Object Extraction in Urban Environments from Large-Scale Dynamic Point Cloud Datasets

Attila Börcs, Oszkár Józsa and Csaba Benedek
Distributed Events Analysis Research Laboratory, Computer and Automation Research Institute (MTA SZTAKI)
H-1111, Kende utca 13-17 Budapest, Hungary, E-mail: firstname.lastname@sztaki.mta.hu

***Abstract*—In this paper, we introduce a system framework which can automatically interpret large point cloud datasets collected from dense urban areas by moving aerial or terrestrial Lidar platforms. We propose novel algorithms for region segmentation, motion analysis, object identification and population level scene analysis which steps can highly contribute to organize the data into a semantically indexed structure, enabling quick responses for content based user queries about the environment. The system is tested on real Lidar data, and for demonstration quantitative evaluation is given on vehicle detection.**

I. Introduction

Efficient data organization and semantic indexing are crucial tasks in Geographic Information Systems (GIS). Available GIS solutions are able to store, manipulate, and manage various geographically referenced data; however they have limited functionalities for automatic data analysis and inference, and for managing the temporal dimension. Nowadays significant research efforts are conducted to obtain intelligent and dynamic GISs filled up with heterogeneous multi-temporal data, which can be able to visualize and answer high level spatio-temporal user queries about the environment.

Mobile mapping systems and aerial Lidar platforms are able to rapidly acquire large-scale 3D point cloud data for a GIS, with jointly providing accurate 3D geometrical information of the scene, and additional features about the reflection properties and compactness of the surfaces. In addition the Lidar measurements are less sensitive on the weather and illumination conditions of the acquisition than optical images or videos. On the other hand by indexing Lidar point sets we have to deal with problems of measurement noise, inhomogeneous point density and mirroring artifacts, while manual evaluation is particularly exhausting and unreliable, because the human visual system is not capable to efficiently interpret unorganized point sets [1].

A group of existing point cloud analysis techniques deal with region segmentation. In [2] a solution has been proposed for efficiently handling data that is continuously streamed from a sensor on a mobile robot, and for separating different semantic regions in the point cloud. [3] presents a set of clustering methods for various types of 3D point clouds, including dense 3D data (e.g. Riegl scans) and sparse point sets (e.g. Velodyne scans). However, these approaches mainly focus on research towards real time point cloud classification for robot navigation and quick intervention rather than complex situation interpretation, visualization or dynamic scene analysis, expected in the GIS application environment.

Object recognition from a segmented point cloud is often performed via machine learning techniques using training samples [4], [5], [6]. The authors of [5] use objects from Google's 3D Warehouse to train an object detection system for 3D point clouds for robots, so that they can safely operate in urban and indoor environments. They extract various object level descriptors for point cloud blobs representing the detected objects, while to obtain similar representations of models in the 3D Warehouse, they perform ray casting on the models to generate point clouds, finally the classification is performed in the descriptor space. However, difficulties in object extraction are not detailed here, and information from object-environment interactions is not exploited.

In this paper, we focus on the interpretation and analysis of large point cloud sets collected by either aerial Lidar systems or a terrestrial Rotating Multi-Beam (RMB) Lidar configuration (Velodyne) which is able to provide 360° point stream with a frequency of 5-15 Hz. In the considered cases the object separation itself is a crucial issue which is highly challenging in crowded urban environments, where the ground cannot be considered planar, the point density is either low (aerial platform) or rapidly decreases as a function of the distance from the sensor (RBM Lidar), and several moving and static objects are present in the scene causing occlusion and contacting effects. We also propose to step over the pure data driven object level classification approach and include strong prior knowledge and contextual information in the analysis with a robust Bayesian tool called Marked Point Process [7]. Finally, regarding the RBM configuration, we highly exploit the temporal information obtained from the Lidar point cloud stream: after frame registration and separating static and moving point cloud regions we analyze the static objects in merged high density point clouds, while attempt to estimate the trajectory of moving vehicles through several consecutive frames.

