TERRAMOBILITA/IQMULUS URBAN POINT CLOUD ANALYSIS BENCHMARK

M. Brédif[1], B. Vallet[1], A. Serna[2], B. Marcotegui[2], N. Paparoditis[1]

[1] Université Paris-Est, IGN Recherche, SRIG, MATIS, 73 avenue de Paris, 94160 Saint Mandé, France
[2] Centre de Morphologie Mathmatique (CMM), 35 rue Saint Honor, 77305 Fontainebleau, France

KEY WORDS: Benchmark, urban scene, laser scanning, mobile mapping, classification, segmentation, detection, analysis

ABSTRACT:

The object of the TerraMobilita/iQmulus 3D urban analysis benchmark is to evaluate the current state of the art in urban scene analysis from mobile laser scanning (MLS). A very detailed semantic tree for urban scenes is proposed (cf Figure 1). We call analysis the capacity of a method to separate the points of the scene into these categories (classification), and to separate the different objects of the same type for object classes (detection). The ground truth is produced manually in two steps using advanced editing tools developed especially for this benchmark. Base on this ground truth, the benchmark will aim at evaluating both the classification, detection and segmentation quality of the submitted results.

1 INTRODUCTION

Nowadays, LiDAR technology (Light Detection And Ranging) has been prospering in the remote sensing community. We can find several developments such as: Aerial Laser Scanning (ALS), useful for large scale buildings survey, roads and forests; Terrestrial Laser Scanning (TLS), for more detailed but slower urban surveys in outdoor and indoor environments; Mobile Laser Scanning (MLS), less precise than TLS but much more productive since the sensors are mounted on a vehicle; and more recently, “stop and go” systems, easily transportable TLS systems making a trade off between precision and productivity.

Thanks to all these technologies, the amount of available 3D geographical data and processing techniques has bloomed in recent years. Many semi-automatic and automatic methods aiming at analyzing 3D urban point clouds can be found in the literature. It is an active research area. However, there is not a general consensus about the best detection, segmentation and classification methods. This choice is application dependent. One of the main drawbacks is the lack of publicly available databases allowing benchmarking.

In the literature, most available urban data consist in close-range images, aerial images, satellite images but few MLS datasets. Moreover, manual annotations and algorithm outputs are rarely found in available 3D repositories. Some annotated 3D MLS datasets publicly available are the Oakland 3D point cloud dataset **?**, and the Paris-rue-Madame dataset Serna et al. (2014)

In this context, this paper presents a benchmark that aims at stimulating researchers from different fields such as Computer Vision, Computer Graphics, Geomatics and Remote Sensing, working on the common goal of processing 3D data, benchmarking segmentation and classification methods working on 3D MLS data. This will provide a ground for cross-fertilization and discussions on the future challenges in this important research area. More information about this benchmark are available on the webpage: `http://data.ign.fr/benchmarks/UrbanAnalysis/index.html`

2 EXPERIMENTAL DATASET

The database contains 3D MLS data from a dense urban environment in Paris (France), approximately 10 km long. The acquisition was made in January 2013.

Figure 1: 3D View of the dataset.

This database is produced in the framework of the iQmulus and TerraMobilita projects. It has been acquired by Stereopolis II, a MLS system developed at the French National Mapping Agency (IGN). Annotation will be carried out in a manually assisted way by the MATIS laboratory at IGN.

In this database, the entire 3D point cloud is segmented and classified, i.e. each point contains a label and a class. Thus, point-wise evaluation of detection-segmentation-classification methods becomes possible. The datasets and their processing results must be presented in PLY format with little endian encoding. All coordinates are geo-referenced (E,N,U) in Lambert 93 and altitude IGN1969 (grid RAF09) reference system, reflectance is the laser intensity. An offset has been subtracted from the XY coordinates with the aim of increasing data precision: X0 = 649000 m and Y0 = 6840000 m. Each vertex contains the attributes presented in table 1.

