Article

Segmentation of Façades from Urban 3D Point Clouds Using Geometrical and Morphological Attribute-Based Operators

Andrés Serna *, Beatriz Marcotegui and Jorge Hernández

MINES ParisTech, CMM—Centre de Morphologie Mathématique, 35 rue St Honoré, Fontainebleau 77305, France; beatriz.marcotegui@mines-paristech.fr (B.M); jorge.hernandez@mines-paristech.fr (J.H.)
* Correspondence: andres.serna_morales@mines-paristech.fr; Tel.: +57-777-491-896

Abstract: 3D building segmentation is an important research issue in the remote sensing community with relevant applications to urban modeling, cloud-to-cloud and cloud-to-model registration, 3D cartography, virtual reality, cultural heritage documentation, among others. In this paper, we propose automatic, parametric and robust approaches to segment façades from 3D point clouds. Processing is carried out using elevation images and 3D decomposition, and the final result can be reprojected onto the 3D point cloud for visualization or evaluation purposes. Our methods are based on geometrical and geodesic constraints. Parameters are related to urban and architectural constraints. Thus, they can be set up to manage façades of any height, length and elongation. We propose two methods based on façade marker extraction and a third method without markers based on the maximal elongation image. This work is developed in the framework of TerraMobilita project. The performance of our methods is proved in our experiments on TerraMobilita databases using 2D and 3D ground truth annotations.

Keywords: 3D urban analysis; laser scanning; segmentation; façades; mathematical morphology; attribute-based operators; geodesic elongation

1. Introduction

Building segmentation can be defined as the process of separating buildings from other objects such as natural and artificial ground, vegetation and urban objects. The first researches on automatic building extraction began in the 1980s. They used aerial imagery and focused on the extraction of high-level 2D and 3D primitives from stereo images. One of the main drawbacks of these methods is that some false positives need to be filtered out since some linear primitives may not correspond to meaningful structures. In computer vision, elevation images were introduced as data structures allowing direct access to 3D geometric features. First elevation images were mostly acquired from small objects and scenes, using active systems. During the 1990s, aerial laser scanning (ALS) became widely available, so elevation images of huge scenes and cities became possible. Accuracy and point density of those systems are in constant progress. More recently, new acquisition systems such as terrestrial (TLS) and mobile laser scanning (MLS) have been developed, adding not only greater geometrical accuracy and extremely high density, but also façade scans, not visible from ALS [1].

Although the processing of 3D urban data has been underway for many years, façade segmentation is still an open problem. Several contributions on this domain are proposed in this paper.

Our processing assumes that ground has been already segmented. In our works we have used the approach already published in [2,3]. Once the ground is segmented, all remaining structures are

considered as façades and objects. Using TLS and MLS data, only building front parts are visible, as shown in Figure 1. It is noteworthy that façades constitute the highest and longest vertical entities in the urban scene.
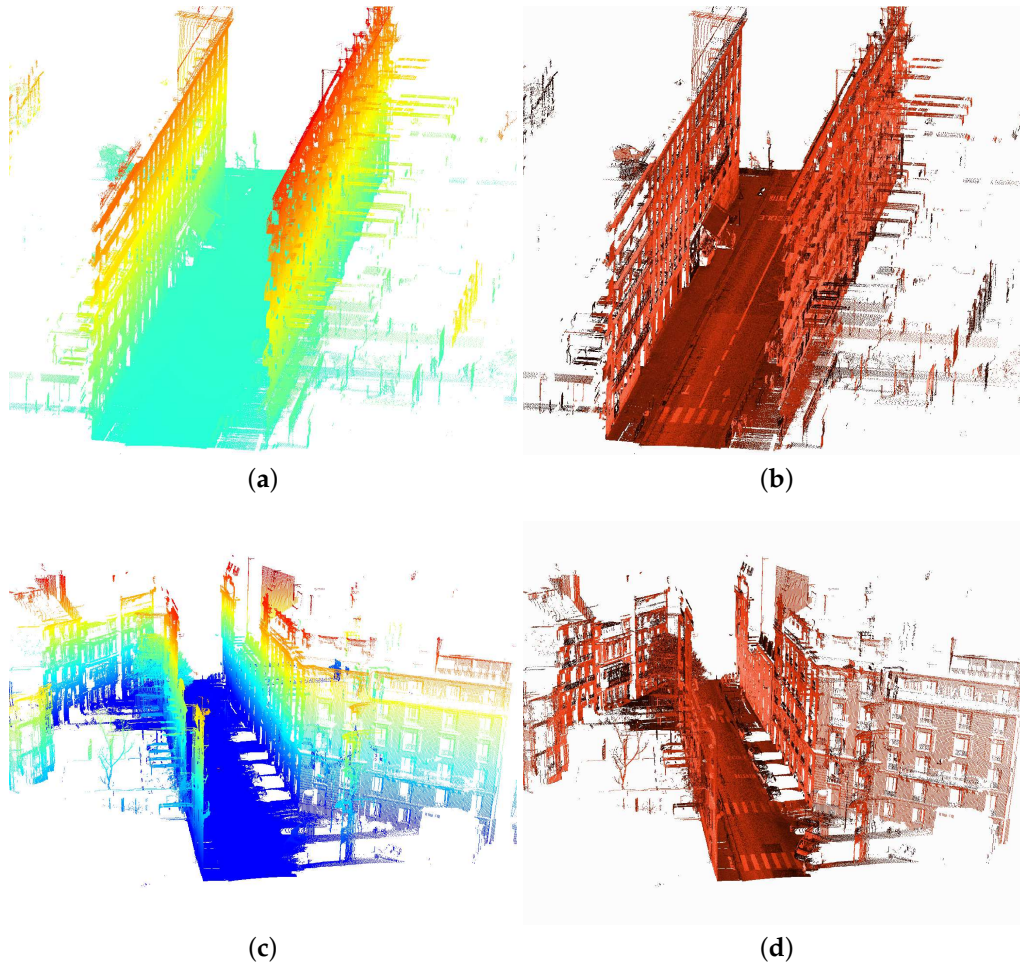

Figure 1. 3D point clouds from two test sites in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. Note that façades constitute the highest and longest vertical entities in the urban scene. (**a**) Point cloud colored by the Z coordinate; (**b**) Point cloud colored by the Intensity; (**c**) Point cloud colored by the Z coordinate; (**d**) Point cloud colored by the Intensity.

This work has been developed in the framework of TerraMobilita project [4] and is organized as follows. Section 2 reviews related works in the state of the art. Sections 4 and 5 introduce two different approaches to segment façades: with and without markers, respectively. Section 6 describes a method to segment city blocks taking advantage of the façade segmentation result. Finally, Section 8 concludes this work.

2. Related Work

Goulette *et al.* [5] develop a MLS system, called LARA-3D, which acquires and segments in real-time ground, façades and objects. Ground and façades are detected fitting horizontal and vertical planes, then remaining points are considered as objects. In a similar way, Boulaassal *et al.* [6] segment building façades using the RANSAC algorithm on TLS data. In general, approaches based on plane extraction are proved to be simple, fast and useful as input for high-level approaches devoted to create accurate geometric models. However, their main drawback is that plane extraction may fail

when ground and façades are not flat enough. Moreover, façades may be over- and under-segmented due to architectural details such as balconies.

Bab-Hadiashar and Gheissari [7] propose a method to segment planar and curved surfaces in range images. Their method consists of selecting the appropriate parametric model that minimizes strain energy of fitted surfaces. The authors applied their methodology to indoor range images of the University of South Florida (USF) database [8]. Several works on the parametric model fitting problem can be found in the literature [9–13]. Those works can be extended in order to segment surfaces such as ground and façades on elevation images. Their main drawbacks are that they require a model estimation which can be different for different images, are time consuming due to minimization procedures and may produce under- and over-segmentation.

Demantke *et al.* [14] propose a method to adapt 3D spherical neighborhoods based on local features. Radius selection is carried out optimizing local entropy. Then, dimensionality features are calculated. These features can be useful to classify 1D structures such as pole-like objects, 2D structures such as ground or façades, and 3D volumetric structures such as trees and urban objects.

Hernández and Marcotegui [2] assume that façades on the same street are aligned, which is verified in their *Paris-rue-Soufflot* database. They use the Hough transform to find façade directions. Then, they analyze the height of each profile in order to detect façades and city blocks. Hammoudi [15] presents a similar technique based on the Progressive Probabilistic Hough Transform in order to detect walls and windows. He assumes that building façades are mainly vertical, so it is possible to generate an accumulation image to compute the number of points projected on each pixel.

Other works aiming at segmenting façades are available in the literature [16–21]. Additionally, color images can be used in order to enrich façade segmentations. For example, several works available in the literature use color images to segment façades keeping architectural details such as balconies and windows. Then, those results are used to colorize 3D point clouds providing a realistic rendering [22–26].

In general, methods based on planar extraction, Hough transform and model fitting are not generic enough since façades can have different shapes and contain different architectural details. Moreover, the problem of objects touching the façade, such as urban furniture or parked motorcycles, is rarely analyzed in the literature. In this paper, we address these problems proposing methods based on geometric and elongation features computed on projection images. These methods are robust to extract façades of several shapes and correctly separate objects touching the façade.
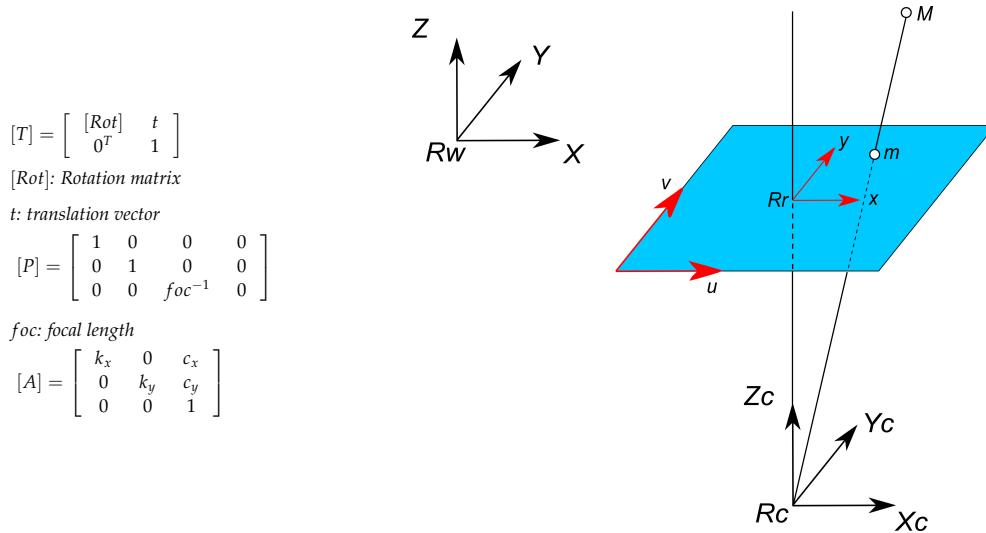
3. Projection Images

Once a 3D point cloud is acquired, a suitable data structure is not only required to inspect and visualize the information quality, but also to process it conveniently. Point clouds are delivered as long lists of (x, y, z) coordinates, possibly with attributes such as intensity and color. Points are usually listed in scan line order, which is not suitable for efficient processing. For example in a (x, y, z) list, it is not possible to quickly determine the neighbors of a point within some radius. Data structures such as projection images, Delaunay triangulation, octrees and k-D trees allow this kind of processing. Choosing the proper data structure is application dependent, moreover it is possible to combine some of them to get better results in specific tasks such as visualization, filtering, segmentation and classification [1].

As aforementioned, our method is based on projection images. 3D point clouds are projected to elevation images because they are convenient structures to visualize and to process data. One can utilize all the large collection of existing image processing tools, in particular mathematical morphology [27,28]. Additionally, images can be processed quickly, implicitly define neighborhood relationships and require less memory than 3D data. In general, the idea of deriving elevation images from 3D point clouds is not new. Nevertheless, the development of accurate and fast urban analysis algorithms is still an open problem.

Elevation images are 2.5D structures that contain depth information at each pixel. They are generated by an orthographic projection to a virtual camera plane, *i.e.*, the depth is the distance from each 3D point to the projection plane. The camera model $\mathcal{P}$ is a projective transformation from $\mathbf{R}^3 \rightarrow \mathbf{N}^2$, and it can be decomposed in three sequential transformations as follows:

Definition 1. *Let $M = (X, Y, Z)$ be a 3D point in $\mathbf{R}^3$ and $m = (u, v)$ a point in the image space $\mathbf{N}^2$. The camera model $\mathcal{P}$ is defined as the successive transformations:*

$$(X, Y, Z) \xrightarrow{T} (X_c, Y_c, Z_c) \xrightarrow{P} (x, y) \xrightarrow{A} (u, v) \tag{1}$$

$$[T] = \begin{bmatrix} [Rot] & t \\ 0^T & 1 \end{bmatrix}$$

[Rot]: Rotation matrix

t: translation vector

$$[P] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & foc^{-1} & 0 \end{bmatrix}$$

foc: focal length

$$[A] = \begin{bmatrix} k_x & 0 & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

where, (c_x, c_y) is the intersection point between the optical axis and the image plane, (k_x, k_y) are the number of pixels per unit of length, $\mathcal{R}_w$ is the real world coordinate system, $\mathcal{R}_c$ is the camera coordinate system with origin in the optical center, and $\mathcal{R}_r$ is the projection plane coordinate system. The virtual camera is chosen such that:

- *It is located on the horizontal plane with normal vector $\overrightarrow{n} = (0, 0, 1)$ and crossing the lowest point $(0, 0, z_{min})$. Therefore, the Rotation matrix $[Rot]$ is equal to the identity.*
- *The point cloud and the projection plane are centered on the gravity center of the point cloud. Thus, the translation vector t is equal to the gravity center, and the intersection point $(c_x, c_y) = (0, 0)$.*
- *The projection is orthographic. Thus, the projection axis is orthogonal to the projection plane and the projection center is located at the infinity. It means, $foc = \infty$, $x = X_c$ and $y = Y_c$.*

According to these assumptions, the number of pixels per unit of length is the only free parameter. In order to simplify this selection, we assume $k_x = k_y = k$. It has to be carefully chosen. On the one hand, if k is too small, fine details are not preserved because too many points would be projected on the same pixel. On the other hand, too large k implies connectivity problems and large image sizes, which implies higher computational time that would no longer justify the use of elevation images instead of 3D point clouds. This parameter is the most critical in terms of quality versus processing time. In our experiments, we have chosen $k = 5$ pix/m thanks to a priori knowledge about point cloud density.

