3D Object-based Classification for Vehicle Extraction from Airborne LiDAR Data by Combining Point Shape Information with Spatial Edge

Wei Yao[1], Stefan Hinz[2] and Uwe Stilla[1]
Photogrammetry and Remote Sensing[1], Technische Universitaet Muenchen,, Germany
Remote Sensing and Computer Vision[2], Karlsruhe Institute of Technology,Germany
{yao, stilla }@tum.de, stefan.hinz@kit.edu

Abstract

The problem of vehicle extraction using airborne laser scanning (ALS) is studied under the framework of object-based point cloud analysis (OBPA). Object extraction relies on the partitioning of raw ALS data into various segments approximating semantic entities followed by classification. A 3D segmentation method working directly on point cloud is used, which features the detection of local arbitrary modes and the globally optimized organization of segments concurrently. To make the segmentation more competent in extracting small-scale objects such as vehicle, the detection of local structures is realized by adaptive mean shift (MS) using variable bandwidths which are determined by the point shape information bounded by spatial edge. The experimental results show that the proposed method performs very well in terms of visual interpretation as well as extraction accuracy.

1. Introduction

The automatic extraction, characterization and monitoring of traffic using remote sensing platforms is an emerging field of research. Approaches rely not only on airborne video but on nearly the whole range of available sensors [1],[2]. Traffic monitoring based on optical systems, however, is only possible at daytime and cloud-free conditions. Besides SAR systems, LiDAR can also work in the night time and have certain ability to penetrate the cloud. Beside the advantage mentioned above, the LiDAR system has following ones especially in view of our application:

1. The completeness of vehicle extraction using LiDAR data could be increased due to the partial penetration ability of laser rays against vegetations.
2. The moving object could generate motion artifacts in LiDAR data which are utilized to distinguish the movement and estimate the velocity.

The research work of [3] used an airborne laser scanner coupled with digital video camera to analyze transportation corridors and acquire traffic flow information. Vehicle velocities were estimated either by analyzing LiDAR motion artifacts or by vehicle tracking in image sequences. The experiment showed that the two sensors complement each other and can provide valuable traffic flow data. The work in [4] formulates the mathematical foundations of object discrimination and object re-identification in range image sequences using Bayesian decision theory. In [5], a context- guided approach is used to delineate single vehicles only with geometric information of LiDAR data and the results demonstrated the application feasibility of extracted vehicle points towards the motion analysis.

In this study, we will address the same key issue in traffic monitoring using ALS data – vehicle extraction. However, this paper is focused on extracting the vehicle points directly from 3D point cloud to overcome inherent deficiencies of the gridding process of LiDAR data and designing a more accurate and reliable approach for motion analysis and velocity estimation. Based on the work of [6], this paper is to propose a methodology that integrates height edge and point shape information in the segmentation substep at local level to improve the performance of vehicle extraction through object-based classification. The direction lines are extended by a joint rule to determine the point-related bandwidth of adaptive mean shift (MS) for detecting local geometric structures, most of which are assumed to correspond to vehicles in the scene.

2. Framework

ALS systems deliver (X, Y, Z) coordinates of the unevenly sampled points over the scene. To maintain the full 3D information in data, the OBPA concept is introduced for object extraction from the raw product of LiDAR sensors – unorganized point cloud to avoid the rasterization to a 2.5D model [7]. Object extraction by the OBPA method is substantially a strategy for object recognition by jointly employing segmentation and classification. This paper is to present an OBPA strategy (Fig.1) for extracting vehicles from ALS data

based on a two-step intelligent 3D segmentation approach working directly on the point cloud, where the point cloud is firstly partitioned into significative segments approximating the spatial extension of various object instances. The classification is then performed on achieved 3D segments to evaluate shape and spatial features to relate semantic classes to each.

