Point Cloud Segmentation towards Urban Ground Modeling

Jorge Hernández and Beatriz Marcotegui
Mines ParisTech
CMM- Centre de morphologie mathématique
Mathématiques et Systèmes
35 rue St Honoré 77305-Fontainebleau-Cedex, France
Email: { hernandez,marcotegui }@cmm.ensmp.fr

***Abstract*—This paper presents a new method for segmentation and interpretation of 3D point clouds from mobile LIDAR data. The main contribution of this work is the automatic detection and classification of artifacts located at the ground level. The detection is based on Top-Hat of hole filling algorithm of range images. Then, several features are extracted from the detected connected components (CCs). Afterward, a stepwise forward variable selection by using Wilk's Lambda criterion is performed. Finally, CCs are classified in four categories (lampposts, pedestrians, cars, the others) by using a SVM machine learning method.**

I. Introduction

In recent years the laser telemetry has been gradually integrated on board systems to digitize the 3D geometry of natural and urban environments. The segmentation and semantic interpretation (3D point cloud) is one of the essential tasks in 3D modeling process. The segmentation consists in separating building façades, roads, pedestrians, trees, and all elements which belong to urban scenes. Furthermore, at the level of building façades, 3D data permits window locations, wall approximation as planar surfaces, etc.

Several approaches are focused on façade modeling and urban scene segmentation. Automatic planar segmentation approaches from 3D façade data are presented in [1]–[4]. Region growing algorithms are used to extract planar surfaces [1], [2] and the planar approximation is carried out using RANSAC paradigm [3], [4]. In addition, a fully automatic approach of point cloud segmentation is presented in [5]. Madhavan and Hong [6] detect and recognize buildings from LIDAR data. Also, Goulette et al. [7] present a segmentation based on the profiles of points, for the following elements in the scene: ground (road and pavement), façades and trees. Nevertheless, those segmentation approaches are performed only on the scene geometry from their acquisition systems. This work is focused on detection and classification of artifacts at the ground level such as: cars, pedestrians, lampposts, etc. This detection is twofold: 1. - Filtering of these structures to facilitate the modeling process from buildings/façades and 2. - Re-introduction of some elements (lampposts, sign boards, bus stop, etc), improving visual realism in modeled scene [8]. Similar approaches of artifact extraction on aerial LIDAR are presented in [9], [10]. A vehicle blob segmentation based on histogram thresholding and edge detection techniques is described in [9]. Yao et al. extract individual vehicles based on morphological operators [10]. Both approaches suppose that all objects are vehicles. As well, they present problems with slopping streets and they have to compensate the inclination angle. In contrast, our method handles these problems, detects all artifacts and classifies them in four categories.

978-1-4244-3461-9/09/$25.00 © 2009 IEEE

Our research is developed in the framework of Cap Digital Business Cluster TerraNumerica project. This project aims to develop a production and exploitation platform, by allowing the definition and visualization of synthetic urban environments. The platform aims in order to increase the productivity and the realism of urban modeling.

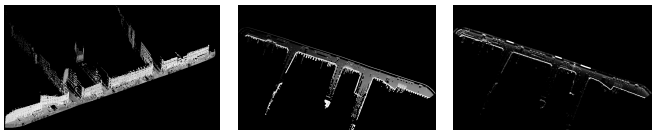
The paper is structured as follows. Data description and range image computation is presented in Section II. Section III describes the detection of artifacts. Then, section IV shows variable selection and classification methods. In Section V, experimental results are shown and the performance of our method is illustrated. Finally, conclusions are drawn in Section VI.

II. 3D Data

The proposed method makes the following assumptions on the 3D data:

- Façade building information is principally located on a vertical plane (z coordinate) and in front of the acquisition system.
- Ground data is perpendicular to façade data.

We propose to exploit 3D data using range and accumulation images. The images are generated by projecting 3D points onto a plane using a " virtual " camera. The camera is placed on the plane with normal vector $\overrightarrow{n} = (0, 0, 1)$ through point $(0, 0, z_{min})$, i.e. parallel to plane XY and positioned on the lowest value of z coordinate. The image coordinates (u, v) are defined by a parametric function related to the camera. The range image is a representation of 3D information where the pixel intensity is a function of the measured distance between 3D points and the camera plane. If several points are projected on the same image coordinates, Z-buffer algorithm [11] is used to determine which distance is stored. The accumulation image counts the number of points which are projected on the same pixel (u, v) of the camera image. Fig. 1 illustrates an example of 3D point clouds and range and accumulation images. Notice that maxima values in the images are in façade pixels, by validating the three initial assumptions.