During projection, several points can be projected to the same pixel. Thus, four elevation images are defined:

- *Maximal elevation image (f_{max})*, or simply elevation image, stores the maximal depth among all projected points on the same pixel.
- *Minimal elevation image (f_{min})*, stores the minimal depth among all projected points on the same pixel.
- *Height image (f_{height})*, contains the difference between maximal and minimal elevation images.

- *Accumulation image* (f_{acc}), stores the number of points projected on each pixel.

In general, processing steps are performed combining these images and resulting image can be reprojected to the 3D point cloud. Figure 2 presents the 3D point clouds and the elevation images for our two Paris test sites.
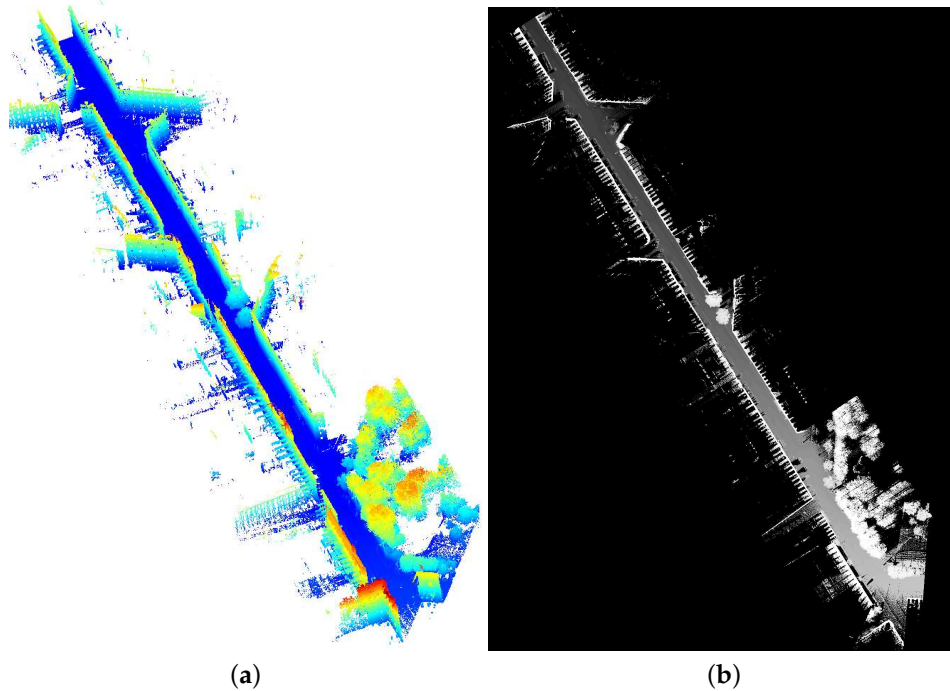

(**a**)　　　　　　　　　　　　　　　　　　　　　(**b**)

Figure 2. 3D point clouds and elevation images for a test site in Paris, France. Data acquired by Stereopolis II, IGN©France. (**a**) *Rue d'Assas:* point cloud; (**b**) *Rue d'Assas:* elevation image.

Image projections imply a reduction in the amount of data to be processed. Therefore, processing 2.5D images using image processing techniques is much faster than process the 3D points directly. Moreover, neighborhood relationships are given in images without any additional computation. In order to avoid connectivity problems due to occlusions or low dense point clouds, an interpolation may be applied. This topic is out of the scope of this paper. For further details on interpolation of elevation images, the reader is encouraged to refer to Serna and Marcotegui [3], Hernández and Marcotegui [29].

4. Façade Segmentation Using Markers

In order to segment façades, we propose a marker-based method using geometrical constraints. After marker extraction, a reconstruction is applied in order to get the entire façade. Let us explain first the façade marker extraction and later in Section 4.2 the reconstruction process.

4.1. Façade Marker Extraction

The height image f_{height} is appropriate to compute façade markers since it contains information about high and vertical structures, as shown in Figure 3.

Figure 3a,b show maximal f_{max} and minimal f_{min} elevation images, respectively. While Figure 3c presents the height image.

From the work by Hernández and Marcotegui [29], we use the two following geometric constraints on f_{height} in order to extract façade markers:

- h_{min}, defining the minimal allowed façade height. In our experiments, this variable has been set to 3.5 m according to architectural characteristics of our databases. This threshold is illustrated in Figure 4a. Note that only the highest objects are preserved.
- L_{min}, defining the minimal allowed façade length. In our experiments, this variable has been set to 5 m according to architectural characteristics of our databases. This threshold is illustrated in Figure 4b. Note that small objects such as lampposts and objects behind façades are not long enough and are then eliminated.
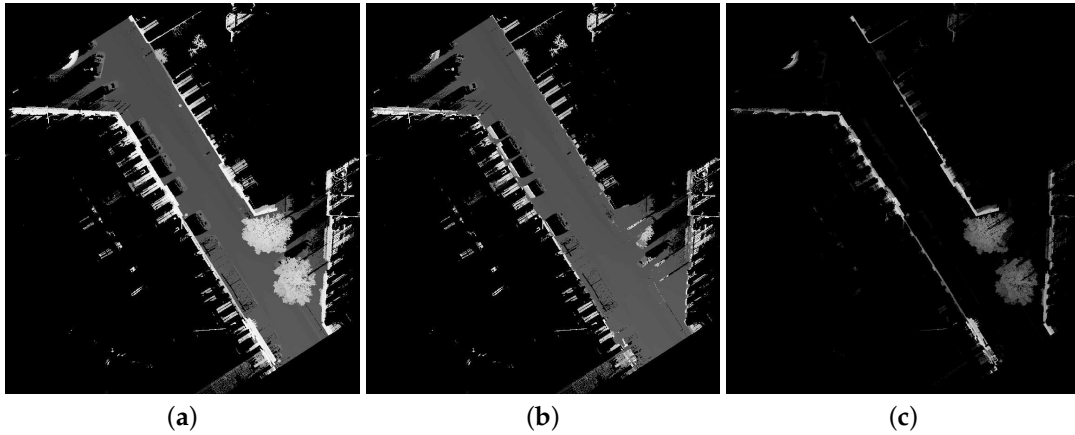

| (a) | (b) | (c) |

Figure 3. Elevation images from a test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Maximal elevation image f_{max}; (**b**) Minimal elevation image f_{min}; (**c**) Height image $f_{\text{height}} = f_{max} - f_{min}$.

In addition, we propose a third constraint in order to eliminate round objects such as trees:

- C_{min}, defining the maximal allowed façade circularity (circularity of an object X is defined as the inverse of its elongation $Circ(X) = 1/E(X)$. In this work, we focus on geodesic elongation, simply called henceforth elongation [30]. The elongation $E(X)$ of an object X is a shape descriptor useful to characterize long and thin structures. It is proportional to the ratio between the square geodesic diameter $L^2(X)$ and the area of the object $A(X)$, as shown in Equation (2). The geodesic diameter $L(X) = \sup_{x \in X}\{l_x(X)\}$ is the length of the longest geodesic arc of X, *i.e.*, the longest internal segment $l_x(X)$ connecting the two end points of X [31]. The longer and narrower the object, the higher the elongation. The lowest bound is reached with the disk, where $E(X) = 1$. An efficient implementation can be found in Morard *et al.* [32].

$$E(X) = \frac{\pi}{4}\frac{L^2(X)}{A(X)} \tag{2}$$

In our experiments, C_{min} has been heuristically set to $1/3$, which corresponds to the circularity of an ellipse whose major axis is 3 times longer than the minor one. This value is appropriate to filter tree-tops since circularity of a circle is equal to 1. This threshold is illustrated in Figure 4c. Note that non-elongated objects such as the two trees in the right street side have been eliminated.

Using these constraints, we extract façade markers as the union of connected components (CC) higher than h_{min}, longer than L_{min} and less circular than C_{min}, as established in Definition 2:

Definition 2. *Let f_{height} be a height image $f_{height}: D \rightarrow V$, with $D \subset Z^2$ the image domain and $V = [0, ..., H]$ the set of gray levels mapping the pixel height. Let $Th(f_{height})$ be the binary image containing the pixels higher than h_{min}:*

$$Th(f_{height}) = \{p \in D | f_{height}(p) > h_{min}\} \tag{3}$$

Let C_1, C_2,..., C_n be the connected components of image $Th(f_{height})$:

$$Th(f_{height}) = \bigcup_{i=1}^{n} C_i, \; i \neq j \Rightarrow C_i \cap C_j = \varnothing \tag{4}$$

Then, façade markers F_{mark} are the connected components C_i of image $Th(f_{height})$ which are longer than L_{min} and less circular than C_{min}:

$$F_{mark} = \{C_j | L(C_j) > L_{min} \wedge Circ(C_j) < C_{min}\} \tag{5}$$

$\forall j \in \{1,...,n\}$, where $L(C_j)$ and $Circ(C_j)$ are respectively the geodesic diameter and the circularity of connected component C_j.

It is noteworthy that these three parameters (h_{min}, L_{min} and C_{min}) are easy to tune since they have a physical meaning and depend on urban/architectural constraints. Figure 4 illustrates this marker selection process.
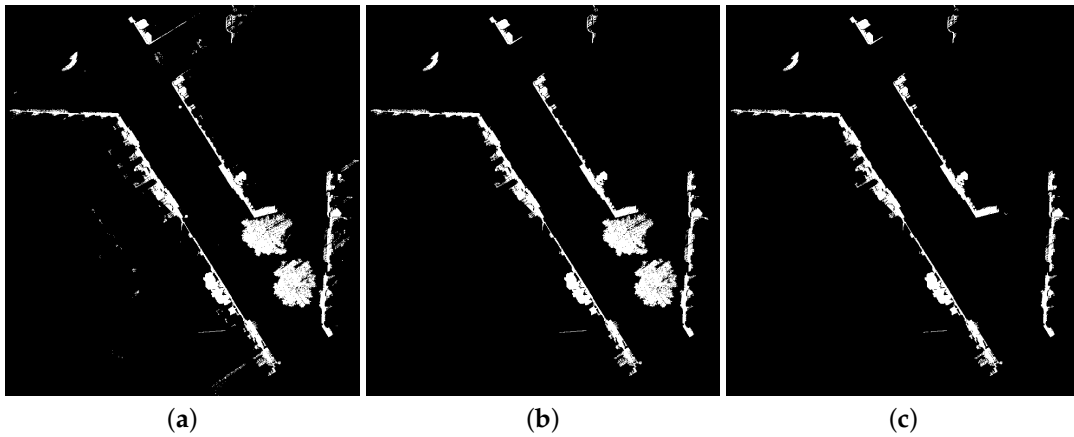

(**a**)   (**b**)   (**c**)

Figure 4. Façade marker extraction. Markers are represented with white pixels: (**a**) Points higher than h_{min}; (**b**) Points higher than h_{min} and longer than L_{min}; (**c**) Façade markers F_{mark}: points higher than h_{min}, longer than L_{min} and less circular than C_{min}. Test site in *rue d'Assas* (Paris).

Due to specific requirements in our TerraMobilita datasets, several 3D point clouds have been acquired with the laser system oriented to the ground. Therefore, structures higher than 2.5 m are out of the laser field of view, as shown in Figure 5. This is challenging for methods using height constraints since high wall parts are not visible.

To solve this problem, we propose a solution taking advantage of the acquisition cycle of the MLS sensor, as shown in Figure 6. In our configuration, the sensor spins scanning vertical lines starting from the top. Thus, the first and the last point of each spin correspond to the highest point on the right and on the left street side, respectively. These highest points are usually located on the façade and can be extracted by detecting sign changes in the *angle of depression* (computed using the sensor position) between consecutive points. Then, these markers are added to image $Th(f_{height})$ and isolated points are filtered out using the same process than before: only markers longer than L_{min} and less circular than C_{min} are considered as façade markers.

Figure 7 illustrates façade marker extraction when the laser sensor is oriented to the ground. Figure 7b presents the façade markers reprojected onto the 3D point cloud. The test site corresponds to a street section in *rue Vaugirard* in Paris, France.
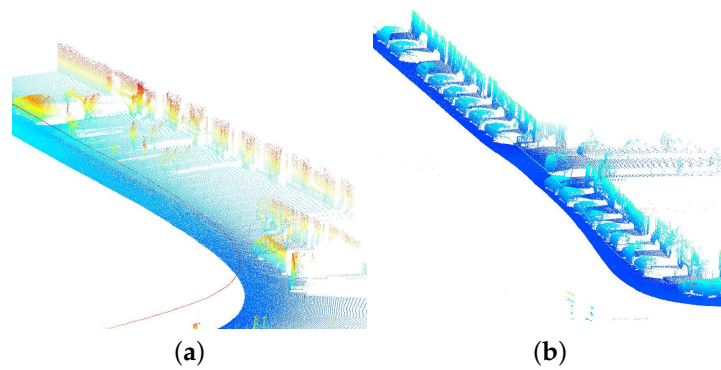
Figure 5. 3D point clouds when laser is oriented to the ground. 3D point clouds are colored by elevation. (**a**) Bend and (**b**) straight street sections in *rue Vaugirard* (Paris). Stereopolis II, IGN©.
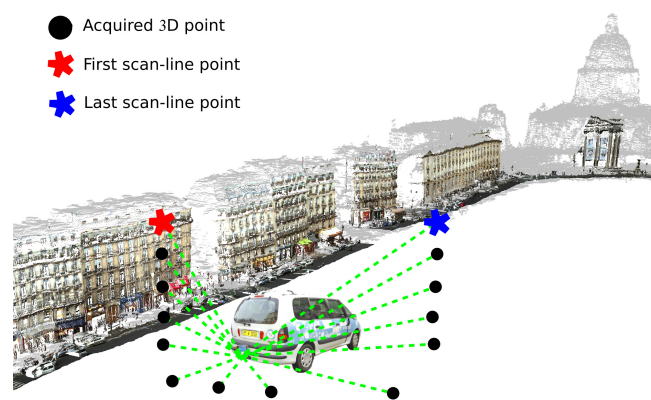

Figure 6. Mobile laser scanning (MLS) acquisition cycle. The first and the last point are taken as façade markers. Test site in *rue Soufflot* in Paris, France. LARA-3D, CAOR-MINES ParisTech©.
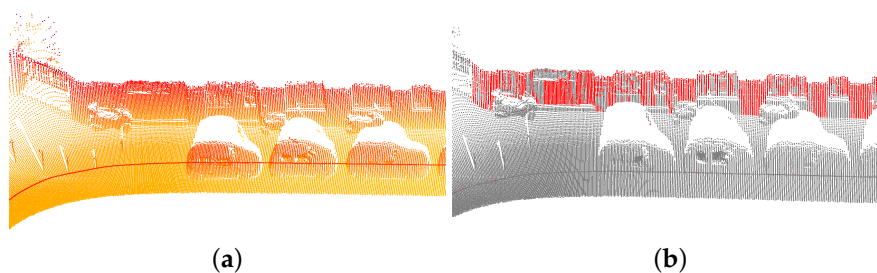

Figure 7. Façade markers reprojected onto the 3D point cloud. During this acquisition, the laser sensor was oriented to the ground. Note that markers only represent partial façades, therefore a façade reconstruction process is required, as explained in Subsection 4.2. Test site in *rue Vaugirard* in Paris, France. Stereopolis II, IGN©. (**a**) 3D point cloud colored by the Z coordinate; (**b**) Façade markers (red), other (gray).