3 ANALYSIS PROBLEM STATEMENT

The problem adressed by this benchmark is to perform a point-wise segmentation and classification. Each processed file, provided by each participant, must be a PLY file containing the original points (in the original order), their original attributes and two additionnal attributes: id and class. All the 3D points belonging to the same object will have the same object identifier (id). Thus, the number of different ids in the 3D point cloud corresponds to the number of objects. In the classification step, a category is assigned to each segmented object. Each class represents a se-

Type	Properties	Description
`float32`	x,y,z	Mesured position
`float32`	x0,y0,z0	Sensor position
`float32`	reflectance	backscattered intensity corrected for distance
`uint8`	num_echo	number of the echo (to handle multiple echoes)
`uint32`	id	object identifier in the segmentation
`uint32`	class	class label assigned to its segmented object.

Table 1: Vertex Properties. Positions are expressed in the Lambert 93 system. Two points having the same id must have the same class. Since each point of the dataset contains an id and a class, the evaluation will be carried out in a point-wise way.

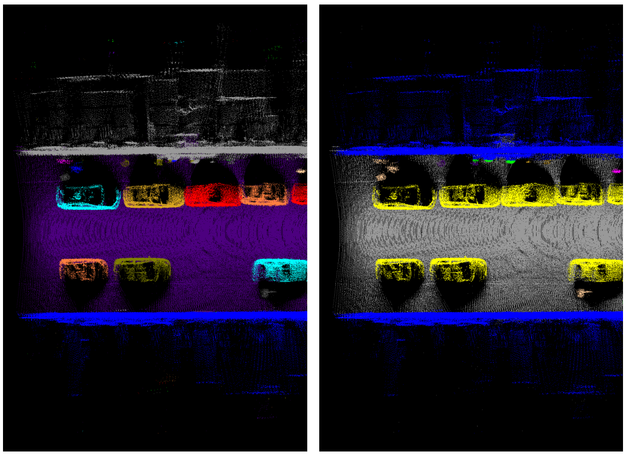


Figure 2: (left) Ids. (right) Classes.

mantic urban entity, as defined in the following subsection. If two points have the same label, they must have the same class.

Figure 2 shows an example of labels and classes on a 3D point cloud. In the left image, note that ground, facades, pedestrians and cars are represented with different colors because they are different objects and have different labels. In the right image, the colors represent the object classes: ground in gray, facades in blue, cars in yellow, pedestrians in skin color, and furniture in red and cyan.

3.1 Classification Ontology

In this contest, a hierarchy of semantic classes has been defined. The class tree (figure 3) is downloadable as an xml file from `http://data.ign.fr/benchmarks/UrbanAnalysis/download/classes.xml` and is composed as follows:

Surface Surface of unbounded or very large objects

- **Ground** Ground surface
- **Building** All points on the outside of the building
- **Other Surface** Surface of unbounded objects that does not fit in one of the following categories

Object All semantic objects

- **Static** Objects that are not made to be easily moved
- **Dynamic** Individuals or objects that can move easily
- **Natural** Natural objects or vegetation
- **Other Object** Objects that do not fit in the other categories

Other Undefined points, outliers, points inside buildings...

Unclassified Not classified yet

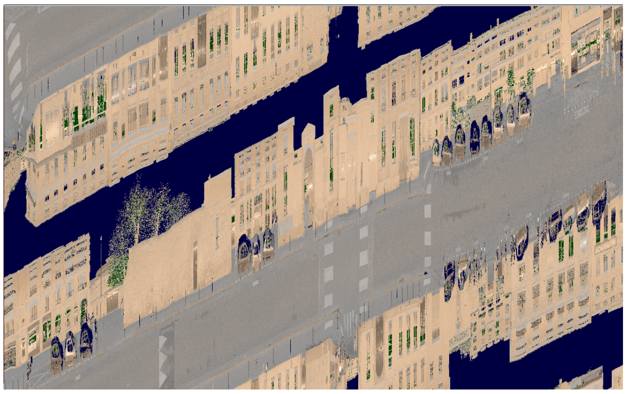


Figure 4: Lidar point cloud viewed in sensor-space : horizontal axis corresponds to time and vertical axis corresponds to rotation angle.