II. Proposed point cloud interpretation system

The proposed approach consists of three steps. First, the Lidar point cloud is segmented into different semantic regions. Second, the RMB Lidar frames are automatically registered, i.e. transformed to a common coordinate system, then the point

This work is connected to the i4D project funded by the internal R&D grant of MTA SZTAKI. The third author also acknowledges the support of the Hungarian Research Fund (OTKA #101598), and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

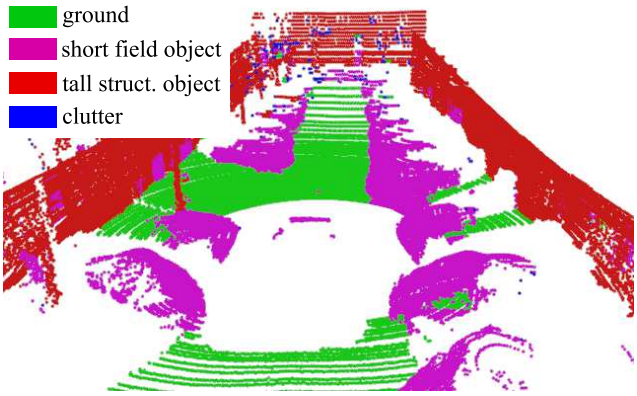


Fig. 1: Segmented frame of the Velodyne point cloud stream. Note: figures of this paper are best viewed in color print.

Fig. 2: Registration of 30 Velodyne frames about the building of the Great Market Hall, Budapest, Hungary

cloud regions corresponding to static and moving objects are separated, and the segmentation of the first step is refined based on features extracted from the merged cloud. This step does not concern the aerial data, which is obtained by a line-scanning technique, thus the Lidar system already provides a registered point cloud. Third, individual objects are extracted from the point cloud and the global scene is analyzed as a population of interacting entities.

A. Point cloud segmentation

The segmentation process assigns to each measured point a class label from the following set: (i) *clutter* (ii) *vegetation*, (iii) *ground*, (iv) *tall structure object* (walls, roofs, lamps posts, traffic lights etc.) and (v) *short street object* (vehicles, pedestrians etc.). As shown in Fig. 1, 2 and 4, walls usually appear in terrestrial, roofs in aerial scans, otherwise they cause sparse *clutter* regions in the point cloud.

First, local point cloud density is calculated to extract points of the *clutter* class. *Vegetation* can be detected in aerial measurement containing multiple laser returns in a straightforward way, exploiting that trees and bushes cause more than one echoes. Regarding the terrestrial data, we only remove the vegetation after frame registration (Sec. II-B). The next step is terrain modeling. Planar *ground* models are frequently adopted in the literature relying on robust plane estimation methods such as RANSAC. However, in the considered urban scenes we experienced significant elevation differences (often up to a few meters) between the opposite sides and central parts of the observed roads and squares. In these cases, planar ground estimation yields significant errors in the extracted object shapes, e.g. bottom parts can be cut off, or the objects may drift over the ground. On the contrary, we apply a locally adaptive approach: we fit a regular 2D grid onto the horizontal plane with *zero elevation* using cells with side length of 50-80cm, and project the points of the cloud to the plane. In each cell, statistical features are calculated among the 'non-outlier' points assigned to the cell, which have been classified neither as clutter, nor as vegetation yet. A cell, with all corresponding points is classified as *ground*, if the difference between the maximal and minimal point elevations is smaller than a threshold (used 25cm), and the average point elevation is within an allowed height range based on a globally estimated digital terrain map. The first criteria ensures the flatness or homogeneity of the points. Given a cell with 60 centimeters if width, this allows $22.6\,^\circ$ of elevation within a cell; higher elevations are rarely expected in an urban scene. The second criteria ensures that this patch of flat surface is under the car. A cell corresponds to *tall structure object*s, if either the difference of the maximal and minimal elevations of the included points is larger than a threshold (used 310 centimeters), or the maximal observed elevation is larger than a predefined value (used 140 centimeters). The first criterion is needed for dealing objects standing on a lower point of the ground. The rest of the point cloud is assigned to class *short street object* like vehicles, pedestrians, short road signs, line posts etc. These entities can be either dynamic or static, which attribute can only be determined later, after further, more complex investigation of the point cloud sequence.

