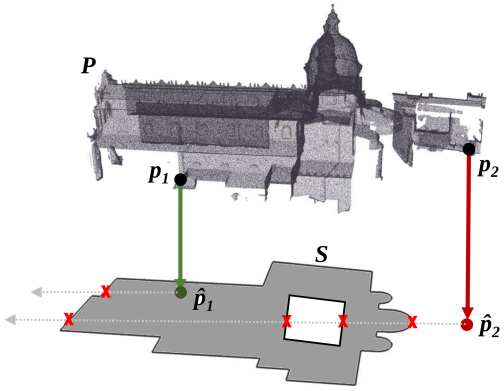


Figure 7: *Point-in-Polygon. The input includes both the point cloud $\mathcal{P}$ and a polygonal representation $\mathcal{S}$ of the footprint of the building of interest. Each point p in $\mathcal{P}$ is projected on the plane of $\mathcal{S}$ and the Jordan Curve Theorem is applied to assess if the point lies within the footprint.*

to Figure 7). Our implementation accommodates both simple polygons and polygons with multiple boundaries, including holes; this is crucial for urban environments, since many buildings exhibit inner courts. If the shapefile contains multiple disjoint polygons (e.g., for a subset of buildings), we apply the same approach concurrently to the entire set of buildings.

In the same way, 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.

6. Case Study

The case study comes from a dataset acquired in the framework of the UISH project [UIS], which aims at developing Digital Twins of Catania, Italy, to improve city monitoring and governance. 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 the city.

Various dataset were acquired, using different acquisition techniques; in particular, an extended area of the city centre (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) and in particular Palazzo degli Elefanti (Elephants' Palace) and Palazzo dei Chierici (Table 2). 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 (see Figure 8).

Referring to the single buildings, each input point cloud has been partitioned into sub-clouds of maximum size of 10M points each,

Figure 8: *Left: extended area showing the aerial planning for aerophotogrammetric acquisition; middle: digitisation of the buildings of interest with fixed laser stations; right: partial scan. Image courtesy of START4.0.*

Input	# Points (millions)
Palazzo degli Elefanti	541
Palazzo dei Chierici	757

Table 2: *Size of the input dataset, in terms of number of points.*

sufficient to be processed by our ordinary PC equipped with an Intel Core i9 processor (at 3.6 GHz) and Windows 11 64 bits operating system.

Figure 9 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 Figure 9(a)). The last step, that is the aggregation of segments belonging to the same primitives (see Figure 9(b)). Finally, Figure 9(c) 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 Figure 11, this dataset also presents cylinders and spheres within the building, on the arcades of the cloister.

Figure 10 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 Figure 10(a)) that are then aggregated (see Figure 10(b)). Finally, in Figure 10(c) 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.

As Figures 9 and 10 show, our pipeline returned a semantic segmentation of both buildings with success. Just in a few cases, the classification given by RANSAC failed, but, thanks to the following HT-based recognition method, the misclassifications is automatically corrected. An example of a planar segment missclassified as cylinder is shown in Figure 12 from two different perspectives. The first one (Figure 12(a)) highlights that the segment presents some missing parts, while the second one shows the presence of noise.

In our experimental configuration, 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. Such a value has been empirically chosen and works properly for our case study. However, it is important to note that this constraint does have limitations in recognizing certain areas within the input point clouds, especially in cases where the cloud's resolution is lower due to the use of multiple acquisition techniques. An instance of such a limitation is depicted in Figure 13, where only a partial detection of the roof occurs, due to the fact that roof and facades were acquired using photogrammetry and terrestrial laser scanning respectively, resulting in different resolutions.

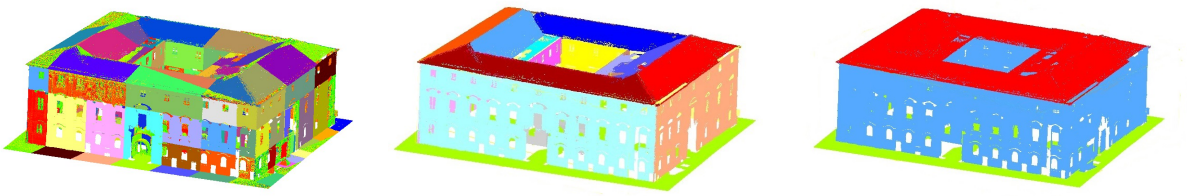
7. Discussion and conclusions

In this paper we have proposed a new method for segmenting high-resolution point clouds representing urban 3D scenarios into geometric primitives to detect some main features of the buildings, such as façades, roofs and arcades. Our approach is able to segment huge point clouds thanks to the application of an out-of-core partitioning, avoids over-segmentation 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 methods 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 could in principle handle larger point clouds, like the extended area of Catania, and be extended to recognise features out of buildings (streets, pavements, ramps, streetlamps, and so on). We are currently tackling this development, deploying a subdivision of the city into blocks or districts to apply a parallel computation to decrease computation time, which is, however, far to be real time.