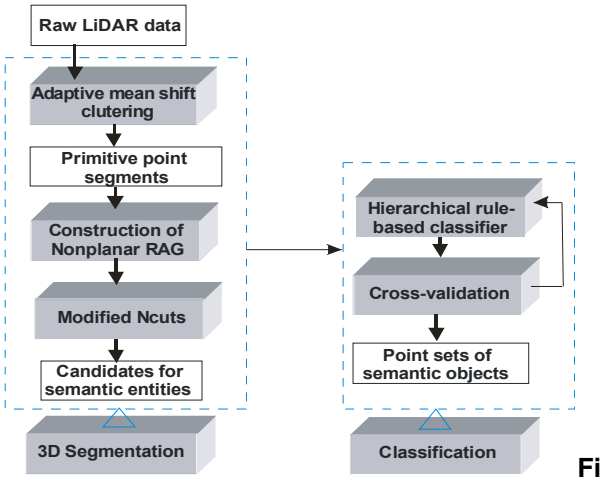


Figure.1 Flowchart for the whole strategy

3. Methodology

3.1. Local modes detection by adaptive MS

MS can perform a genuine clustering directly on LiDAR data. It is intuitive for MS to find arbitrary geometric modes which tend to emerge around local distinct structures to form weak primitives for scene perception. The major challenge of applying the MS algorithm is to adaptively determine the kernel bandwidth in the spatial domain based on different applications. For our application, a multivariate cylinder-shaped kernel is used where the horizontal bandwidth is assumed to vary and need to be determined at each point whereas the vertical bandwidth is kept as constant and set to $2m$. This is because that most of objects in ALS data exhibit stronger scale-variability within the horizontal dimensions. In [8], The Pixel Shape Index is developed to capture the pixel-wise structural information in images by extending direction lines and can easily be adapted to the determination of the point shape information (PSI) in point cloud. This paper is to propose a rule that fuses the spatial edge and shape information when the direction lines are being extended to determine the bandwidth for adaptive MS.

The point cloud is firstly spatially indexed by Geo-tiling process which can establish the corresponding relation between single points and the grid cell [5]. Edge detection for the resulting height grid data is performed using the Canny filter; it is an optimal edge detector having a low probability of false or missing edges and a high accuracy of edge positioning. After applying the Canny detector, it results in binary edge map (Fig.2) where each pixel $edge[x]$ is represented by a value of either 0 (non-edge) or 1 (edge).

PSI is then computed point by point by extending a number of direction lines radiating from the central point based on measuring the point homogeneity which is defined as:

$$PH_d(k,x) = \left| p(k) - p(x) \right| \quad (1)$$

where $PH_d(k,x)$ represents the spatial homogeneity of the dth direction line between the central point k and its surrounding point x. $p(\cdot)$ is the height value of the point.

After introducing the spatial constraint, the condition for extension of direction lines can be formulated as:

$$edge[x] \neq 1 \text{ and } PH_d(k,x) \leq T_1 \text{ and } L_d(k) \leq T_2 \quad (2)$$

where $L_d(k)$ is the length of the dth direction-line, T_1 is a pre-defined threshold for $PH_d(k,x)$ and T_2 is maximal number of points allowed in this direction line. If $edge[x]=1$, it may show an obvious height discontinuity in the local area and hence the extension of direction-lines should be terminated there.

Upon completion of the extension of all direction-lines based on edge and PSI information, the kernel bandwidth can be determined as follows:

$$PSI(k) = \frac{1}{D}\sum_{i=1}^{D} L_d(k) \quad (3)$$

$$h_h(k) = (PSI(k)-1)/2 \quad (4)$$

D is the total number of direction-lines and is invariable. $h_h(k)$ is horizontal bandwidth of the kernel for MS operation centering at point k. Fig.3 illustrates the segmentation result by applying the proposed adaptive MS to one point cloud where local structures are well detected.