The tree is voluntarily very detailed as we aim at producing a ground truth that can be usefull to a wide range of methods. Participants can choose the classes that they want in this tree and the evaluation will be performed accordingly. The **Unclassified** label is used to only focus the evaluation on the portion of the point cloud that has been classified only. **Other X** labels are scattered throughout the hierarchy so that the classifications may differentiate between a classification that did not try to distinguish among the more specialized child labels of an inner node of the classification hierarchy, and a classification that expresses that the relevant label is not part of the classification hierarchy.

4 SEMI-AUTOMATIC GROUND TRUTH PRODUCTION

As soon as any form of algorithm is used to produce the Ground Truth that will be used to evaluate a classification result, this particular algorithm will bias the results to favor similar approaches. That is why we have strived to propose an approach that is as manual as possible, thereby reducing algorithmic bias, while providing an editing tool that enables an efficient segmentation and classification of the benchmark dataset. Therefore, we have set up a semi-automatic approach where the user has a full control over the resulting segmentation and classification.

4.1 Segmentation

Navigating and selecting through a point cloud is a counter-intuitive task, due to the absence of proper occlusion and its sparse nature. As the segmentation was to be performed at the point level rather than alternative representations such as bounding boxes, an efficient browsing and segmentation of the point cloud was a key issue. We tackled this problem by proposing an interface that shows the pointcloud in sensor space (figure 4). This was made possible and convenient by the geometry of acquisition and the availability of its two parameters in the raw dataset : the constant time step dt_{pulse} between each emitted pulse and the constant time step dt_{turn} between each rotation of the lidar. Please not that we do not require that each turn is composed of an integral number of pulses, which yield overall shear in figure 4. This parameter allow to recover a regular topology out of the point cloud stream : the pulse neighbors are the immediately preceding