At the current status, the proposed pipeline focuses on planes, cylinders and spheres; it enables the possibility to detect building features fitting such primitives. Future work will be addressed to extend the number of geometric primitives to be fitted, in order to enable the possibility to detect some additional features, such as cones that would be useful to detect and label some specific types of roofs.

Acknowledgements

This work has been 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

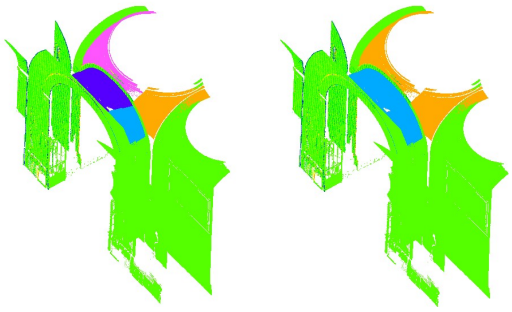


(a) (b) (c)

Figure 9: *Palazzo degli Elefanti: in (a) the resulting preliminary segmentation, where different colors 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.*

(a) (b) (c)

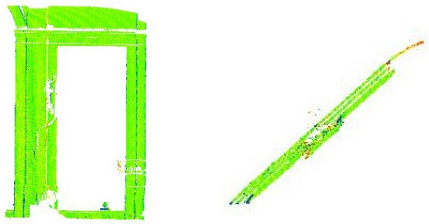
Figure 10: *Palazzo dei Chierici: in (a) the resulting preliminary segmentation, where different colors correspond to different segments; in (b) the segments that belong to the same plane are grouped; in (c) the planes classified as façade, roof and floor are grouped.*



(a) (b)

Figure 11: *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).*

for Socio-economic Empowerment" (ECS00000035). We wish to thank the city and municipality of Catania, direct beneficiary of outcomes of the UISH project, for their support during the 3D survey. The authors also thank their colleagues at CNR-IMATI Silvia Biasotti and Bianca Falcidieno for their support and fruitful discussions.



(a) (b)

Figure 12: *Example of planar segment missclassified as cylinder by RANSAC, due to missing data and noise, view from two different perspectives: (a) front and (b) top*

References

[AP10] ATTENE, MARCO and PATANÈ, GIUSEPPE. "Hierarchical Structure Recovery of Point-Sampled Surfaces". *Computer Graphics Forum* 29.6 (2010), 1905–1920. DOI: https://doi.org/10.1111/j.1467-8659.2010.01658.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2010.01658.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2010.01658.x 3.

[BBN*20] BIRDAL, TOLGA, BUSAM, BENJAMIN, NAVAB, NASSIR, et al. "Generic Primitive Detection in Point Clouds Using Novel Minimal Quadric Fits". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.6 (2020), 1333–1347. DOI: 10.1109/TPAMI.2019.2900309 3.

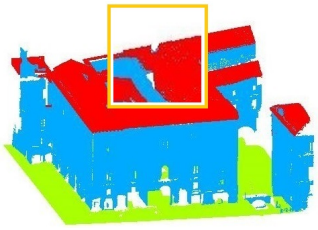


Figure 13: *Focus on a missing part of the roof in Palazzo dei Chierici.*

[Ber] BERLIN. *Berlin 3D - Download Portal*. https://www.businesslocationcenter.de/en/economic-atlas/download-portal 2.

[BPAT20] BERGAMASCO, FILIPPO, PISTELLATO, MARA, ALBARELLI, ANDREA, and TORSELLO, ANDREA. "Cylinders extraction in non-oriented point clouds as a clustering problem". *Pattern Recognition* 107 (2020), 107443 3.

[BR12] BELTRAMETTI, MAURO C and ROBBIANO, LORENZO. "An algebraic approach to Hough transforms". *J. of Algebra* 37 (2012), 669–681 3.

[CA15] CABIDDU, DANIELA and ATTENE, MARCO. "Large mesh simplification for distributed environments". *Computers & Graphics* 51 (2015). International Conference Shape Modeling International, 81–89. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2015.05.015. URL: https://www.sciencedirect.com/science/article/pii/S009784931500062X 4.