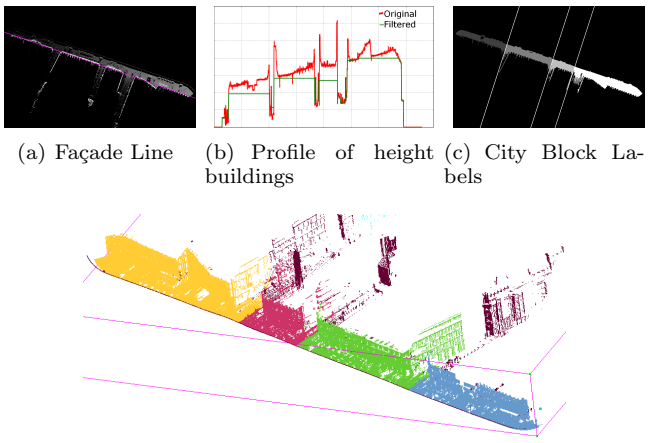
(a) Point Cloud (b) Range Image (c) Acc. Image

Figure 1. Example of the Range and Accumulation Images. 3D Data Rue Soufflot Paris ©IGN.

Range images are a $\mathbf{R}^3 \rightarrow \mathbf{N}^2$ projection. In order to avoid sampling problems, their dimensions should be carefully chosen. If they are too small, there will be an important information loss. Otherwise if they are too large, pixel connectivity is not ensured (required by method). Hence, the ideal choice of dimensions is $1:1$ i.e. a pixel by a 3D point in the camera plane. In our case, point clouds have a resolution of approximately $20[cm]$, and for this reason the selected resolution must be approximately equal to $5[pix/m]$.

A. City Block Separation

In order to handle dense point clouds, we start separating into city blocks. We assume that the façades of the same street are aligned. This assumption is verified on our data. The algorithm uses Hough transform to detect the façade direction (Figure 2(a)). On this line, we analyze a profile of building heights. We detect the most important profile variations, which correspond to the city block separation streets. In order to detect these variations, the profile is initially filtered (Figure 2(b)) and then a watershed segmentation is performed. The divisions are traced in perpendicular direction to the façade direction to split them into city blocks (Figure 2(c)). Urban divisions are reprojected onto 3D data via parametric function of a "virtual" camera (Figure 2(d)) and once the 3D data is split, a detection step is applied to each urban block separately.



(a) Façade Line (b) Profile of height buildings (c) City Block Labels

(d) 3D City Blocks

Figure 2. Procedure of City Block Separation.

III. Detection

By assuming that the artifacts are on the ground, a first ground detection is necessary. Then, we look for artifacts inside the mask.

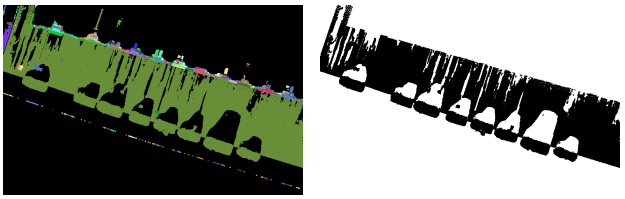
A. Ground Mask

The method is based on a raw segmentation of range image, by using $\lambda-$ flat zones labeling algorithm. Quasi-flat zones are introduced by Meyer in [12].

Definition 1: Two neighboring pixels p, q belong to the same quasi-flat zone of a function f, if their difference $|f_p - f_q|$ is smaller than or equal to a given λ value.

$$\forall\,(p,q)\,neighbors : |f_p - f_q| \leq \lambda \qquad (1)$$

As the height variations on the ground level are small, a range image segmentation with a height $\lambda = 2m$ will cluster ground pixels into the same region (Figure 3(a)). By choosing the largest region we obtain the ground mask (Figure 3(b)). Then, the range image is masked by the ground mask, in order to remove façade information.



(a) $\lambda-$ flat zones labeling (b) Ground Mask

Figure 3. Ground Mask Detection.