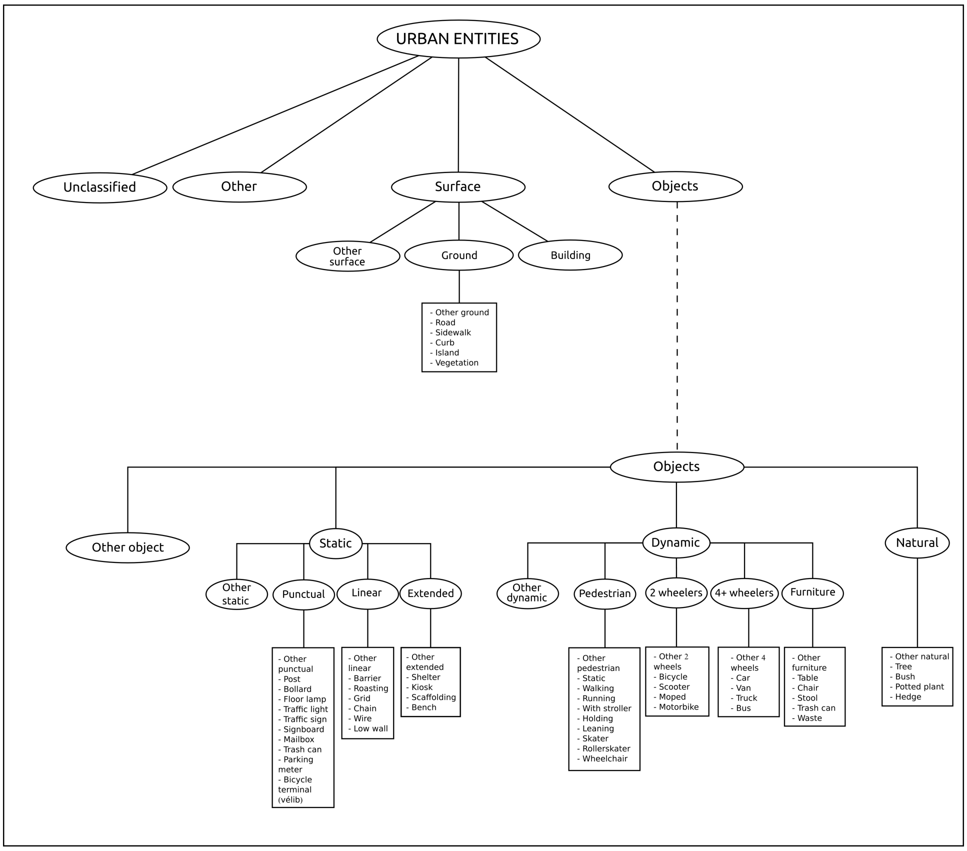


Figure 3: Hierarchy of class annotations.

and succeeding pulses and the closest ones on the preceding and succeeding rotations. The recovery of such a 2D topology with 4-connectivity enables the use of traditionnal image processing techniques. A particularity is on its topological boundary : it is composed of two topological circles corresponding to the pulses of the first and last rotations. This is due to the continuous sampling of the particular scanner used to acquire this dataset : the last pulse of a rotation is connected to the first pulse of the following one.

In this 2D sensor-space, it is then trivial to create, maintain a 2D segmentation by providing to the user a graph edition tool : node creation and deletion, edge creation between existing nodes, polyline input, insertion of a node by splitting an edge... In order for this graph to define a partition of the sensor measurements, nodes are placed at pixel corners and edges are manhattan curves along pixel boundaries. The user experience is improved by allowing zooming, panning, and snapping.

The only automatic task used in this editing tool applies when a user adds an edge between two nodes or when one is modified by a moving end node. Two modes are available to rasterize a path between the two end nodes of the path on the pixel boundaries :

Sensor-space Line This is a simple rasterization of the sensor-space line segment as a Manhattan curve passing through pixel corners.

Optimal Path This mode performs an $A*$ search to provide the rasterized path which is optimal with respect to the sum of weights of the traversed pixel boundaries. This $cost$ is expressed in terms of the absolute differences of the measured quantities $\Delta_\alpha, \Delta_d, \Delta_r$, where α is the angle between the normal and the vertical direction, d is the measured distance and r is the measured reflectance, and $\lambda, \alpha_0, d_0, r_0$ are empirical weights :

$$cost = \max\left(1, \lambda\left(1 - \max\left(\frac{\Delta_\alpha}{\alpha_0}, \frac{\Delta_d}{d_0}, \frac{\Delta_r}{r_0}\right)\right)\right) \quad (1)$$

The user experience is improved by preprocessing the shortest path from the selected node as a background process, so that the optimal path may be updated efficiently as the mouse moves. The user click then only validates the already optimized and displayed path, increasing the productivity.

4.2 Classification

As the number of objects is limited compared to the massive size of the point cloud, assignation of class labels to an already segmented lidar point cloud is performed by a paint-bucket-like interface : the user selects a class label from a drop-down list of available labels and then clicks on the pointcloud to assign this label to the segmented object id. The display mode may be either set to segmentation mode that applies a random palette according to the segment ids or a classification mode where the visualization depends on the class label. To further help this process, points to be rendered may be filtered base on there class labels.