B. Lidar scan registration

In this section, we propose a method for automatic registration of sparse terrestrial Lidar, yielding dense and detailed point clouds of large street scenes.

Although various established techniques do exist for point cloud registration, such as Iterative Closest Point (ICP) and Normal Distribution Transform (NDT) [8] these methods fail, if we try to apply them for the raw Velodyne LIDAR point clouds for two reasons: 1) All moving points appear as outliers for the matching process, and since in a crowded street scene we expect several moving objects, many frames are erroneously aligned. 2) Due to the strongly inhomogeneous density of the LIDAR clouds, even the static ground points mislead the registration process. The above algorithms often match the concentric circles of the ground (see Fig. 1), which yields that the best match erroneously corresponds to a near zero displacement between two consecutive frames. However, we have also observed that the point density is quite uniform in local wall regions which are perpendicular to the ground.

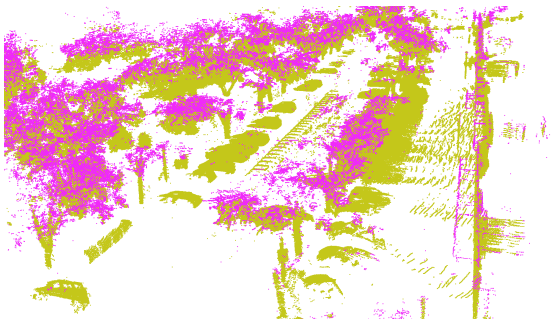


Fig. 3: Vegetation detection (marked with purple).

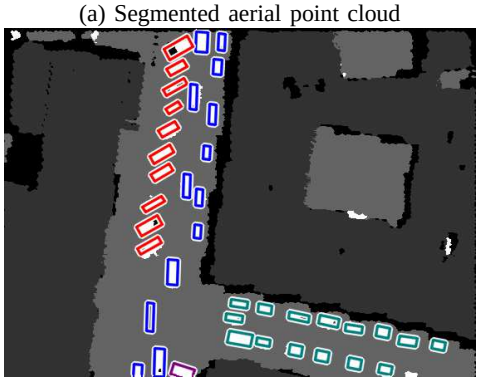
Our key idea is to utilize the point classification result from the previous section to support the registration process. We only use as input of the registration algorithm the points segmented as *tall structure objects*, since we expect that in majority, these points correspond to stationary objects (such as buildings), thus they provide stable features for registration. The NDT algorithm was applied to match the selected regions of the consecutive frames of the point cloud [8], since it proved to be efficient with the considered data and it is significantly quicker than the ICP. After calculating the optimal transformation, the whole point cloud of each frame is registered to a joint world coordinate system. The method is also able to deal with tilted sensor configurations which may result in complete models of tall building facades based on the RMB-Lidar data, as shown in Fig. 2.

As mentioned in Sec. II-A we have also developed a vegetation removal algorithm for the merged point cloud, which calculates a statistical feature for each point based on the distance and irregularity of its neighbors, and also exploits the intensity channel which is a strong indicator of vegetation (see Fig. 3). In this way, we can improve the segmentation of Sec. II-A in cases of trees hanging over parking cars. Thereafter, we also refine the separation of *ground*, *tall* and *short objects* in the registered cloud, using the previously introduced classification steps. However, the regions of moving objects in the merged point cloud cause blurred object blobs (Fig. 6), which should be indicated. Although dynamic regions have generally a lower point density, in our experiments region-level local features proved to be inefficient for motion separation. Instead, we utilize blob-level features: We extract connected blobs of the *short object* regions in the merged cloud with floodfill propagation, then within each blob we separate the points corresponding to the different time stamps and determine their centroids. Assuming that the centroids of the same object follow lines or elongated curves if the object is moving, and make small random movements in a certain region if the object is static, we can cluster the moving and static object regions as shown in Fig. 6.