4.2. Façade Reconstruction from Markers

As mentioned earlier, façade markers only contain partial façades. Therefore, a reconstruction should be applied from those markers in order to retrieve the whole façade. With this purpose, we use a reconstruction constrained by the ground residue. Ground G is assumed to be previously computed using either our ground segmentation method proposed in [3,29] or any other method reported in the literature [33–36].

Ground residue G^c is computed as the difference between the elevation image and the ground image: $G^c = f_{max} - G$. Then, a first solution for the reconstruction process consists in a set of increasing geodesic dilations applied until idempotence. This transformation is called reconstruction by dilation [31] and is defined as follows:

Definition 3. *Reconstruction by dilation. The reconstruction by dilation of a mask image G^c from a marker $F_{mark} \leq G^c$ is defined as the geodesic dilation of F_{mark} within G^c until idempotence and it is denoted by:*

$$R_{G^c}^{\delta}(F_{mark}) = \delta_{G^c}^{(i)}(F_{mark}) \tag{6}$$

where i is the geodesic dilation size for which idempotence is reached, i.e., $\delta_{G^c}^{(i+1)}(F_{mark}) = \delta_{G^c}^{(i)}(F_{mark})$

Figure 8 illustrates the façade segmentation based on reconstruction by dilation. Figure 8a an elevation image and Figure 8b its corresponding height image. Figure 8c shows the façade markers F_{mark} computed by the method explained above in Section 4.1. Figure 8d presents the ground segmentation result G, while Figure 8e presents the ground residue G^c. Finally, Figure 8f shows the façade segmentation obtained by reconstruction by dilation of the ground residue from the façade markers. Note that pixels behind façades have been included in the segmentation result. This method is fast and easy to implement. However, the main drawback is that objects connected to the façade, e.g., motorcycles or pedestrians leaning on walls or objects behind façades, are reconstructed as well.
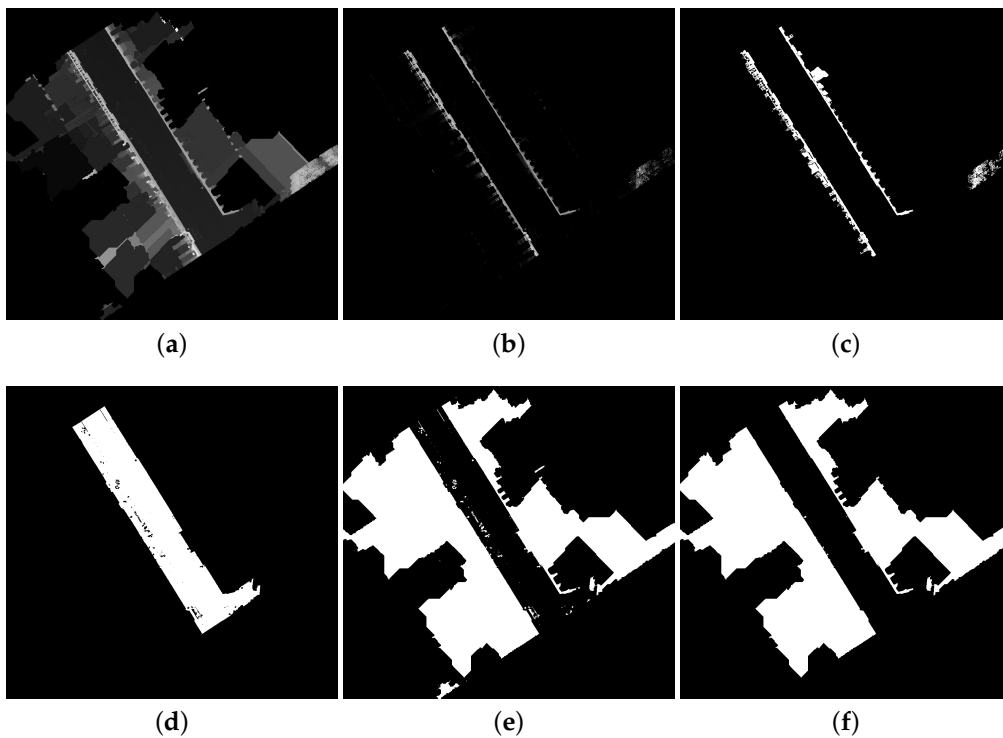

(a)　　　　　　　　(b)　　　　　　　　(c)

(d)　　　　　　　　(e)　　　　　　　　(f)

Figure 8. Façade segmentation using reconstruction by dilation on the ground residue image from façade markers. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Max. elevation image f_{max}; (**b**) Height image f_{height}; (**c**) Façade markers F_{mark}; (**d**) Ground segmentation G; (**e**) Ground residue $G^c = f_{max} - G$; (**f**) Façade segmentation using reconstruction by dilation: $R_{G^c}^{\delta}(F_{mark})$.

Figures 9 and 10 present this segmentation result reprojected onto the 3D point cloud. Note that the reconstruction by dilation retrieves not only the entire façade but also other objects connected to it. For example in Figure 10, a store, a pedestrian and a traffic sign on the right sidewalk have been wrongly reconstructed.

In order to solve the problem of objects touching the façade, we propose an attribute controlled reconstruction from façade markers. This procedure has been previously published in Serna and Marcotegui [19]. It consists of appending nearby points with similar height. The propagation process stops when a given attribute reaches its maximum value. In our case, we use increasing propagations from façade markers over λ-flat zones, called also in the literature quasi-flat zones [37]. As façades are the longest and most elongated structures in the elevation image, we keep the propagation that maximizes the geodesic elongation.

Let us introduce a formal definition for the façade segmentation process using attribute controlled reconstruction:

Definition 4. *Attribute controlled reconstruction. Let G^c be a digital elevation image containing the ground residue $G^c : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., H]$ the set of gray levels mapping the elevation values. Two neighboring pixels p, q belong to the same λ–flat zone of G^c, if their absolute difference $|G^c(p) - G^c(q)|$ is smaller than or equal to a given λ value.*

For all $x \in F_{mark} \subseteq D$, let Λ be the set of increasing regions containing marker pixel x. For all $\lambda \in V$ and $j = [1, ..., n - 1]$, we define $A_\lambda(F_{mark}) \in \Lambda$ as the λ–flat zone of image G^c containing marker F_{mark}:

$$A_\lambda(F_{mark}) = \{x\} \cup \{q | \exists \wp = (p_1 = x, ..., p_n = q) \text{ such that } |G^c(p_j) - G^c(p_{j+1})| \leq \lambda\} \tag{7}$$

Let $E(A_\lambda(F_{mark}))$ be the geodesic elongation of λ–flat zone $A_\lambda(F_{mark})$. For all $\lambda_i \in V$ and $i = [0, ..., R]$, we define λ_M as the value for which the elongation is maximum:

$$\lambda_M = argmax_{\lambda_i \in V} |E(A_{\lambda_i}(F_{mark}))| \tag{8}$$

Then, we define $A_{\lambda_M}(F_{mark})$ as the attribute controlled reconstruction of the façade from marker F_{mark}.

Using this controlled reconstruction maximizing the geodesic elongation, it is possible to reconstruct the façade without merging adjacent objects. Figure 9 compares façade segmentation methods using reconstruction by dilation $R^\delta_{G^c}(F_{mark})$ and attribute controlled reconstruction $A_{\lambda_M}(F_{mark})$, both from marker F_{mark}. It is noteworthy that neither objects connected to the façade nor objects behind it have been reached by the attribute controlled reconstruction. Figure 10d presents the segmentation result using attribute controlled reconstruction reprojected onto the 3D point cloud. Compared to Figure 10b, note that the store, the pedestrians and the traffic sign have not been included in the façade. Additionally, several objects such as wall lamps and objects behind façades have been correctly separated.
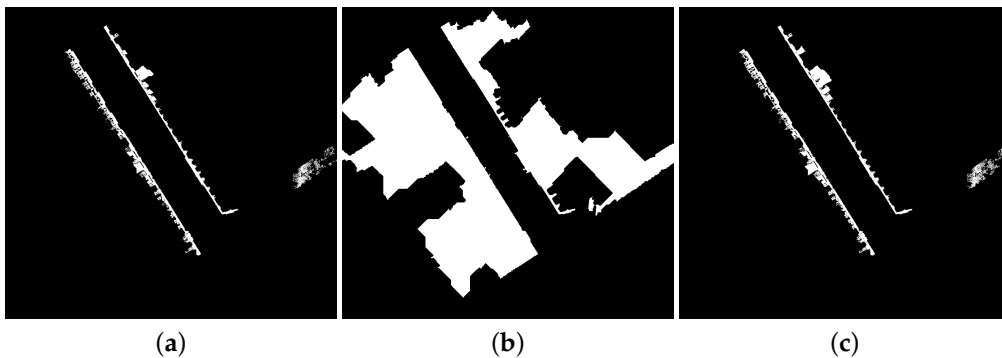

(**a**)　　　　　　　　　(**b**)　　　　　　　　　(**c**)

Figure 9. Comparison of façade segmentation methods using reconstruction by dilation and attribute controlled reconstruction on the ground residue image. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Façade markers F_{mark}; (**b**) Reconstruction by dilation $R^\delta_{G^c}(F_{mark})$; (**c**) Attribute controlled reconstruction $A_{\lambda_M}(F_{mark})$.

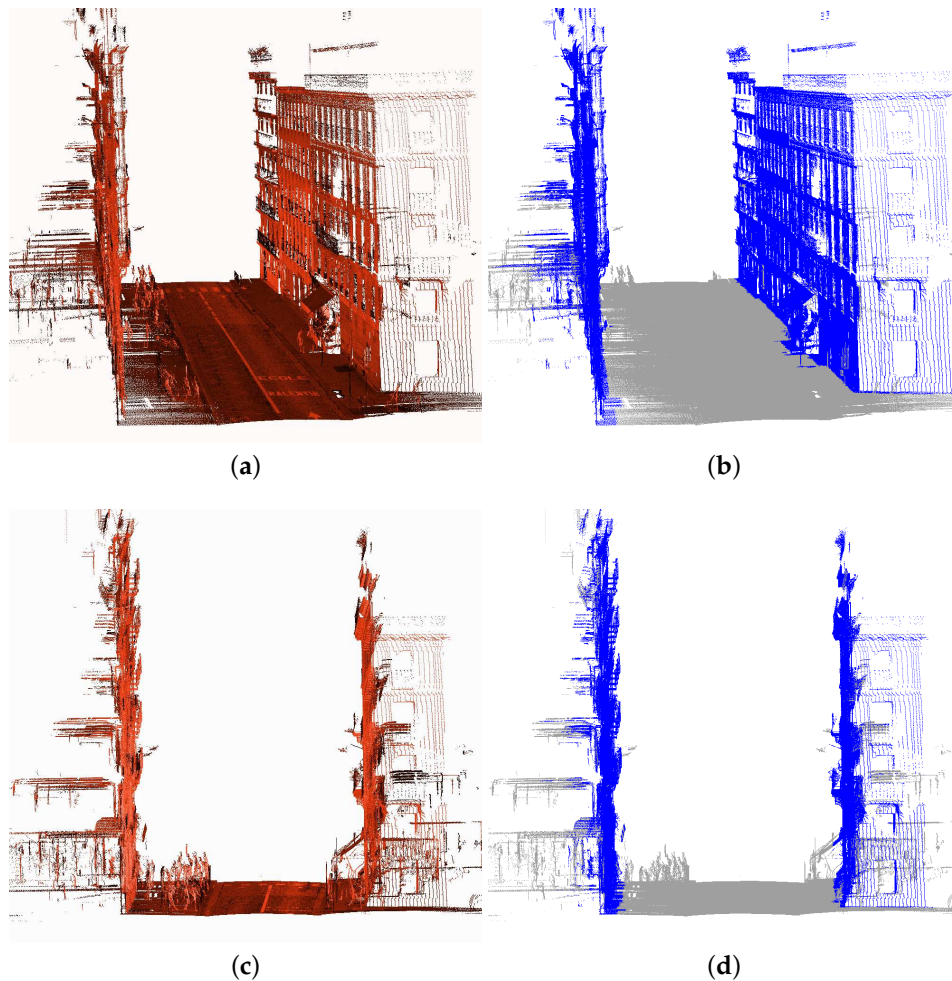(**a**)   (**b**)

(**c**)   (**d**)

Figure 10. Façade segmentation result reprojected onto the 3D point cloud. Images correspond to two different views of the same test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Point cloud colored by the laser intensity; (**b**) Segmentation result: façade (blue), other (gray). Façade has been segmented using reconstruction by dilation on the ground residue image from façade markers; (**c**) Point cloud colored by the laser intensity; (**d**) Segmentation result: façade (blue), other (gray). Façade segmentation using attribute controlled reconstruction on the ground residue image.