5 SUBMISSION EVALUATION

The proposed semantic tree is very detailed and probably no existing method treats the whole problem. This is why, the participants to the benchmark can choose to analyse the scene using any subtree of the tree. In this case, they will simply apply the "other"

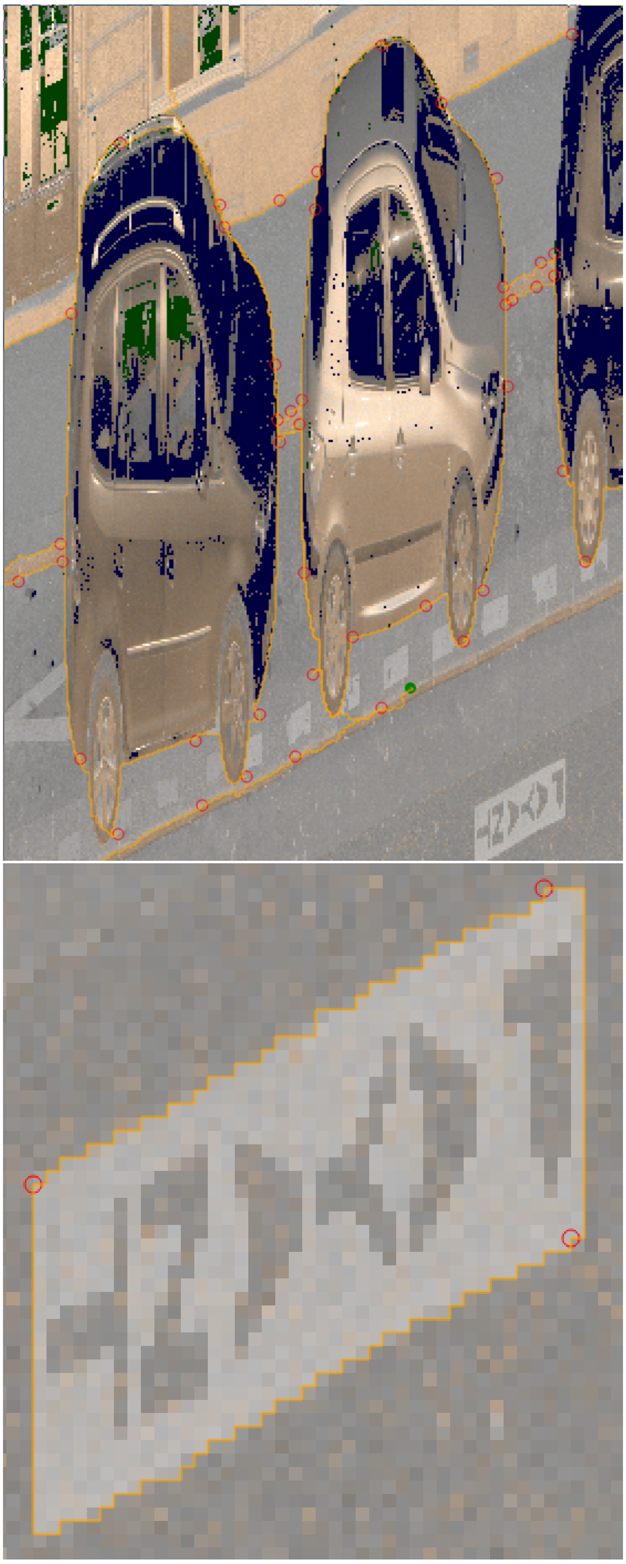


Figure 5: Segmentation-inducing graph.

label to the nodes that they do not wish to detail. The evaluation will be performed accordingly and only the relevant metrics will be given.

The benchmark does not aim at ranking the participants but at providing insights on the strengths and weaknesses of each methods. We consider that the quality of a method is subjective and application dependent, and the results of this benchmark should only help a user choosing one approach depending on its own specific requirements. Results will be evaluated at three levels: as a classification, as a detection and as a segmentation. Details of the evaluation metrics used are given in the Evaluation Protocol document.