C. Object Extraction and Population analysis

Marked Point Processes (MPP) provide an efficient Bayesian tool to characterize object populations, through jointly describing individual objects by various data terms,

(a) Segmented aerial point cloud



(b) Vehicle extraction and traffic segmentation

Fig. 4: Traffic analyis from aerial Lidar data

and using information from entity interactions by prior geometric constraints. In addition, with utilizing a two-layer extension of the MPP models [7] we can partition the entity populations into groups of semantically corresponding objects, called configuration segments, and extract the objects and the optimal segments simultaneously by a joint energy minimization process. For example, performing complex traffic analysis in the considered GIS data needs a hierarchical modeling approach: at low level individual vehicles should be detected and separated, meanwhile at a higher level we need to extract coherent traffic segments, by identifying groups of corresponding vehicles, such as cars in a parking lot, or a vehicle queue waiting in front of a traffic light. In our system framework, various configurable data models can be included about different objects such as vehicles and pedestrians. The module is also able to consider flexible user defined prior constraints about expected vehicle alignment patterns and shapes. System parameters can be set by sample based training or user settings. For model optimization, a Hierarchical extension of the Multiple Birth and Death Algorithm is adopted.

III. Experiments and Conclusions

We have tested the proposed framework in real aerial and terrestrial Lidar measurements. Fig. 4 shows result on joint vehicle extraction and grouping from aerial Lidar scans. Detected traffic segments – displayed with different colors

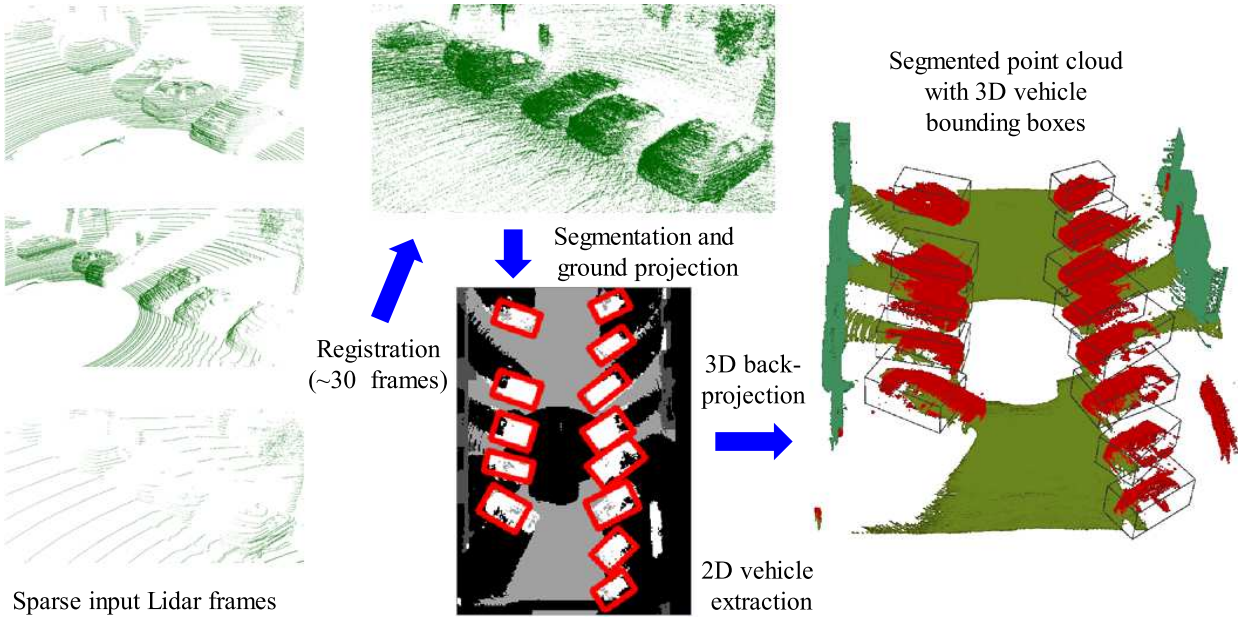


Fig. 5: Workflow of parking vehicle extraction from the terrestrial point cloud sequence