4.3. Discussion

Methods based on markers are robust to segment non-planar façades and façades with several architectural details and balconies, such it is the case in Parisian buildings. Moreover, these methods are simple and fast since they are based on geometric constraints, easily translated into threshold values. In the case of low façades or when the laser sensor is oriented to the ground, additional markers corresponding to the highest points of each laser cycle are added. Since markers only contain partial façades, a reconstruction is required in order to get the entire façade. In our case, reconstruction is constrained to the ground residue.

In this section, we have proposed two reconstructions: (i) a classic reconstruction by dilation; (ii) a reconstruction based on attribute controlled propagation. The performance of these methods strongly depends on the markers selection. A wrong marker may produce errors since it will reconstruct the marked object, even if it is not a façade. In our experiments, our marker selection method has proved to be efficient in many cases. However, objects such as tree alignments may produce false markers, therefore wrong segmentations, as shown in Figure 11. The reason for this problem is that our circularity criterion fails filtering trees alignments since their footprints may be

elongated enough. Changing the value of circularity does not seem to be a good idea. On the one hand, reducing the circularity value would affect the recall of the method, *i.e.*, less façades would be detected. On the other hand, increasing the circularity value would increase the number of false positives.
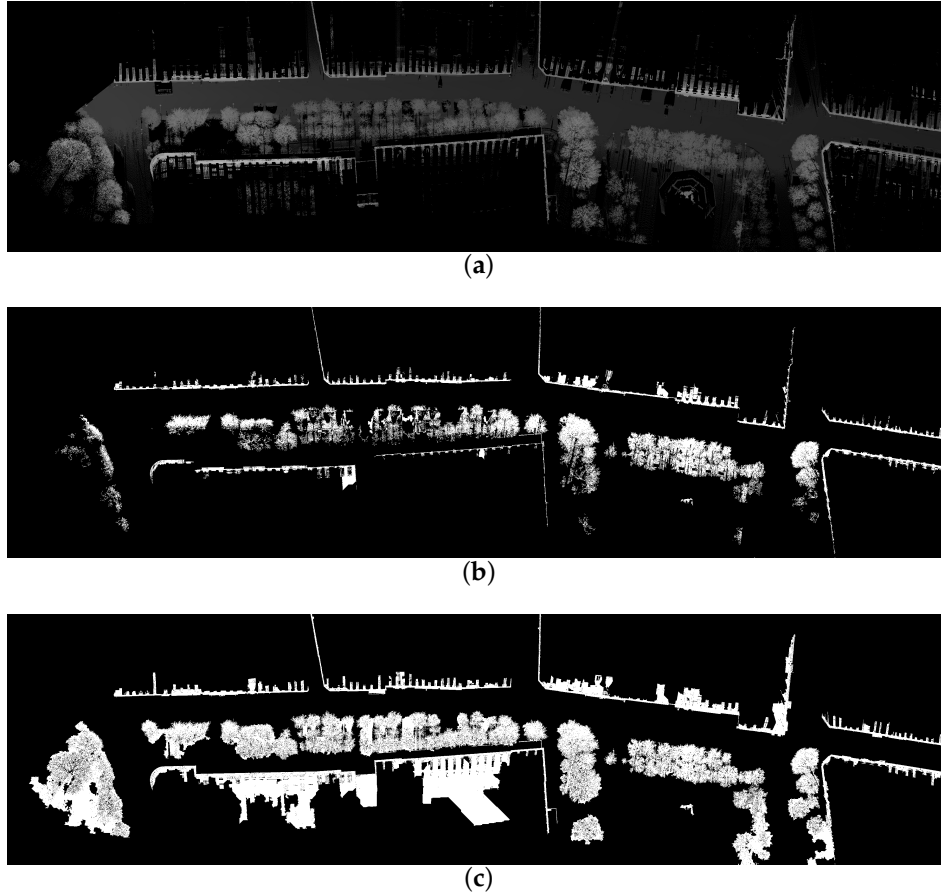

(**a**)

(**b**)

(**c**)

Figure 11. Errors in façade segmentation due to tree alignments wrongly extracted as façade markers. Test site in *St. Sulpice square* in Paris, France. Stereopolis II, IGN©. (**a**) Elevation image; (**b**) Façade markers. Several incorrect markers have been detected due to tree alignments; (**c**) Façade segmentation using attribute controlled reconstruction.

This is our motivation to propose a third method, more robust than the two proposed previously, avoiding the use of façade markers. Our idea is analyzing the elongation and its evolution over the height decomposition of the 3D scene. Let us introduce such a method in the following section.

5. Façade Segmentation without Markers

In order to segment façades avoiding the use of markers, we propose a method based on threshold decomposition and attribute profiles. In [38], we have introduced this method in order to segment elongated objects on gray-scale images. Let us remind its definition in the 2D case:

Definition 5. *2D Attribute profile.* *Let I be a digital gray-scale image $I : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels. A decomposition of I can be obtained considering successive thresholds:*

$$T_t(I) = \{p \in D | I(p) > t\} \, \forall t = [0, ..., R-1] \tag{9}$$

Since this decomposition satisfies the inclusion property $T_t(I) \subseteq T_{t-1}(I), \forall t \in [1, ..., R-1]$, it is possible to build a tree, called the component tree, with level sets $T_t(I)$. Each branch of the tree represents the evolution

of a single connected component X_t. An attribute profile is the evolution of an attribute (e.g., area, perimeter, elongation, average gray-level, etc.) of a given CC along a branch of the tree.

Figure 12 illustrates the threshold decomposition for a 1D function, its component tree and the attribute (width) profiles for the two function maxima (p_A and p_B). Events on this attribute profile are useful to segment objects [39], extract features [40–42] and define adaptive structuring elements [19].
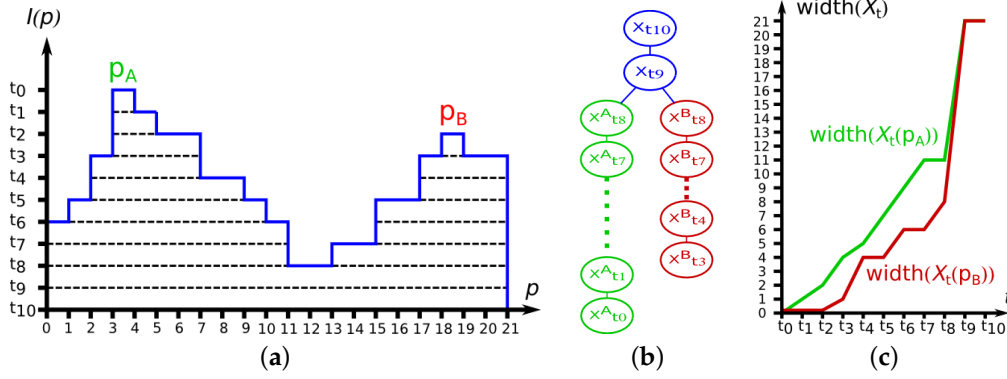

Figure 12. (**a**) 1D threshold decomposition; (**b**) component tree; (**c**) attribute profile.

Now, let us extend this definition to the 3D case:

Definition 6. *3D Attribute profile. Let P be a 3D point cloud containing a list of N points $p_i = (x_i, y_i, z_i) \subset \Re^3$, where $i = [0, ..., N]$. Let $m_i = (u_i, v_i) \subset Z^2$ the projection of point p_i in the elevation image f_{max}. A decomposition of P in horizontal slices can be obtained considering successive thresholds on the Z axis separated by a given height Δz:*

$$T_t^Z(P) = \{m_i \in D \mid t\Delta z < z_i < (t+1)\Delta z\}; \; \forall t = [0, ..., R-1]; \; \forall i = [0, ..., N] \tag{10}$$

Contrarily to the 2D case, this decomposition does not satisfy any inclusion property. However, it is always possible to analyze the evolution of a single connected component X_t over horizontal slices $T_t^Z(P)$. A 3D attribute profile is the evolution of an attribute (e.g., number of points, density, average elevation, etc.) of a given CC along the decomposition.

More adapted to our 3D urban data, let us define an adaptive decomposition using slices parallel to the ground. From Definition 6, we propose the following decomposition:

Definition 7. *Adaptive voxelization using ground information. Let G be a digital gray-scale image $G : D \to V$, with $D \subset Z^2$ the image domain and $V = [0, ..., R]$ the set of gray levels mapping the ground elevation, resulting from a given ground segmentation process. A decomposition of a 3D point cloud P using slices parallel to the ground can be obtained considering successive thresholds from the ground separated by a given height Δz:*

$$T_t^{gr}(P) = \{m_i \in D \mid G(m_i) + t\Delta z < z_i < G(m_i) + (t+1)\Delta z\} \tag{11}$$

$\forall t = [0, ..., R-1]; \; \forall i = [0, ..., N]$

This decomposition is equivalent to an adaptive voxelization, as shown in Figure 13. Dashed lines represent slices parallel to the ground. For each slice, an occupancy grid is defined according to the elevation image pixel size $1/k$, where k is the number of pixels per unit length (For further details, see [43]). Each voxel is labeled *full* if there is at least one 3D point inside, or *empty* otherwise. Finally, these occupancy grids are stacked in a binary 3D image. For each slice, attributes are computed on each binary CC.
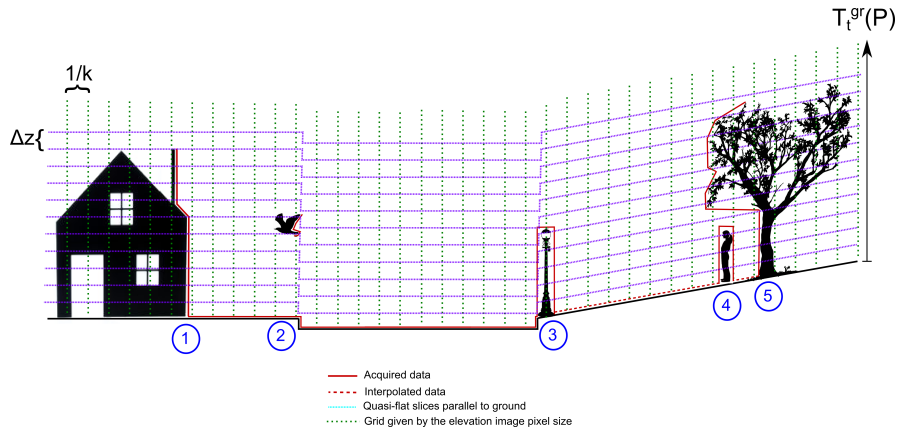
Figure 13. Adaptive voxelization using slices parallel to the ground. This example contains five objects: ① façade, ② bird, ③ lamppost, ④ pedestrian, and ⑤ tree.

We propose to segment façades using the maximal elongation image computed from the attribute profile of decomposition T_t^{gr}. With this aim, we compute the geodesic elongation $E(X_t)$ for each CC on each slice parallel to the ground. Then, for each pixel m_i, we store the maximal elongation over the whole decomposition:

$$E_{max}(m_i) = max|E(X_t(m_i))|; \ \forall X_t \in T_t^{gr}; \ \forall t \in [0, ..., R-1] \tag{12}$$

Such feature image is a partition of the space where each pixel contains information about elongation of its neighborhood. Then, it is useful in segmentation tasks where some prior shape knowledge is available. This decomposition is used to segment façades while filtering out other structures, such as objects connected to it. The slice height has been set to $\Delta z = \Delta \lambda = 1.0$ m since we are only interested in connected objects higher than 1.0 m (motorcycles, pedestrians, urban furniture, *etc.*). Additionally, it offers a trade off between processing time and performance since only a few tens of slices are required to decompose an urban scenario without skyscrapers.

Figure 14 illustrates an example of façade segmentation using this approach. Figure 14a shows the elevation image. Figure 14b presents the elongation image computed from the threshold decomposition of the 3D point cloud. Figure 14c presents the segmentation result applying a simple threshold on the elongation image. We define E_{min} as the minimal elongation allowed for a façade. In our experiments, we have heuristically set $E_{min} = 20$, which corresponds to the elongation of a rectangle whose length is approximately 25 times longer than its width. It is noteworthy that this elongation threshold is more restrictive than that used for marker extraction (Section 4.1) since here we are interested in detecting the maximal elongation value over all the decomposition.
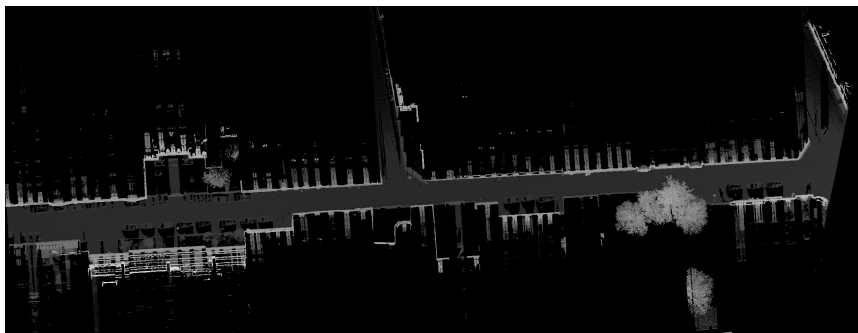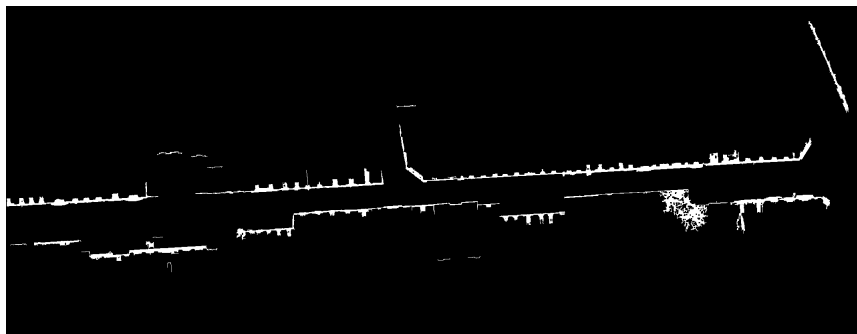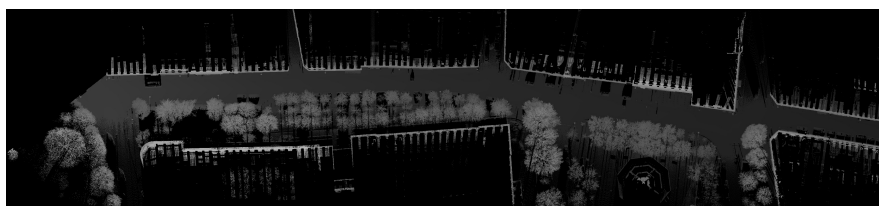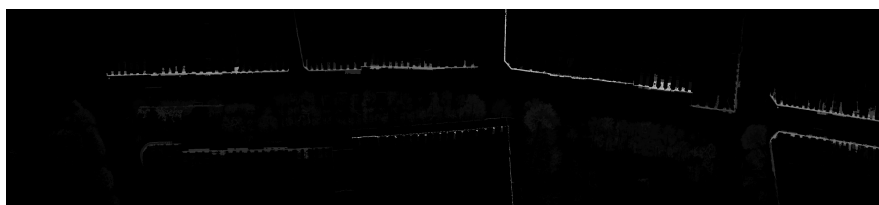