B. Artifacts

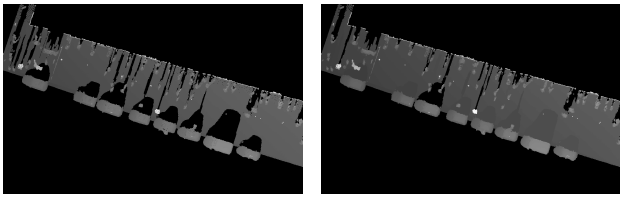
The proposed method for artifact detection is based on hole filling algorithm [13]. Holes of an image correspond to sets of pixels whose minima are not connected to the image border [14]. The algorithm consists in removing all minima which are not connected to the border by using the morphological reconstruction by erosion (Eq. 2). The image marker (mk) is the maximum image value except along its border (Eq. 3).

$$\text{Fill}\,(f) = R_f^{\varepsilon}\,(mk) \tag{2}$$

where,

$$mk = \begin{cases} f_p & \text{if } p \text{ lies on the border} \\ \max\,(f) & \text{otherwise} \end{cases} \tag{3}$$

Hole filling algorithm is used twice: First, to reduce several shadows and concavities inside of the artifacts (missing data) produced by occlusions (Fig. 4). Second, to extract artifacts separately. An artifact can be seen as a hump on the road. Hence, if we invert the range image, those humps become new holes (concavities) to be filled. Then, the artifacts are detected by making the difference between the inverted range image and the filled image . The method allows to handle slopping streets because they are linked to the border.



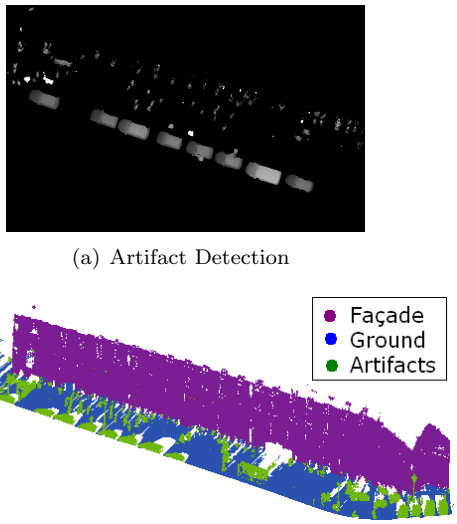
(a) Masked Range Image (b) Image with Filled Holes

Figure 4. First Hole Filling: Shadow Reduction.

Nevertheless, two additional details must be considered:

- **Black pixels:** Once the image is inverted, the image background (black) becomes the highest value. In this case, whole depth information is a big hole. To solve this problem, we invert only the points different from zero (those pixels remain in black).
- **Border:** The artifacts which are close to the border, especially cars, lack side information caused by the sensor proximity and occlusions. By applying Top-Hat operator, the artifact will only have upper information. The problem is solved, by putting the minimum value on mask image surrounding.

A threshold of 10[cm] is applied to the Top-Hat result in order to eliminate artifacts produced by ground roughness and noisy surfaces. At this stage of the method, the artifacts are satisfactorily detected (see Fig. 5(a)). The artifacts are reprojected onto the point cloud. Thus, they may be filtered out form the 3D data allowing data filtering for urban scene modeling. Fig. 5(b) shows the segmentation result: façade, ground and artifacts.



(a) Artifact Detection

(b) Data Segmentation

Figure 5. Detection of Artifacts and Data Segmentation.

IV. Classification of Artifacts

A. Connected Artifacts

Most artifacts are easily separable with a simple labeling. However, some of them are connected to each other. Fig. 6 shows three situations where this problem appears: pedestrians close to a lamppost (Fig. 6(a)), two cars too close (Fig. 6(b)) and a pedestrian opening car door (Fig. 6(c)).

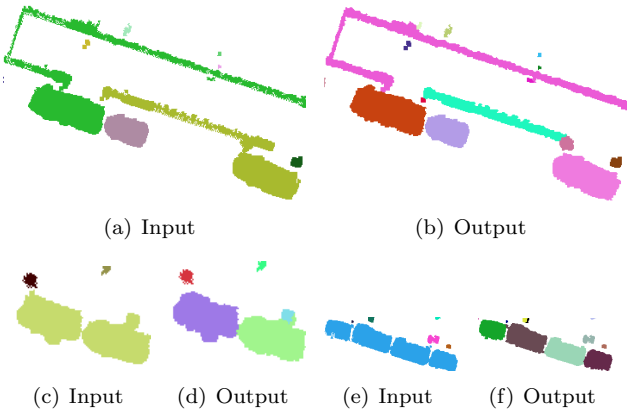
(a) (b) (c)

Figure 6. Connected Artifacts.