5.1 Classification quality

The classification quality will be evaluated point-wise. The results of the evaluation will be a confusion matrix for each node of the tree that the evaluated method handles. Rows and lines will be the subclass labels from the ground truth and the evaluated method respectively, and matrix values are the percentage of points with the corresponding labels. All nodes from the semantic tree have an "other" class, so participants can classify into less classes than what is given in the tree. For non root nodes, an additional label "not in class" will be given for point that were not classified correctly at a lower level.

5.2 Detection quality

The detection quality work measures the capacity of the method to detect the objects present in the scene. Thus it requires to choose a criterion to decide if an object from the ground truth is detected or not. This biases the evaluation as this choice will impact the result. The solution that we propose is to give the evaluation result for a varying threshold m on the minimum object overlap. In this benchmark, an object is defined by the subset of points with the same object identifier. For a such subsets S^{GT} of the ground truth and S^{AR} of the evaluated algorithm result, we will validate S^{AR} as a correct detection of S^{GT} (a match) iff:

$$\frac{|S^{GT}|}{|S^{GT} \cup S^{AR}|} > m \text{ and } \frac{|S^{AR}|}{|S^{GT} \cup S^{AR}|} > m \tag{2}$$

where $|\cdot|$ denotes the cardinal (number of objects) of a set. The standard precision/recall are then functions of m:

$$precision(m) = \frac{\text{number of detected objects matched}}{\text{number of detected objects}}$$

$$recall(m) = \frac{\text{number of detected objects matched}}{\text{number of ground truth objects}}$$

Precision/Recall will be evaluated for each object types at each level of the semantic tree that the participants have handled and results will be presented as two curves. Precision/Recall are decreasing in m and this decay indicates the geometric quality of the detection (good geometry implies slower decay).

5.3 Segmentation quality

When the threshold m is below 0.5, the criterion (2) does not guarantee that objects are uniquely matched. When $m < 1/n$, n objects from the ground truth (GT) can be matched to a single object of the algorithm result (AR), or the opposite. Thus for $m < 0.5$ we will also give the curves of over-segmentation (1-to-n) and under-segmentation (n-to-1) by averaging n over the matches defined by (2).

6 PARTICIPANTS

For the moment, two participants have shown interest, but the benchmark will stay open at least until the end of 2014.

6.1 IPF - KIT

The institute of Photogrammetry and Remote Sensing (IPF) from the Karlsruhe Institute of Technology (KIT) will participate with a method that they present as follows:

We propose a new methodology for large-scale urban 3D scene analysis in terms of automatically assigning 3D points respective semantic labels. The methodology focuses on simplicity and reproducibility of the involved components as well as performance in terms of accuracy and computational effort. Exploiting a variety of low-level geometric features and considering recent advancements in order to improve their distinctiveness, the methodology is in principal designed to process point clouds with a few millions of 3D points. For analysing huge 3D point clouds with possibly billions of points for a whole city like Paris, however, an adaptation has to be introduced. For this purpose, we propose an adaptation which is based on a tiling of the scene and thus allows a successive processing in reasonable time without affecting the quality of the classification results. We demonstrate the performance of our methodology on two adequate, labelled 3D point cloud datasets with respect to robustness, efficiency and scalability.

For further details, the reader is encouraged to review Weinmann et al. (2013).

6.2 CMM - MINES ParisTech

The Centre de Morphologie Mathématiques (CMM) from MINES ParisTech will participate with a method that they present as follows:

Our method is based on elevation images and it uses image processing techniques, specially Mathematical Morphology, and machine learning techniques. Our general workflow is presented in Fig. 6 and it consists in the following steps: i) the 3D point cloud is projected onto an elevation image; ii) ground is segmented using the λ–flat zones labeling algorithm, curbs are segmented and their accessibility is analyzed; iii) facades are segmented as the highest vertical structures in the urban scenario; iv) objects are detected as discontinuities on the ground; v) small and isolated structures are eliminated and connected objects are separated using a constrained watershed; vi) objects are classified in several categories using a SVM approach with geometrical and contextual features; and, vii) the segmented and classified images can be reprojected to the 3D point cloud for visualization purposes.