(**a**)

Figure 14. *Cont.*
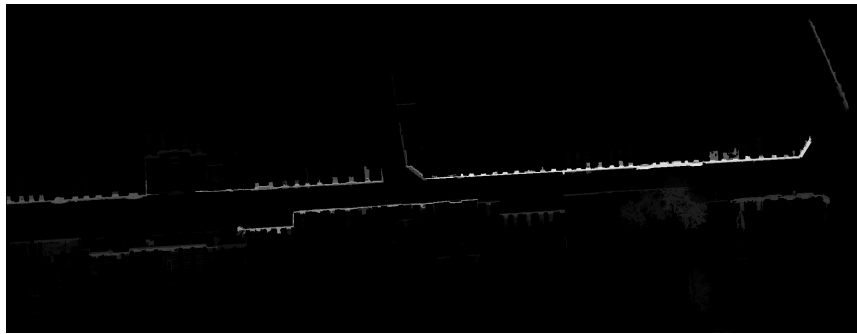
(**b**)

(**c**)

Figure 14. Façade segmentation using the maximal elongation image (without markers). Test site in *rue Cassette* in Paris, France. Stereopolis II, IGN©. (**a**) Elevation image; (**b**) Maximal elongation image computed from the 3D point cloud; (**c**) Façade segmentation using the maximal elongation image.

Figure 15 presents a façade segmentation result on a test site in *rue Bonaparte* in Paris, France. With respect to Figure 11 most façades are correctly segmented. Some problems appear in the left part: zone 1, where the side part of a bus has been wrongly detected as façade; and zone B, where bushes and vegetation over a low wall could not be separated (Figure 15d). These objects present a high elongation, then they have been wrongly segmented as façades.

(**a**)

(**b**)

Figure 15. *Cont.*
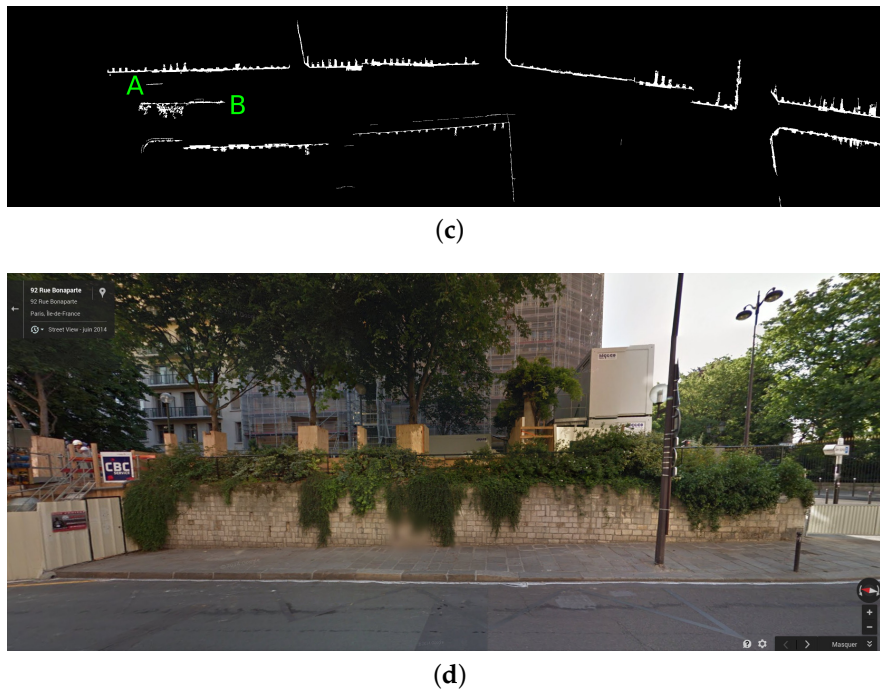
(**c**)

(**d**)

Figure 15. Façade segmentation using the maximal elongation image (without markers). Test site in *rue Bonaparte* in Paris, France. Stereopolis II, IGN©. (**a**) Elevation image; (**b**) Maximal elongation image computed from the 3D point cloud; (**c**) Façade segmentation using the maximal elongation image. Two segmentation errors have been found in zones A and B; (**d**) Façade segmentation error due to vegetation and bushes over a low wall. This is an unusual case presented in zone B in Figure 15c.

6. City Block Segmentation

A city block is the smallest area that is surrounded by streets. A wide variety of sizes and shapes can be found in urban environments. In general, it depends on historic, demographic and geographic constraints. For example, many pre-industrial cities tend to have irregular city blocks, while newer cities have usually much more regular arrangements [44].

In our application, city blocks are considered as the biggest semantic entities in the urban environment. Their segmentation is useful for individual city block analysis. Additionally, each city block can be processed separately and their results joined at the end of the analysis, reducing memory requirements and allowing parallelization.

Once façades have been segmented using one of our three proposed methods, we compute the influence zone (IZ) of each façade in order to define city blocks. The IZ algorithm was one of the first morphological operators applied to image segmentation. It was discovered in the 1970s by the iterative application of basic operators such erosion and dilation [45,46]. The IZ of a given CC is defined by the set of pixels of a binary image that are closer to this CC than to any other CC on the image. Let us introduce its formal definition:

Definition 8. *Influence zones (IZ).* *Let X be a binary image and K_1, K_2,..., K_n the CC of X. The influence zone of K_i is the set of pixels of image X that are closer to K_i than to any other CC of image X:*

$$IZ(K_i) = \{p | \forall j \in \{1, ..., n\}, i \neq j \Rightarrow d(p, K_i) < d(p, K_j)\} \tag{13}$$

It is noteworthy that this is an alternative definition of Voronoi diagrams [47]. In practice, the IZ of a binary image is computed using a constrained watershed on the distance function of the binary image. Figure 16 illustrates this city block segmentation. Figure 16a shows the elevation image. Figure 16b presents the binary image containing the façade segmentation result. Figure 16c shows

a morphological closing of size S_{min} in order to reconnect near façades belonging to the same city block, *i.e.*, S_{min} stands for the minimal separation between city blocks. Figure 16d shows the medial road axes useful to avoid defining city block crossing the street. This information is used if available and it can be obtained from the vehicle trajectory or from an external 2D map. Figure 16e illustrates the distance function computed from façades. The distance function is constrained to be maximum on both medial road axes and empty pixels. Finally, Figure 16f presents the IZ as the result of a constrained watershed on the distance function. Each color represents a different city block.
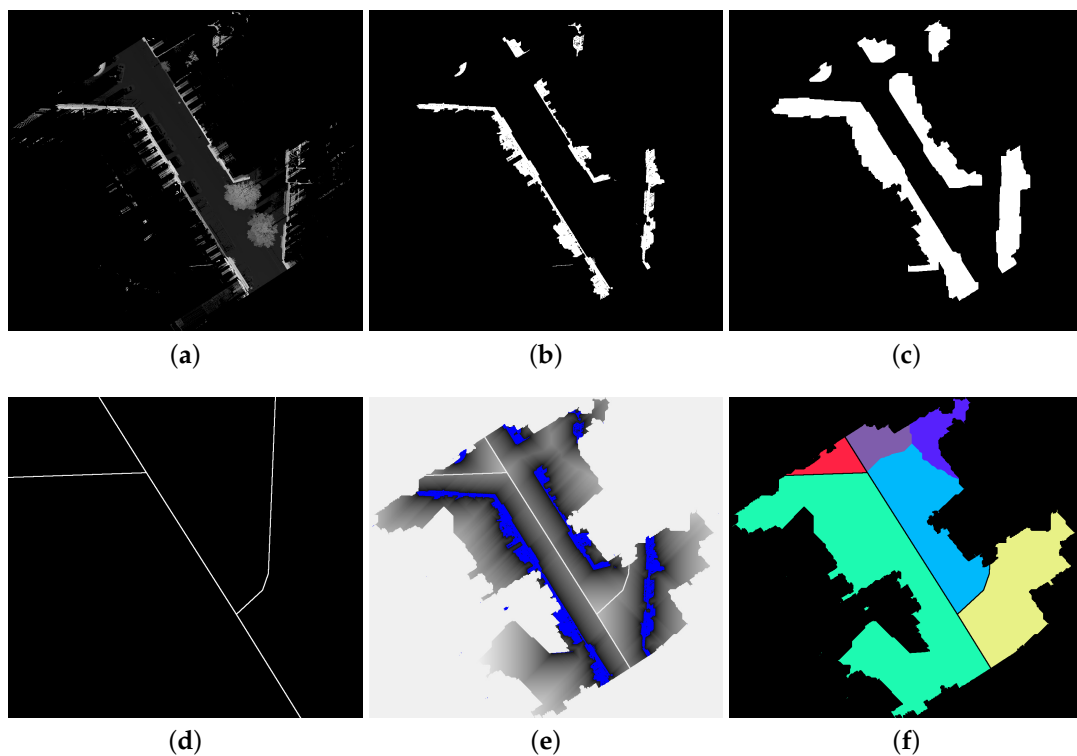

Figure 16. City block segmentation using the influence zones of the façade. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Elevation image; (**b**) Binary image containing the façade segmentation result; (**c**) Closing of size S_{min} to reconnect near façades; (**d**) Medial road axes (used if available); (**e**) Distance function from façades (blue). Both empty pixels and medial road axes are constrained to have the maximal distance; (**f**) Influence zones. Each color represents a different city block.

Figures 17 and 18 present two city block segmentation results reprojected onto the 3D point cloud. In those experiments, façades have been segmented using the elongation based method explained in Section 5.
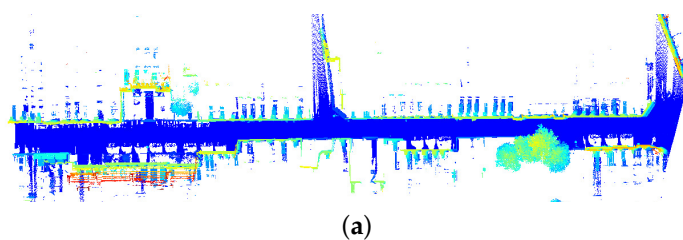

(**a**)

Figure 17. *Cont.*
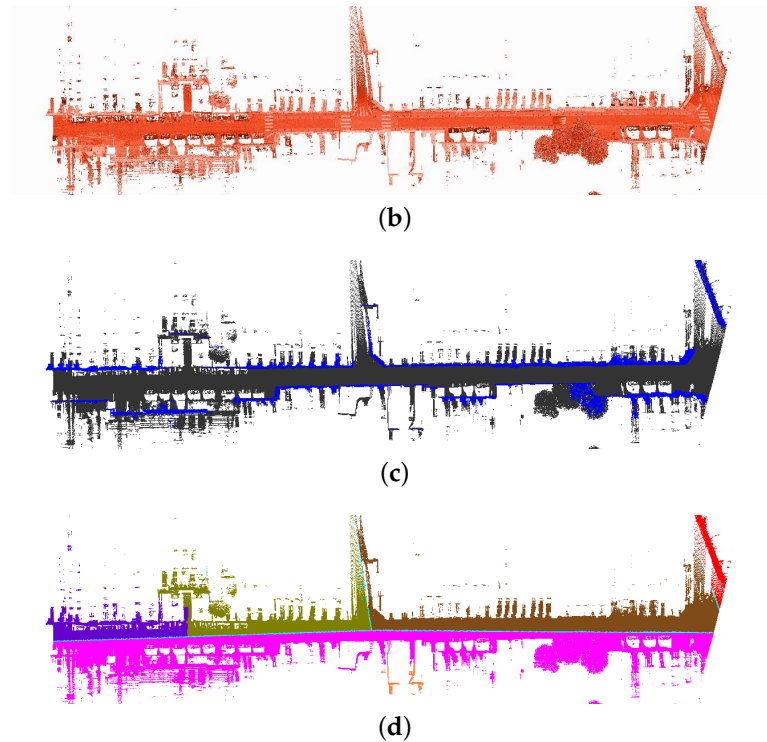
(**b**)

(**c**)

(**d**)

Figure 17. City block segmentation using the influence zones of the façade. Reprojection onto the 3D point cloud. Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©. (**a**) Point cloud colored by the Z coordinate; (**b**) Point cloud colored by the laser intensity; (**c**) Façade segmentation using the elongation image (method without markers). Façade (blue) and other (gray); (**d**) City block segmentation. Each color represents a different city block.