[CCC*22] CASTELLI, GIORDANA, CESTA, AMEDEO, CIAMPI, MARIO, et al. "Urban Intelligence: Toward the Digital Twin of Matera and Catania". *2022 Workshop on Blockchain for Renewables Integration (BLORIN)*. 2022, 132–137. DOI: 10.1109/BLORIN54731.2022.10028437 2.

[CCD*19] CASTELLI, GIORDANA, CESTA, AMEDEO, DIEZ, MATTEO, et al. "Urban intelligence: a modular, fully integrated, and evolving model for cities digital twinning". *2019 IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT)*. IEEE. 2019, 033–037 2.

[CY00] COUGHLAN, JAMES M. and YUILLE, A. L. "The Manhattan World Assumption: Regularities in Scene Statistics Which Enable Bayesian Inference". *Proceedings of the 13th International Conference on Neural Information Processing Systems*. NIPS'00. Denver, CO: MIT Press, 2000, 809–815 3.

[CZMH14] CHEN, DONG, ZHANG, LIQIANG, MATHIOPOULOS, P. TAKIS, and HUANG, XIANFENG. "A Methodology for Automated Segmentation and Reconstruction of Urban 3-D Buildings from ALS Point Clouds". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.10 (2014), 4199–4217. DOI: 10.1109/JSTARS.2014.2349003 3.

[DWL*20] DEMBSKI, FABIAN, WÖSSNER, UWE, LETZGUS, MIKE, et al. "Urban Digital Twins for Smart Cities and Citizens: The Case Study of Herrenberg, Germany". *Sustainability* 12.6 (2020). ISSN: 2071-1050. DOI: 10.3390/su12062307. URL: https://www.mdpi.com/2071-1050/12/6/2307 2.

[ESR98] 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 6.

[Fra] FRANKLIN, WM. RANDOLPH. *Libgenua: Ptinpoly.h*. URL: https://www.larosterna.com/documentation/apidoc/genua/apidoc/html/ptinpoly_8h_source.xhtml 6.

[Hai94] HAINES, ERIC. "Point in Polygon Strategies". *Graphics Gems IV*. USA: Academic Press Professional, Inc., 1994, 24–46. ISBN: 0123361559 6.

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

[Jor93] JORDAN, CAMILLE. *Cours d'analyse de l'École polytechnique*. Vol. 1. Gauthier-Villars et fils, 1893 6.

[KYB19] KAISER, ADRIEN, YBANEZ ZEPEDA, JOSE ALONSO, and BOUBEKEUR, TAMY. "A Survey of Simple Geometric Primitives Detection Methods for Captured 3D Data". *Computer Graphics Forum* 38.1 (2019), 167–196 2.

[LD17] LE, TRUC and DUAN, YE. "A primitive-based 3D segmentation algorithm for mechanical CAD models". *Computer Aided Geometric Design* 52-53 (2017), 231–246 3, 5.

[LH19] LIU, CHANG and HU, WEIDUO. "Real-time geometric fitting and pose estimation for surface of revolution". *Pattern Recognition* 85 (2019), 90–108 3.

[LO15] LIMBERGER, FREDERICO A. and OLIVEIRA, MANUEL M. "Real-time detection of planar regions in unorganized point clouds". *Pattern Recognition* 48.6 (2015), 2043–2053. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2014.12.020. URL: https://www.sciencedirect.com/science/article/pii/S0031320315000072 3.

[LSD*19] LI, LINGXIAO, SUNG, MINHYUK, DUBROVINA, ANASTASIA, et al. "Supervised Fitting of Geometric Primitives to 3D Point Clouds". *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, 2647–2655. DOI: 10.1109/CVPR.2019.00276 3.

[LWN16] LI, MINGLEI, WONKA, PETER, and NAN, LIANGLIANG. "Manhattan-world Urban Reconstruction from Point Clouds". *ECCV*. 2016 3.

[MI16] MALTEZOS, E. and IOANNIDIS, C. "AUTOMATIC EXTRACTION OF BUILDING ROOF PLANES FROM AIRBORNE LIDAR DATA APPLYING AN EXTENDED 3D RANDOMIZED HOUGH TRANSFORM". *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* III-3 (2016), 209–216. DOI: 10.5194/isprs-annals-III-3-209-2016. URL: https://isprs-annals.copernicus.org/articles/III-3/209/2016/ 3.

[MJB21] MARKOVIC, VELJKO, JAKOVLJEVIC, ZIVANA, and BUDAK, IGOR. "Automatic recognition of cylinders and planes from unstructured point clouds". *The Visual Computer* (2021), 1–24 3.