For further details and complete analyses in each step, the reader is encouraged to review Hernández, J. and Marcotegui, B. (2009); Serna, A., Marcotegui, B. (2013a,b, 2014).

Fig. 7 presents a results of our proposed method on a datasets in the *rue Cassette* in Paris.

7 CONCLUSIONS

Results of the contest will be presented at the IQmulus workshop taking place on July 8th, 2014 in Cardiff (UK), in conjunction with SGP'14.

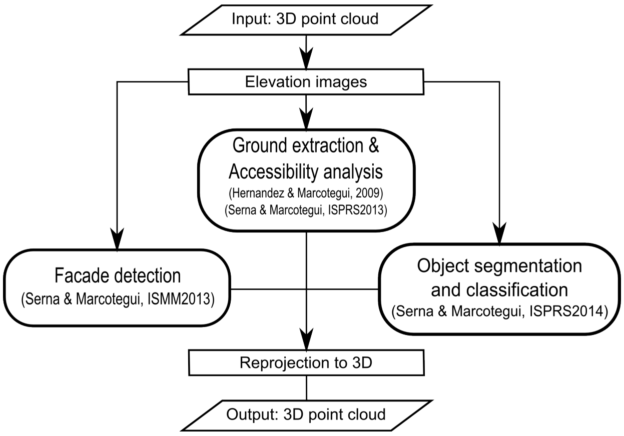
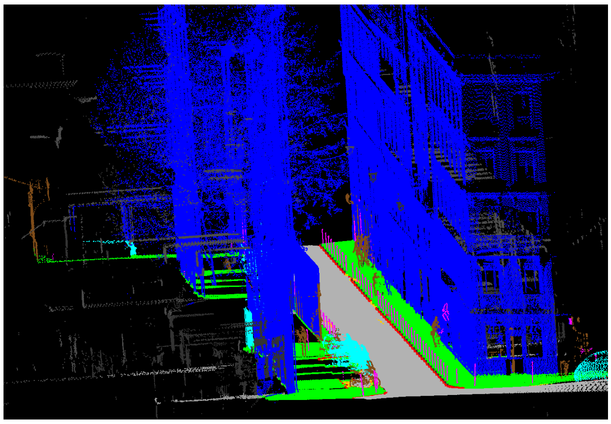


Figure 6: Work-flow of our proposed segmentation methodology.



(a) Rue Cassette, Paris. Acquired by Stereopolis II, IGN France.

Figure 7: 3D object segmentation and classification using our proposed methodology. Facades (blue), road (gray), sidewalk (green), accessible curbs (yellow), non-accessible curbs (red), cars (cyan), pole-like objects (magenta), other (brown) and undefined (dark-gray).

References

Hernández, J., Marcotegui, B.. Filtering of artifacts and pavement segmentation from mobile LiDAR data. In: Bretar, F., Pierrot-Deseilligny, M., Vosselman, M.G., editors. ISPRS workshop Laser scanning '09; vol. XXXVIII-3/W8 of *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Paris, France; 2009, p. 329–333.

Serna, A., Marcotegui, B.. Attribute controlled reconstruction and adaptive mathematical morphology. In: 11th International Symposium on Mathematical Morphology. Uppsala, Sweden; 2013a, p. 205–216.

Serna, A., Marcotegui, B.. Urban accessibility diagnosis from mobile laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 2013b;84:23–32.

Serna, A., Marcotegui, B.. Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning. ISPRS Journal of Photogrammetry and Remote Sensing 2014;93:243–255.

Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E. Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. ICPRAM 2014. `http://cmm.ensmp.fr/~serna/rueMadameDataset.html`

Weinmann, M., Jutzi, B., Mallet, C. Feature relevance assessment for the semantic interpretation of 3D point cloud data. ISPRS Workshop on Laser Scanning, Antalya, 2013.