(**a**)

(**b**)

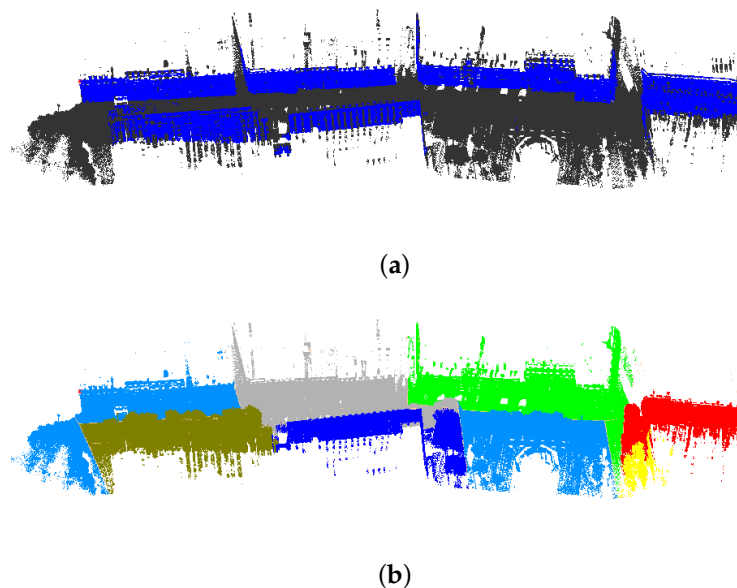**Figure 18.** City block segmentation using the façades influence zones. Reprojection onto the 3D point cloud. Test site in *rue Bonaparte* and *St. Sulpice* square in Paris, France. Stereopolis II, IGN©. (**a**) Façade segmentation using the elongation image (method without markers). Façade (blue) and other (gray); (**b**) City block segmentation. Each color represents a different city block.

7. Results

Our façade segmentation methods have been tested on TerraMobilita datasets in order to get qualitative and quantitative results [4]. Two types of ground truth (GT) annotations are available:

- 2D lines indicating the separation between sidewalks and buildings. These 2D manual annotations are usually provided by local authorities. In our case, they have been obtained from Open Data Paris [48], a project from Paris city hall (*Mairie de Paris*, in French) in order to make urban data available to the community. Evaluations using 2D lines are commonly used in the state of the art when 3D annotations are not available [3,49,50]. These evaluations give an idea on the segmentation method performance. However, results should be carefully interpreted since the evaluation is only carried out on the 2D space at the ground level, the performance segmenting 3D features such as tilted façades, architectural details and balconies cannot be directly evaluated, as illustrated in Figure 19.
- 3D point-wise annotations, *i.e.*, a class is assigned to each 3D point. These point-wise annotations allow a global evaluation taking all façade points into account. In our opinion, this evaluation is the most appropriate, however complete 3D manual annotations are rarely available in the state of the art. In our experiments, we have used 3D annotations and evaluation methods developed in the framework of TerraMobilita/iQmulus benchmark [51].

Evaluations based on each type of GT annotation are presented in the two following subsections.

7.1. Results: Evaluation Using Open Data Paris

We have chosen 4 MLS datasets from 6th Parisian district in France. Data have been acquired by Stereopolis II [52], a MLS system by the National French Mapping Agency (IGN—In French *Institut National de l'Information Géographique et Forestière*). Table 1 describes the datasets used in this evaluation.

Table 1. Datasets used for evaluation of our façade segmentation methods. All datasets have been acquired in January 2013 by Stereopolis II system in the 6th Parisian district, France.

	Filename	Description
site I	TerMob2_LAMB93_0020.ply	*rue d'Assas*, approx. 90 m. Short street section with straight façades. See Figure 20
site II	TerMob2_LAMB93_0021.ply	*rue d'Assas*, approx. 80 m. Short street section with two intersections and two big trees. See Figure 21
site III	Cassette_idclass.ply	*rue Cassette*, approx. 200 m. Long street with mainly straight façades and a small tree alignment. See Figure 22
site IV	Z2.ply	*rue Bonaparte* and *St. Sulpice* square, approx. 400 m. Long street section with non straight façades and several tree alignments. See Figure 23

GT annotations have been obtained from Open Data Paris and correspond to 2D lines indicating the junction between sidewalks and buildings at the ground level. Quantitative analyses are performed by comparison between automatic and GT lines on a 2D image. When 3D façade points are projected to a 2D plane, they are usually wider than a single line due to façade inclination, architectural details and balconies. Therefore, buffers around GT lines and segmented façades are required in order to evaluate the results.

On the one hand, a segmented façade is labeled as true positive or false positive if it is located inside or outside a GT buffer, respectively. On the other hand, a GT façade is labeled as segmented or missed if it is located inside or outside a segmented façade buffer, respectively.

In our datasets, we consider a buffer width of 1.0 m, which corresponds to the width of a typical Haussmannian balcony in Paris. This buffer is appropriate to quantify true positives without overestimating false positives, as shown in Figure 19. This buffer-based evaluation is commonly used in other works reported in the literature [3,49,50].
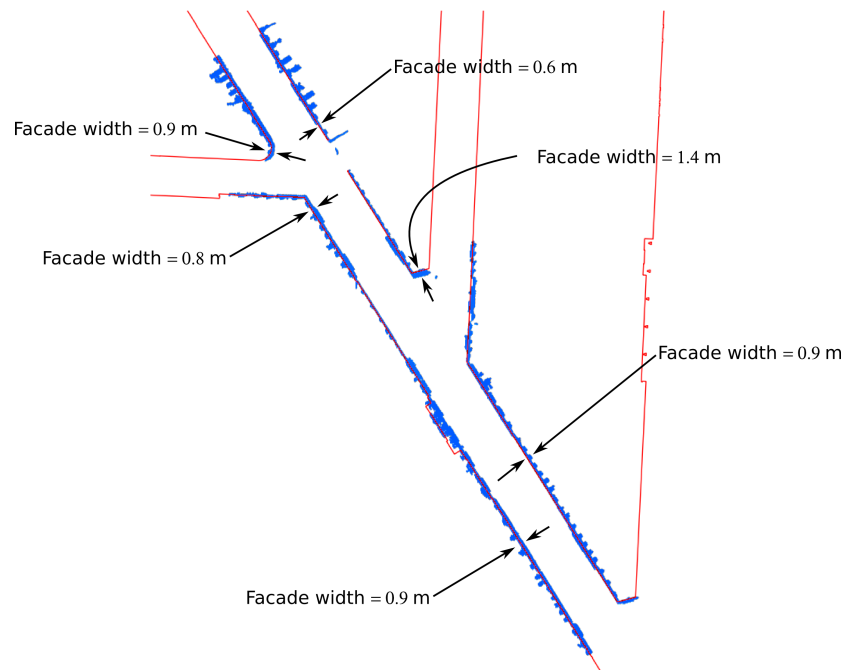

Figure 19. Ground truth lines and 3D façade points projected onto the 2D plane. GT lines (red) and segmented façades (blue). Test site in *rue d'Assas* in Paris, France. Stereopolis II, IGN©.

The classic Precision (P), Recall (R) and f_{mean} criteria are computed. Recall is defined as the number of GT pixels correctly segmented divided by the total number of GT pixels; Precision is defined as the number of true positive pixels divided by total number of segmented pixels (true positives + false positives); and $F_{mean} = 2PR/(P+R)$.

Table 2 presents a quantitative comparison between our façade segmentation methods. As aforementioned, these results should be carefully interpreted since the evaluation is only carried out on the 2D space at the ground level. Then, the performance segmenting inclined façades, façades with architectural details and balconies cannot be directly quantified.

Figures 20–23 present elevation images, GT annotations and segmentation results on our four test sites. It is noteworthy that several ground truth annotations are missing. For example, in site III (Figure 22b) several façades in the street left side are missing, while in site IV (Figure 23b) façades behind trees in the street right side have not been annotated. Therefore, several correct segmentations have been incorrectly labeled as false positives. As a result, the performance of our methods is under-estimated in these two test sites. Let us analyze results of each method in detail.

Method 1, based on reconstruction by dilation from markers, presents the highest Recall retrieving 100% of façades in the four test sites. However, this method presents also the highest number of false positives (Precision ranges between 13.6% and 45.1% for all test sites). As aforementioned, this method is based on iterative geodesic dilation, then any object touching the façade is segmented as part of it. This method is the fastest one and its use may be justified in an application with strict time constraints or if only a rough segmentation is required. For example, if we are only interested in defining the public space boundary (e.g., for a urban mobility application), all objects touching or behind the façade are not required to be segmented.

Table 2. Quantitative comparison between our façade segmentation methods on 4 test sites.

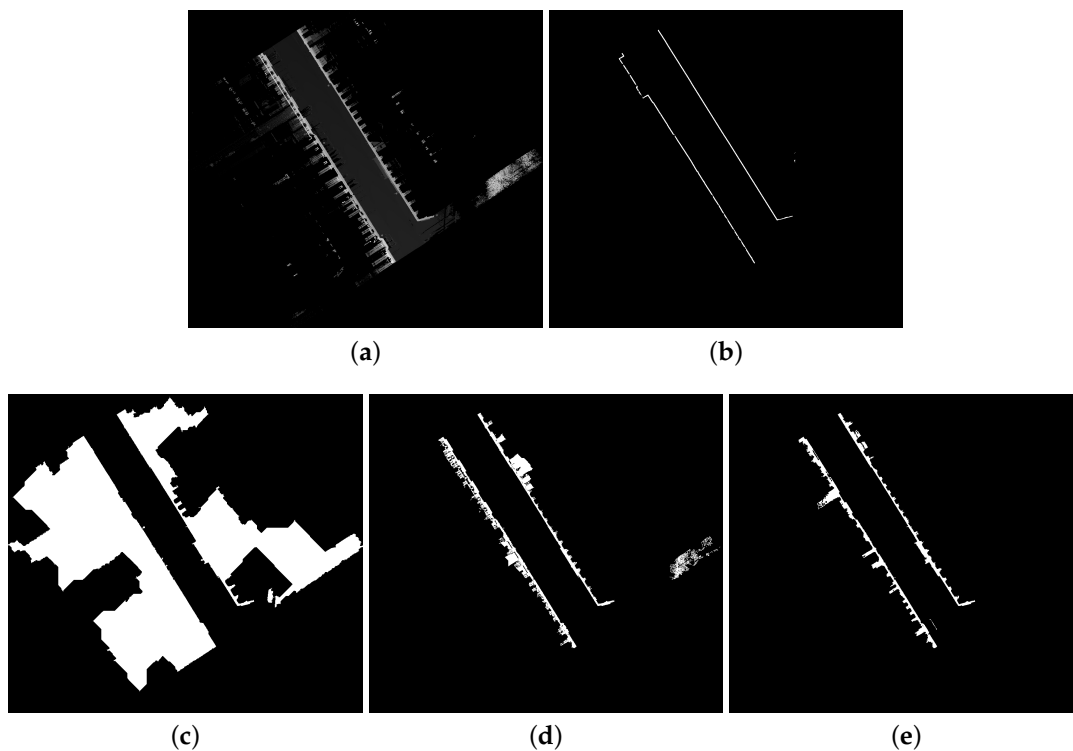|	Method 1: Reconstruction by Dilation (Section 4, Definition 3)	Method 2: Attribute Controlled Reconstruction (Section 4, Definition 4)	Method 3: Maximal Elongation Image (Section 5)
site I	Precision = 45.1% Recall = 100.0% f_{mean} = 62.2% time = 2.2 s	Precision = 73.2% Recall = 99.5% f_{mean} = 84.4% time = 10.7 s	Precision = 80.9% Recall = 99.3% f_{mean} = 89.2% time = 20.0 s
site II	Precision = 47.7% Recall = 100.0% f_{mean} = 64.6% time = 2.1 s	Precision = 87.8% Recall = 97.9% f_{mean} = 92.6% time = 7.8 s	Precision = 92.3% Recall = 93.2% f_{mean} = 92.8% time = 16.5 s
site III	Precision = 27.5% Recall = 99.3% f_{mean} = 43.0% time = 3.5 s	Precision = 44.8% Recall = 98.6% f_{mean} = 61.6% time = 22.5 s	Precision = 67.9 % Recall = 98.2% f_{mean} = 80.3 % time = 44.6 s
site IV	Precision = 13.6% Recall = 100.0% f_{mean} = 23.8% time = 6.8 s	Precision = 13.9% Recall = 99.8% f_{mean} = 27.4% time = 68.6 s	Precision = 64.7% Recall = 88.9% f_{mean} = 74.9% time = 108.4 s

(**a**)   (**b**)

(**c**)   (**d**)   (**e**)

Figure 20. Façade segmentation results for site I (TerMob2_LAMB93_0020.ply). (**a**) Elevation image; (**b**) Ground truth. Façade lines taken from ODParis; (**c**) Façade segmentation using reconstruction by dilation; (**d**) Façade segmentation using attribute controlled reconstruction; (**e**) Façade segmentation using the maximal elongation image.