As we have described an artifact as a hump, we analyze how many maxima each CC has. The maxima number corresponds to the number of artifacts which are contained by a CC. Nevertheless, some small noise peaks and the top roughness of artifacts produce a lot of maxima. In order to reduce this problem, we use an area opening [15] and h-Maxima operator [14]. h-Maxima eliminates maxima which have a depth equal to or lower than a threshold h. The operator use a morphological reconstruction by dilation of f from a $f - h$ (Eq. 4).

$$\text{HMax}\,(f) = R_f^{\delta}\,(f - h) \tag{4}$$

Once the artifact image is filtered, we use maxima and background images as markers and a constrained watershed on the range image gradient is carried out to segment the new artifacts. Fig. 7 shows two examples of the new segmentation, by showing the performance of our method.



(a) Input (b) Output

(c) Input (d) Output (e) Input (f) Output

Figure 7. **(a)**,**(c)**,**(e)** Initial Detection and **(b)**,**(d)**,**(f)** Separation of Connected Artifacts.

B. Feature Extraction

Each connected component from the image of artifacts is characterized to determine artifact types. Four categories of artifacts are analyzed: lampposts, cars, pedestrians and the others (garbages, motorcycles, parking-meters, bus stops, etc). The extracted measures are:

- Mean, standard deviation, maximum, minimum and mode of artifact height [meters].
- Mean, standard deviation, maximum, minimum and mode of artifact accumulation (number of points projected onto the same pixel) [times] .
- Area. Initially, this feature is calculated in pixels. However, by using the camera information, the value is converted to square meters. This transformation homogenizes this feature in whole images (each range image is independently generated).

We annotated our database to establish a ground truth of 442 artifacts: 67 cars, 33 lampposts, 198 pedestrians and 144 the others. In order to reduce our space of eleven characteristics, a stepwise forward variable selection is carried out. The method is based on Wilk's lambda criterion [16], [17]. It starts with the variable which best separates the classes, and adds one by one if any is statistically significant (p-value). Fig. 8 shows ranked variables by using this method.

The distribution of classes with the first three characteristics is illustrated in Fig. 9, where the area is the main feature of cars and mean height of lampposts. However,

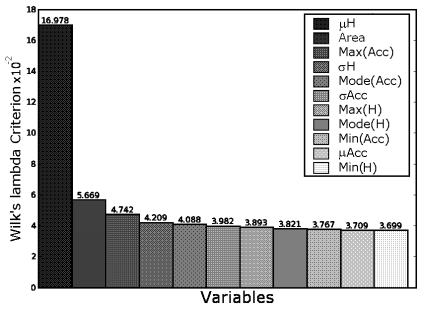


Figure 8. Selection Step of Features (Ranked via Wilk's lambda criterion).

the pedestrian class and the other one are intersected. If we set 0.01 as the maximum permissible error (p-value) by reduced model compared to model with all variables, we can reduce the model to 6 features.

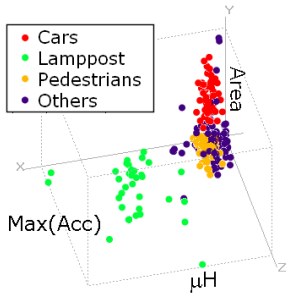


Figure 9. Class Distribution: μHeight *vs* Area *vs* max *Acc.*

C. Classification

Firstly, we eliminate CCs smaller than 10 pixels and which maximum accumulation value is lower than 3. Then, we use a supervised classification method, called support vector machine (SVM) [18], [19]. The SVM finds a linear separating hyperplane with the maximal margin in a higher dimensional space. Finally, the classification is validated by using K-fold cross validation.

Classification results are shown in the confusion matrix of Table I. The table shows a good classification with the use of only six characteristics. All lampposts are properly classified, and cars and pedestrians have good classification rates 92.54% and 96.97% respectively. However, a large quantity of artifacts from the other class were classified as pedestrians, an example of this problem is presented with parking-meters which are structures with similar features to pedestrians (area, height). The total classification error is 14.93%.

V. Experimental Results

The presented method was tested on 3D point clouds acquired by two different mobile systems (LARA 3D of

	Detected			
GT	**Cars**	**Lampposts**	**Pedestrians**	**Others**
Cars	92.54%	0.0%	0.0%	7.46%
Lampposts	0.0%	100.0%	0.0%	0.0%
Pedestrians	0.0%	0.0%	96.97%	3.03%
Others	7.64%	1.39%	29.17%	61.80%

Table I
CONFUSION MATRIX FOR TEST SET.

CAOR Lab-Research [1], and IGN Stereopolis [2]). The point clouds correspond to approximately 900 street meters (ten urban blocks) of the 5th Paris district. The results obtained are satisfactory on the whole dataset. A classification example on 3D point clouds is shown in Fig. 10. Extended tests on large databases are foreseen in the framework of TerraNumerica project.