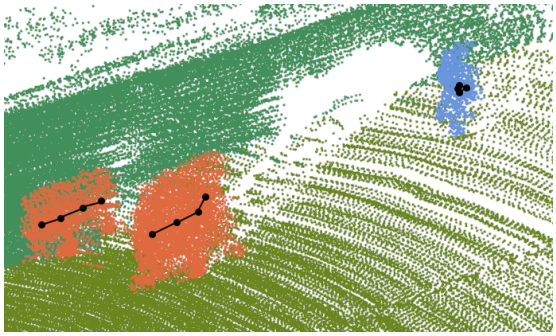


Fig. 6: Blobs of two moving pedestrians (orange) and a standing person (blue) on the merged and segmented point cloud. Trajectories of object centroids in the separate time frames are marked with black.

TABLE I: Evaluation of vehicle detection in an aerial and terrestrial Lidar test set.

Data set	NV	TP	FP	FN	F-rate
Aerial	471	443	64	28	91%
Terrestrial	141	140	1	2	99%

– are formed by parallelly parking or traveling cars. Fig. 5 demonstrates the workflow of vehicle detection from terrestrial Lidar sequences: first the sparse frames are registered and merged, then the point labels of the segmented cloud are projected to the ground plane, where the vehicle population is a modeled as a configuration of 2D rectangles, finally the MPP detection results are back projected to the 3D point cloud space. In Fig. 6, we can find an example for pedestrian identification and tracking, considering two walking (orange) and one standing person. Regarding vehicle detection, we also performed quantitative evaluation, by counting the total Number of Vehicles (NV) in the tests sets, the True Positive (TP), False Positive (FP) and False Negative (FN) detected objects, and the F-rate of the detection. Results in Table I confirm that the proposed approach is notably accurate at object level, especially regarding the high resolution terrestrial data. As a conclusion, we have introduced a system for automatic analysis of large Lidar point cloud sets collected from dense urban areas, and demonstrated its usability by selected application examples. Our future work includes a more extensive evaluation dealing with several types of objects, and extraction of various dynamic parameters of the scene.

REFERENCES

[1] A. Velizhev, R. Shapovalov, and K. Schindler, "An implicit shape model for object detection in 3D point clouds," in *ISPRS Congress*, Melbourne, Australia, 2012.

[2] H. Hu, D. Munoz, J. A. Bagnell, and M. Hebert, "Efficient 3-D scene analysis from streaming data," in *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.

[3] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel, "On the segmentation of 3D Lidar point clouds," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011, pp. 2798 –2805.

[4] M. Samples and M. R. James, "Learning a real-time 3D point cloud obstacle discriminator via bootstrapping," in *ICRA10 Workshop on Robotics and Intelligent Transportation System*, May 2010.

[5] K. Lai and D. Fox, "Object recognition in 3D point clouds using web data and domain adaptation," *I. J. Robotic Res.*, vol. 29, no. 8, pp. 1019–1037, 2010.

[6] A. J. Quadros, J. Underwood, and B. Douillard, "An occlusion-aware feature for range images," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 4428–4435.

[7] A. Börcs and C. Benedek, "Urban traffic monitoring from aerial Lidar data with a two-level marked point process model," in *International Conference on Pattern Recognition (ICPR)*, Tsukuba City, Japan, 2012, pp. 1379–1382.

[8] P. Biber and W. Strasser, "The normal distributions transform: A new approach to laser scan matching," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, USA, October 2003, pp. 2743–2748.