| (**a**) | (**b**) |

| (**c**) | (**d**) | (**e**) |

Figure 21. Façade segmentation results for site II (TerMob2_LAMB93_0021.ply). (**a**) Elevation image; (**b**) Ground truth. Façade lines taken from ODParis; (**c**) Façade segmentation using reconstruction by dilation; (**d**) Façade segmentation using attribute controlled reconstruction; (**e**) Façade segmentation using the maximal elongation image.
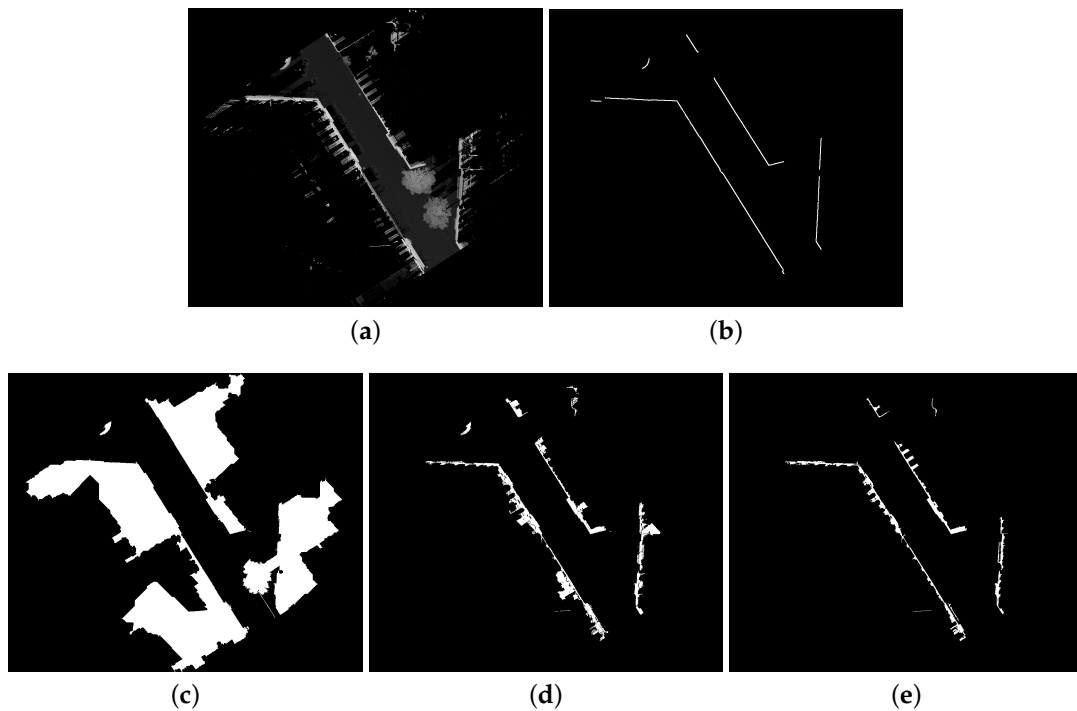

| (**a**) | (**b**) | (**c**) | (**d**) | (**e**) |

Figure 22. Façade segmentation results for site III (Cassette_idclass.ply). (**a**) Elevation image; (**b**) Ground truth. Façade lines taken from ODParis; (**c**) Façade segmentation using reconstruction by dilation; (**d**) Façade segmentation using attribute controlled reconstruction; (**e**) Façade segmentation using the maximal elongation image.
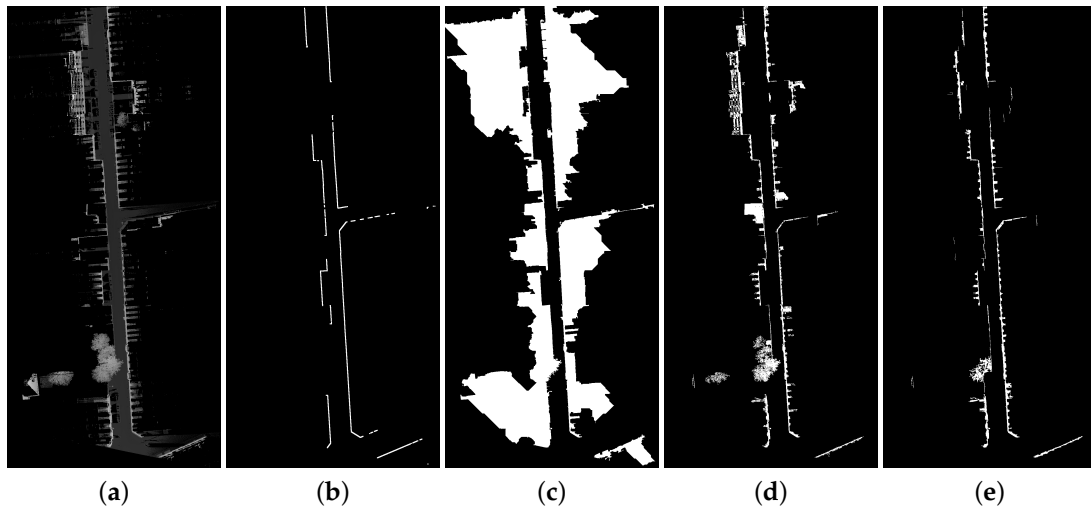
As explained earlier, the main problem of this method is that connected objects, such as motorcycles parked next to the façade or leaning pedestrians, are reconstructed in the façade mask. In order to solve this problem, we have proposed Method 2, based on attribute controlled reconstruction from markers. Since connected objects usually reduce the global façade elongation, this method offers better results than the first one: Recall is higher than 97% for all test sites while

Precision increases up to 73.2% and 87.8% in sites I and II, respectively, sites for which all façades have been annotated. This method presents the best trade-off between performance and processing time.
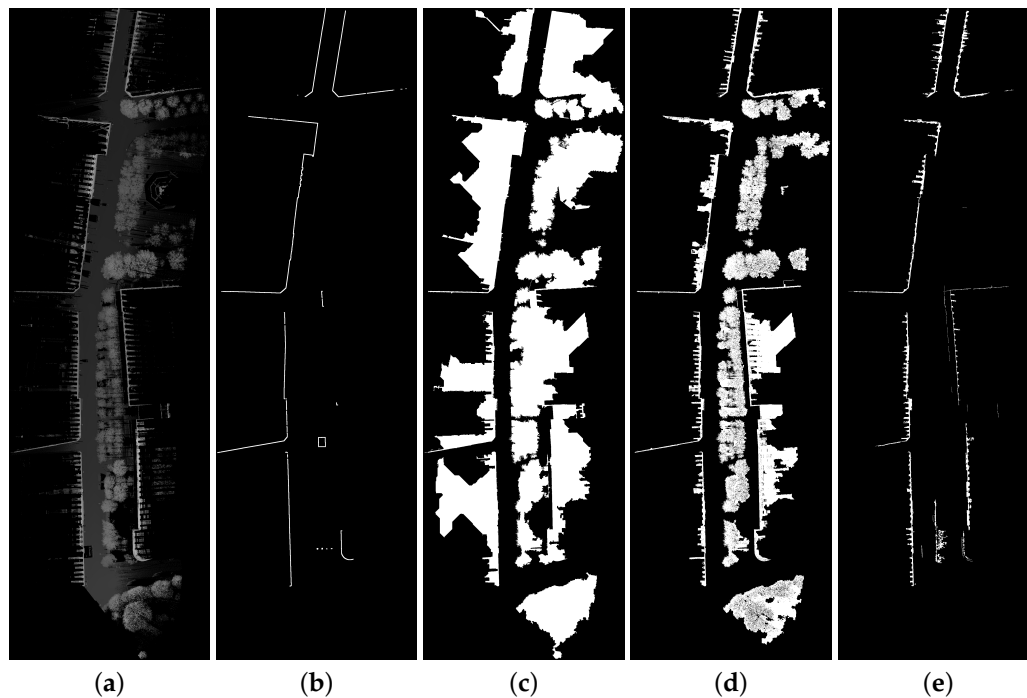

Figure 23. Façade segmentation results for site IV (Z2.ply). (**a**) Elevation image; (**b**) Ground truth. Façade lines taken from ODParis; (**c**) Façade segmentation using reconstruction by dilation; (**d**) Façade segmentation using attribute controlled reconstruction; (**e**) Façade segmentation using the maximal elongation image.

In general, methods based on façade markers are strongly influenced by markers extraction. The main drawback is that bad located markers may produce errors reconstructing non-façade objects. Particularly in site IV, marker-based methods fail segmenting the tree alignments (Precision is 13.6% and 13.9% for methods 1 and 2, respectively). Method 3 does not require markers and leads to a more robust segmentation based on the maximal elongation image. This method produces the best results for all test sites: f_{mean} equal to 80.9% and 92.8% for sites I and II. In spite of missing GT annotations, f_{mean} is equal to 80.3% and 74.9% for sites III and IV, proving the performance of this method even in the presence of trees. The main drawback is that its implementation is slower, then it is not suitable for real-time applications. However, time computing is reasonable for large scale applications, where time constraints are less strict. Note that processing time is only a few tens of seconds for an acquisition of several hundreds of meters, using a non-optimized implementation.

7.2. Results: TerraMobilita/iQmulus Database

TerraMobilita/iQmulus database [51] has been created with the aim of benchmarking semantic analysis methods working on 3D dense urban data. This database has been created in the framework of TerraMobilita project. It consists in 11 annotated 3D point clouds acquired by Stereopolis II system in the 6th Parisian district in January 2013. Annotation has been carried out in a manually assisted way by MATIS laboratory at IGN.

Our evaluation consists of one of those point clouds. For this experiment, file "Cassette_idclass.ply" has been used. The manual annotated 3D point cloud is available at [51] and the 3D point cloud processed by our method is available at [53]. It contains 12 million points from a street section approximately 200 m long in *rue Cassette* in Paris, France. Manual annotations and

point-wise evaluations have been independently carried out by the IGN. Results of this benchmark have been previously published in [54].

Figure 24 presents the façade segmentation result projected onto the 3D point cloud. In this experiment, only our method based on the maximal elongation image has been applied. As a general remark, errors of our segmentation method are due to an incomplete detected façade (zone A) and a tree alignment connected to a low wall (zone B).
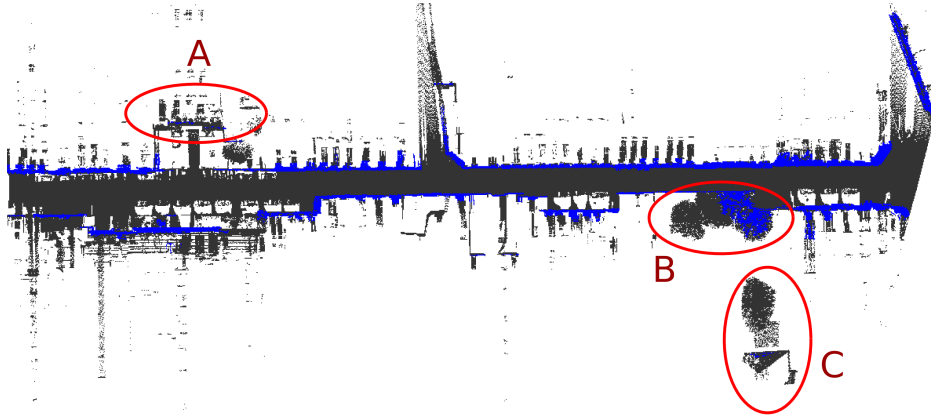

Figure 24. Façade segmentation result using the maximal elongation image on "Cassette_idclass.ply" file. Ground (gray), façades (blue). Input file taken from TerraMobilita/iQmulus database. Stereopolis II, IGN©. Errors are due to an incomplete detected façade (**zone A**), a tree alignment connected to a low wall (**zone B**) and some non-annotated points (**zone C**).

As cited previously, our results are evaluated point-by-point using the TerraMobilita/iQmulus evaluation protocol [51]. First, we classify the 3D point cloud in two main categories: *surface* (containing façades and ground) and *other*. Moreover, the *unclassified* category has been defined for non-annotated points in the GT, which are ambiguous points difficult to annotate. They correspond to 18.31% of total number of points in the dataset. For example, consider the tree and the wall in zone C in Figure 24. These points have been manually marked as *unclassified*, then they have not been taken into account in the evaluation.

Table 3 presents the confusion matrix and our classification results for these two categories. Results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points). Segmentation has been carried out using the method without markers (Section 5). This classification is useful to evaluate the ability of our method segmenting surfaces (façades and ground) while separating other objects connected to them.

Table 3. Classification results on TerraMobilita/iQmulus database. GT: ground truth, AR: automatic result. Segmentation using the method without markers (Section 5).

GT/AR	Unclassif.	Surface	Other	Sum	Recall	Precision	f_{mean}
unclassif.	-	-	-	18.31%	-	-	-
surface	1.90%	70.81%	3.10%	75.81%	93.40%	98.82%	96.03%
other	0.09%	0.85 %	4.94%	5.88%	84.16%	61.52%	71.08%
Sum	1.99%	71.66%	8.04%	100.0%	**Overall accuracy: 92.73%**

Using our method, f_{mean} for the *surface* class is equal to 96.03% while *other* are correctly separated from them with f_{mean} equal to 71.08%. In this experiment, we are mainly interested in separating façades and ground from other structures such as connected objects. Note that the *surface* class includes façades and ground, which represent the biggest categories in the scene with 75.81% of

total 3D points, while the *other* class represents 5.88% of total 3D points. The overall accuracy of our method considering these categories is 92.73%.

Table 4 presents our segmentation results for the *surface* class (façades and ground). In the confusion matrix, results are presented as percentages with respect to the total number of points in the 3D point cloud (12 million points). Note that our method correctly separates façades and ground giving F_{mean} equal to 97.25% and 98.72%, respectively. Figure 25 shows that small errors are due to the façade-ground junction, where some points may be wrongly assigned. The overall accuracy in this case is 98.26%. These results prove the performance of our method.

Table 4. Classification results for surface class (façades and ground) on TerraMobilita/iQmulus database. GT: ground truth, AR: Automatic result.

GT/AR	Ground	Façade	Sum	Recall	Precision	F_{mean}
ground	30.77%	0.01%	30.78%	99.96%	94.69%	97.25%
façade	1.73%	67.49%	69.22%	97.51%	99.98%	98.72%
Sum	32.50%	67.50%	100.0%	**Overall accuracy: 98.26%**		

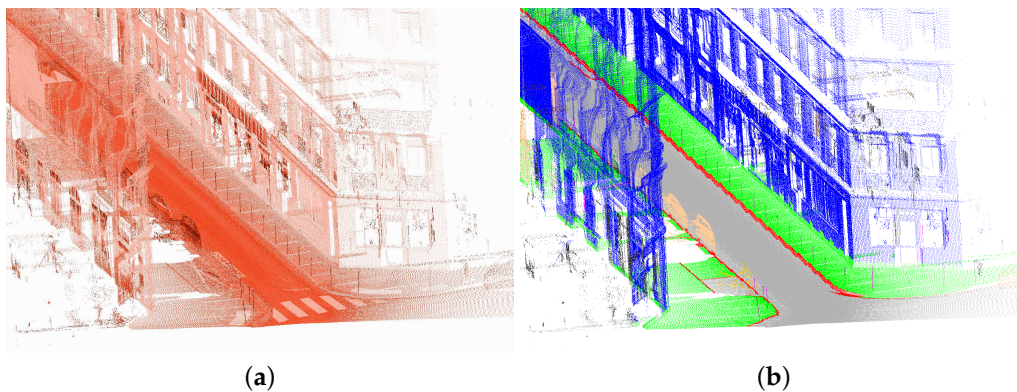(**a**)　　　　　　　　　　　　　　　　　　　　　　　(**b**)

Figure 25. Minor errors in the façade–ground junction (between blue and green colors). Acquisition by IGN©France. (**a**) Point cloud colored by the laser intensity; (**b**) Segmentation result projected onto the 3D point cloud. Façade (blue), sidewalk (green), road (gray), curbs (red), pole-like objects (magenta), other (pink).