Figure 10. Classification of Artifacts.

VI. CONCLUSIONS

A method of 3D point cloud analysis and segmentation on ground level urban environments has been presented. The detection of artifacts (without classification) is useful to filter data in order to facilitate urban scene modeling (buildings, façades, roads). The method detects and classifies ground level artifacts as lampposts, cars and pedestrian as well as the location of other artifacts not recognized yet.

At present we are working on the detection of signboards and we will extend the method to the detection of other artifacts such as motorbikes and waste garbage containers. The introduction of other classes in order to reduce the rate of wrong classification in the class "the others".

ACKNOWLEDGMENT

The work reported in this paper has been performed as part of Cap Digital Business Cluster Terra Numerica project.

[1]caor.ensmp.fr/french/recherche/rvra/3Dscanner.php
[2]recherche.ign.fr/labos/matis/accueilMATIS.php

REFERENCES

[1] C. Dold and C. Brenner, "Registration of terrestrial laser scanning data using planar patches and image data," in *International Archives of Photogrammetry and Remote Sensing*, 2006, pp. 25–27.

[2] I. Stamos, G. Yu, G. Wolberg, and S. Zokai, "3d modeling using planar segments and mesh elements," in *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 599–606.

[3] S. Becker and N. Haala, "Combined feature extraction for façade reconstruction," in *ISPRS Workshop on Laser Scanning and SilviLaser, Espoo, Finland*, September 2007, pp. 44–50.

[4] H. Boulaassal, T. Landes, P. Grussenmeyer, and F. Tarsha-Kurdi, "Automatic segmentation of building facades using terrestrial laser data," *Laser*, p. 65, 2007.

[5] V. Verma, R. Kumar, and S. Hsu, "3d building detection and modeling from aerial lidar data," in *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2213–2220.

[6] R. Madhavan and T. Hong, "Robust detection and recognition of buildings in urban environments from ladar data," *International Conference on Artificial Intelligence and Pattern Recognition*, pp. 39–44, 2004.

[7] F. Goulette, F. Nashashibi, S. Ammoun, and C. Laurgeau, "An integrated on-board laser range sensing system for on-the-way city and road modelling," *Revue française de photogrammétrie et de télédétection*, vol. 185, p. 78, 2007.

[8] N. Cornelis, B. Leibe, K. Cornelis, and L. V. Gool, "3d city modeling using cognitive loops," in *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 9–16.

[9] Ádam Rakusz, T. Lovas, and Áarpard Barsi, "Lidar-based vehicle segmentation," Department of Photogrammetry and Geoinformatics, Budapest University of Technology and Economics, Tech. Rep., 2004.

[10] W. Yao, S. Hinz, and U. Stilla, "Automatic vehicle extraction from airborne lidar data of urban areas using morphological reconstruction," in *IAPR Workshop on Pattern Recognition in Remote Sensing*, 2008.

[11] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice in C*, 2nd ed. Addison-Wesley Professional, August 1995.

[12] F. Meyer, "From connected operators to levelings," in *ISMM '98: Proceedings of the fourth International Symposium on Mathematical Morphology and its Applications to Image and Signal Processing*. Norwell, MA, USA: Kluwer Academic Publishers, 1998, pp. 191–198.

[13] P. Soille and C. Gratin, "An efficient algorithm for drainage networks extraction on DEMs," *Journal of Visual Communication and Image Representation*, vol. 5, no. 2, pp. 181–189, June 1994.

[14] P. Soille, *Morphological Image Analysis: Principles and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.

[15] L. Vincent, "Morphological area openings and closings for grey-scale images," in *Proceedings of the Workshop "Shape in Picture", 7–11 September 1992, Driebergen, The Netherlands*, Y.-L. O, A. Toet, D. Foster, H. J. A. M. Heijmans, and P. Meer, Eds., Berlin, 1994, pp. 197–208.

[16] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Academic Press, 1979.

[17] C. Roever, N. Raabe, K. Luebke, U. Ligges, G. Szepannek, and M. Zentgraf, *Classification and visualization (klaR Package)*, Fakultaet Statistik, Technische Universitaet Dortmund, February 2009.

[18] C. chung Chang and C. jen Lin, "Libsvm: a library for support vector machines," 2001.

[19] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, and A. Weingessel, *Misc Functions of the Department of Statistics (e1071 Package)*, TU Wien, February 2009.