[Nex] NEXTGENERATIONEU. *NextGenerationEU*. https://next-generation-eu.europa.eu/index_en 1.

[Ope17] OPENSTREETMAP CONTRIBUTORS. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017 6.

[PGE21] PEREZ-PEREZ, YERITZA, GOLPARVAR-FARD, MANI, and EL-RAYES, KHALED. "Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment". *Automation in Construction* 125 (2021), 103584 3.

[PMSK22] POUX, F., MATTES, C., SELMAN, Z., and KOBBELT, L. "Automatic region-growing system for the segmentation of large point clouds". *Automation in Construction* 138 (2022), 104250 3.

[RRBF23] ROMANENGO, CHIARA, RAFFO, ANDREA, BIASOTTI, SILVIA, and FALCIDIENO, BIANCA. "Recognizing geometric primitives in 3D point clouds of mechanical CAD objects". *Computer-Aided Design* 157 (2023), 103479 3.

[RRFB22] RAFFO, ANDREA, ROMANENGO, CHIARA, FALCIDIENO, BIANCA, and BIASOTTI, SILVIA. "Fitting and recognition of geometric primitives in segmented 3D point clouds using a localized voting procedure". *Computer Aided Geometric Design* 97 (2022), 102123. ISSN: 0167-8396. DOI: https://doi.org/10.1016/j.cagd.2022.102123. URL: https://www.sciencedirect.com/science/article/pii/S0167839622000590 3–5.

[SCMS22] SCALAS, ANDREAS, CABIDDU, DANIELA, MORTARA, MICHELA, and SPAGNUOLO, MICHELA. "Potential of the geometric layer in urban digital twins". *ISPRS International Journal of Geo-Information* 11.6 (2022), 343 2.

[SH20] SCHROTTER, G. and HʹURZELER, C. "The Digital Twin of the City of Zurich for Urban Planning." *PFG* 88 (2020), 99–112. DOI: https://doi.org/10.1007/s41064-020-00092-2 2.

[Sin] SINGAPORE. *Virtual Singapore*. https://www.nrf.gov.sg/programmes/virtual-singapore 2.

[SLM*20] SHARMA, GOPAL, LIU, DIFAN, MAJI, SUBHRANSU, et al. "ParSeNet: A Parametric Surface Fitting Network for 3D Point Clouds". *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII*. Glasgow, United Kingdom: Springer-Verlag, 2020, 261–276. DOI: 10.1007/978-3-030-58571-6_16 3.

[SWK07] SCHNABEL, R., WAHL, R., and KLEIN, R. "Efficient RANSAC for Point-Cloud Shape Detection". *Computer Graphics Forum* 26.2 (2007), 214–226. DOI: https://doi.org/10.1111/j.1467-8659.2007.01016.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2007.01016.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2007.01016.x 2, 4, 5.

[TLG07] TARSHA-KURDI, FAYEZ, LANDES, TANIA, and GRUSSENMEYER, PIERRE. "Hough-Transform and Extended RANSAC Algorithms for Automatic Detection of 3D Building Roof Planes from Lidar Data". *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*. Vol. XXXVI. Espoo, Finland, Sept. 2007, 407–412. URL: https://shs.hal.science/halshs-00264843 3.

[UIS] 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/ 2, 6.

[WXW*23] WYSOCKI, O., XIA, Y., WYSOCKI, M., et al. "Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks". *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE Computer Society, 2023, 6548–6558 3.

[XCW*20] XIA, SHAOBO, CHEN, DONG, WANG, RUISHENG, et al. "Geometric Primitives in LiDAR Point Clouds: A Review". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020), 685–707 2.

[YWLY12] YAN, DONG-MING, WANG, WENPING, LIU, YANG, and YANG, ZHOUWANG. "Variational mesh segmentation via quadric surface fitting". *Computer-Aided Design* 44.11 (2012), 1072–1082. ISSN: 0010-4485. DOI: https://doi.org/10.1016/j.cad.2012.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0010448512000887 3.

[YYM*21] YAN, SIMING, YANG, ZHENPEI, MA, CHONGYANG, et al. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, 2753–2762. DOI: 10.1109/ICCV48922.2021.00275 3.

[ZJ16] ZHOU, QINGNAN and JACOBSON, ALEC. *Thingi10K: A Dataset of 10,000 3D-Printing Models*. cite arxiv:1605.04797. 2016. URL: http://arxiv.org/abs/1605.04797 1.