8. Conclusions

In this paper, we have proposed automatic and robust approaches to segment façades from 3D point clouds. Processing is carried out using elevation images and 3D decomposition, and the final result can be reprojected onto the 3D point cloud for visualization or evaluation purposes.

Our methods are based on geometrical and geodesic constraints. Parameters are related to urban and architectural constraints. Thus, they are intuitive to tune. The performance of our methods have been proved in our experiments on TerraMobilita databases using 2D and 3D ground truth annotations. Three approaches have been proposed: (i) reconstruction by dilation from markers; (ii) attribute controlled reconstruction from markers and; (iii) maximal elongation image from 3D decomposition (without markers). The method based on reconstruction by dilation from markers is the fastest one since it is based on simple thresholds and use reconstructions constrained by the ground residue in order to get the entire façade. The main problem is that objects connected to the façade are reconstructed as well. In order to solve this problem, we have proposed an attribute controlled reconstruction using the geodesic elongation. Since connected objects usually appear at low heights and reduce the global façade elongation, this method offers better results than the first one. In our experiments, we have used geometric and geodesic constraints in order to extract façade

markers. In the case of low façades or when the laser sensor is oriented to the ground, additional markers have been proposed.

In general, methods based on façade markers are strongly influenced by the markers extraction method. The main drawback is that bad located markers produce errors since they may reconstruct non-façade objects. For this reason, we have proposed a more robust method avoiding the use of façade markers. In such method, both the elongation and its evolution over the height decomposition of the scene are analyzed. This method is based on the maximal elongation image computed from 3D decomposition. It has been proved to produce the best results.

Since the three proposed methods are based on geometric and geodesic constraints, they can be set up to manage façades of any length, height and elongation. However, these methods offer different results and level of detail. In that sense, the selection of the best segmentation method should be a trade off between quality of the results and computational cost. This choice remains application dependent: On the one hand, for applications with strict time constraints where only rough results are required, the user should prefer the first method based on façade markers and morphological reconstruction by dilation. However, in the presence of objects touching the façades, such as low bushes or parked motorcycles, the user should prefer the second method based on attribute controlled reconstruction. On the other hand, for mapping applications, where the quality of results is of the utmost importance while keeping a reasonable computation time able to address large-scale applications, the third method based on the elongation image should be preferred. This method is the most accurate and appropriate to manage high objects, such as trees, connected to the façade.

Our approach is a research prototype, mainly based on Morph-M library [55], the image processing library of our laboratory. This library allows fast prototyping but it is not intended to be a fast library. Optimization of our base operators (erosion, dilation, opening, reconstruction, watershed, and so on), currently under development at CMM, aims at providing optimized operators for real time and/or big image developments. Software (hierarchical queues, structuring elements decomposition, among others) and hardware (SIMD-Single Instruction Multiple Data and parallelization) optimizations are available in SMIL library [56] and will be integrated in our future developments.

Author Contributions: Andrés Serna is the leading author of this work. He performed the research, developed the software prototype and wrote the paper together with Beatriz Marcotegui. The paper was further revised by Jorge Hernández who gave substantial contributions to the design and analysis of this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vosselman, G.; Maas, H.G. *Airborne and Terrestrial Laser Scanning*; Whittles Publishing: Caithness, UK, 2010; p. 318.
2. Hernández, J.; Marcotegui, B. Point cloud segmentation towards urban ground modeling. In Proceedings of the 5th GRSS/ISPRS Joint Urban Remote Sensing Event (URBAN 2009), Shanghai, China, 20–22 May 2009; pp. 1–5.
3. Serna, A.; Marcotegui, B. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS J. Photogramm. Remote Sens.* **2013**, *84*, 23–32.
4. CapDigital. TerraMobilita: 3D Mapping of Roads and Urban Public Space, Accessibility and Soft Mobility. Available online: http://cmm.ensmp.fr/TerraMobilita/ (accessed on 11 Septembre 2014).
5. Goulette, F.; Nashashibi, F.; Ammoun, S.; Laurgeau, C. An integrated on-board laser range sensing system for on-the-way city and road modelling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *34*, 3–5.

6.　Boulaassal, H.; Grussenmeyer, P.; Tarsha-kurdi, F. Automatic segmentation of building facades using terrestrial laser data. In Proceedings of the ISPRS Workshop on "Laser Scanning 2007 and SilviLaser 2007", Espoo, Finland, 12–14 September 2007; pp. 65–70.

7.　Bab-Hadiashar, A.; Gheissari, N. Range image segmentation using surface selection criterion. *IEEE Trans. Image Process.* **2006**, *15*, 2006–2018.

8.　Hoover, A. University of South Florida (USF) Range Image Database, Available online: http://marathon. csee.usf.edu/range/DataBase.html (accessed on 15 May 2014).

9.　Boyer, K.; Mirza, M.; Ganguly, G. The robust sequential estimator: A general approach and its application to surface organization in range data. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 987–1001.

10.　Werghi, N.; Fisher, R.; Robertson, C.; Ashbrook, A. Modelling objects having quadric surfaces incorporating geometric constraints. In Proceedings of the 5th European Conference on Computer Vision ECCV'98, Freiburg, Germany, 2–6 June 1998; pp. 185–201.

11.　Marshall, D.; Lukacs, G.; Martin, R. Robust segmentation of primitives from range data in the presence of geometric degeneracy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 304–314.

12.　Chaperon, T.; Goulette, F. Extracting cylinders in full 3D data using a random sampling method and the Gaussian image. In Proceedings of the 2001 Vision Modeling and Visualization Conference, Stuttgart, Germany, 21–23 November 2001.

13.　Lari, Z.; Habib, A. An adaptive approach for the segmentation and extraction of planar and linear/cylindrical features from laser scanning data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 192–212.

14.　Demantke, J.; Mallet, C.; David, N.; Vallet, B. Dimensionality based scale selection in 3D LiDAR point clouds. In Proceedings of 2011 ISPRS Workshop on Laser Scanning, Calgary, AB, Canada, 29–31 August 2011.

15.　Hammoudi, K. Contributions to the 3D City Modeling. Ph.D. Thesis, Université Paris-Est, Champs-sur-Marne, France, 2011.

16.　Sevcik, C.; Studnicka, N. Documentation of Complex Facades and City Modelling through the Combination of Laserscanning and Photogrammetry. Available online: http://www.riegl.com/uploads/tx_pxpriegldownloads/Documentation_of_complex_facades_and_city_ modelling_through_the_combination_of_Laserscanning_and_photogrammetry.pdf (accessed on 22 March 2012)

17.　Rutzinger, M.; Pratihast, A.K.; Oude Elberink, S.J.; Vosselman, G. Tree modelling from mobile laser scanning data-sets. *Photogramm. Rec.* **2011**, *26*, 361–372.

18.　Poreba, M.; Goulette, F. RANSAC algorithm and elements of graph theory for automatic plane detection in 3D point clouds. *Arch. Photogramm. Cartogr. Remote Sens.* **2012**, *24*, 301–310.

19.　Serna, A.; Marcotegui, B. Attribute controlled reconstruction and adaptive mathematical morphology. In Proceedings of the 11th International Symposium on Mathematical Morphology (ISMM 2013) , Uppsala, Sweden, 27–29 May 2013; pp. 205–216.

20.　Weinmann, M.; Jutzi, B.; Mallet, C. Feature relevance assessment for the semantic interpretation of 3D point cloud data. In Proceedings of 2013 ISPRS Workshop Laser Scanning, 11–13 November 2013, Antalya, Turkey; pp. 313–318.

21.　Weinmann, M.; Jutzi, B.; Mallet, C. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features. In Proceedings of 2014 ISPRS Technical Commission III Symposium, 5–7 September 2014, Zurich, Switzerland; pp. 181–188.

22.　Shao, H.; Svodoba, T.; Gool, L.V. *ZuBuD—Zurich Building Database for Image Based Recognition*; ETH Zurich: Zurich, Switzerland, 2003.

23.　Hernández, J.; Marcotegui, B. Morphological segmentation of building façade images. In Proceedings of the 16th IEEE International Conference on Image Processing, Cairo, Egypt, 7–10 November 2009; pp. 4029–4032.

24.　Teboul, O.; Simon, L.; Koutsourakis, P.; Paragios, N. Segmentation of building facades using procedural shape priors. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010; pp. 3105–3112.

25.　Serna, A.; Hernández, J.; Marcotegui, B. Adaptive parameter tuning for morphological segmentation of building facade images. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO2012), Bucharest, Romania, 27–31 August 2012.

26. Teeravech, K.; Nagai, M.; Honda, K.; Dailey, M. Discovering repetitive patterns in facade images using a RANSAC style algorithm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 38–53.

27. Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press: Orlando, FL, USA, 1982.

28. Soille, P. *Morphological Image Analysis: Principles and Applications*; Springer: Secaucus, NJ, USA, 2003.

29. Hernández, J.; Marcotegui, B. Filtering of artifacts and pavement segmentation from mobile LiDAR data. In Proceedings of 2009 ISPRS Workshop on Laser Scanning, Paris, France, 1–2 September, 2009; pp. 329–333.

30. Lantuéjoul, C.; Maisonneuve, F. Geodesic methods in quantitative image analysis. *Pattern Recognit.* **1984**, *17*, 177–187.

31. Lantuéjoul, C.; Beucher, S. On the use of the geodesic metric in image analysis. *J. Microsc.* **1981**, *121*, 39–49.

32. Morard, V.; Decencière, E.; Dokládal, P. Efficient geodesic attribute thinnings based on the barycentric diameter. *J. Math. Imaging Vis.* **2013**, *46*, 128–142.

33. Golovinskiy, A.; Kim, V.G.; Funkhouser, T. Shape-based recognition of 3D point clouds in urban environments. In Proceedings of the 12th IEEE International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 2154–2161.

34. Munoz, D.; Vandapel, N.D.; Hebert, M. Onboard contextual classification of 3-D point clouds with learned high-order Markov Random Fields. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, 12–17 May 2009; pp. 2009–2016.

35. Mallet, C.; Bretar, F.; Roux, M.; Soergel, U.; Heipke, C. Relevance assessment of full-waveform LiDAR data for urban area classification. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 71–84.

36. Pu, S.; Rutzinger, M.; Vosselman, G.; Elberink, S.O. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 28–39.

37. Meyer, F. From connected operators to levelings. In *Mathematical Morphology and Its Applications to Image and Signal Processing*; Kluwer Academic Publishers: Berlin, Germany, 1998; Volume 12, pp. 191–198.

38. Serna, A.; Marcotegui, B.; Decencière, E.; Baldeweck, T.; Pena, A.M.; Brizion, S. Segmentation of elongated objects using attribute profiles and area stability: Application to melanocyte segmentation in engineered skin. *Pattern Recognit. Lett.* **2014**, *47*, 172–182.

39. Jones, R. Connected filtering and segmentation using component trees. *Comput. Vis. Image Underst.* **1999**, *75*, 215–228.

40. Pesaresi, M.; Benediktsson, J. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 309–320.

41. Beucher, S. Numerical residues. *Image Vis. Comput.* **2007**, *25*, 405–415.

42. Morard, V.; Decencière, E.; Dokládal, P. Region growing structuring elements and new operators based on their shape. In Proceedings of the 2011 International Conference on Signal and Image Processing (ICSIPA), Kuala Lumpur, Malaysia, 16–18 November 2011, pp. 1–8.

43. Serna, A. Semantic Analysis of 3D Point Clouds from Urban Environments: Ground, Facades, Urban Objects and Accessibility. Ph.D. Thesis, MINES ParisTech, Paris, France, 2014.

44. Wikipedia. City Blocks. Available online: http://en.wikipedia.org/w/index.php?title=City_block&oldid=617153038 (accessed on 8 October 2014).

45. Matheron, G. *Random Sets and Integral Geometry*; John Wiley & Sons: New York, NY, USA, 1975.

46. Serra, J.; Soille, P. *Mathematical Morphology and Its Applications to Image Processing*; Kluwer Academic Publishers: Berlin, Germany, 1994.

47. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *J. Die Reine Angew. Math.* **1908**, *133*, 97–178.

48. Marie de Paris. Open Data Paris. Available online: http://opendata.paris.fr/ (accessed on 22 March 2012)

49. Vosselman, G.; Zhou, L. Detection of curbstones in airborne laser scanning data. In Proceedings of 2009 ISPRS Workshop on Laser Scanning, Paris, France, 1–2 September, 2009; pp. 11–116.

50. Zhou, L.; Vosselman, G. Mapping curbstones in airborne and mobile laser scanning data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 293–304.

51. Brédif, M.; Vallet, B.; Serna, A.; Marcotegui, B.; Paparoditis, N. TerraMobilita/iQmulus Urban Point Cloud Classification Benchmark. Available online: http://data.ign.fr/benchmarks/UrbanAnalysis/ (accessed on 26 June 2014).

52. Paparoditis, N.; Papelard, J.P.; Cannelle, B.; Devaux, A.; Soheilian, B.; David, N.; Houzay, E. Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. *Rev. Fr. Photogramm. Télédétec.* **2012**, *200*, 69–79.

53. Serna, A.; Marcotegui, B. Results of TerraMobilita/iQmulus Urban Point Cloud Classification Benchmark. Available online: https://partage.mines-telecom.fr/public.php?service=files &t=294aed38d48c8ddd03a528069f1b2e51 (accessed on 14 December 2014)

54. Vallet, B.; Brédif, M.; Serna, A.; Marcotegui, B.; Paparoditis, N. TerraMobilita/iQmulus Urban Point Cloud Analysis Benchmark. *Comput. Graph.* **2015**, *49*, 126–133.

55. Project Management Platform of the CMM. Morph-M: Image Processing Software Specialized in Mathematical Morphology. Available online: http://morphm.ensmp.fr (accessed on 8 October 2014).

56. Faessel, M.; Bilodeau, M. SMIL: Simple Morphological Image Library. Séminaire Performance et Généricité, LRDE. Available online: http://cmm.ensmp.fr/ faessel/smil (accessed on 8 October 2014).