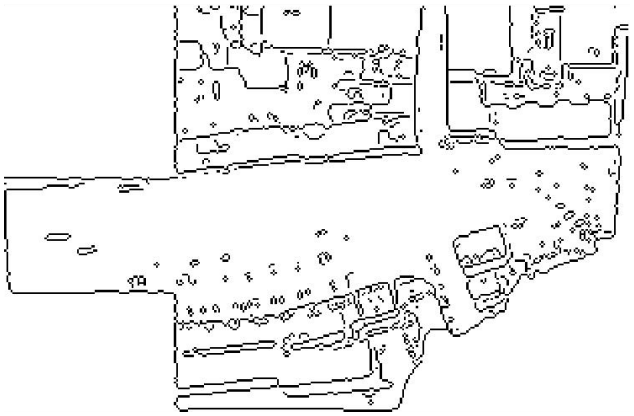


Figure.2 Spatial edge map of an ALS dataset

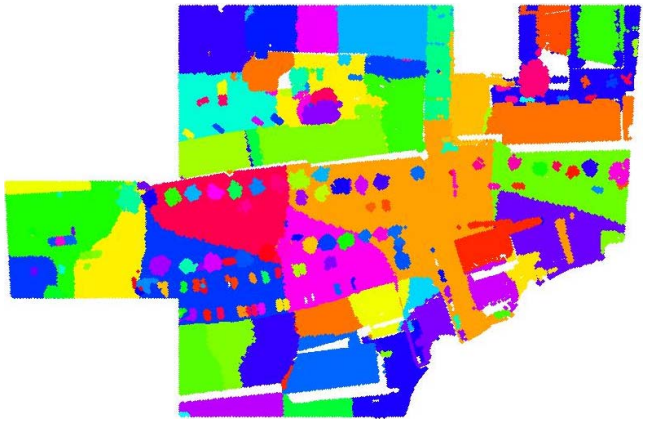


Figure.3 Initial segmentation by Adaptive MS

3.2. Global grouping by Normalized Cut

MS segmentation often produces fragmented but quasi-homogenous sets of point (Fig.3), which are merely meant to approximating the local geometric (sub) primitives and can hardly represent perceivable large-scale objects directly. Consequently, an enhanced segmentation step is introduced in the sense of perceptual organization and global optimization, attempting to group the point segments generated by MS operation to form large semantic objects while keeping small ones left alone. Based on [9], A modified Normalized Cut method (*MNCut*) for image segmentation is chosen to accomplish the task.

Instead of the single points, the point segments produced by MS operation become the unit for grouping and can be represented by a non-planar Region Adjacency Graph (**RAG**) *G*. The subdivision of the G into several segments is realized in the hierarchical procedure:

- Step 1: Create the **RAG** *G* by constructing two edge types – spatial and horizontal edge.
- Step 2:Generate multi-child nodes for *G* and compute weights for all nodes.
- Step 3: Find the solution of the eigenvalue problem and cut the graph *G* into two subgraphs *G1* and *G2* by binarizing the solution vector.
- Step 4:Apply step 3 to the graphs *G1* and *G2*, respectively. Stop if the value for *NCut* exceeds the threshold $NCut_{thres}$.

Single large segments generated by performing *MNCuts* can be used to represent significative (parts of) large-size semantic objects such as ground, road and building, while many small-size segments towards local details like vehicle and tree are also preserved.

3.3. Semantic labeling by classification

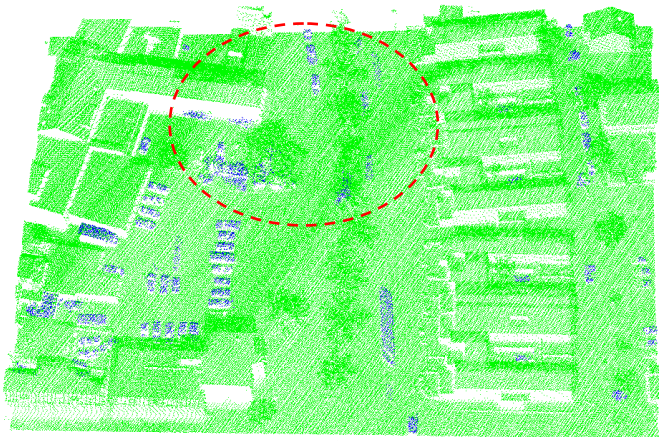
Classification is performed on the point segments resulting from the segmentation step to assign them semantic labels for extracting desirable objects. A rule-based hierarchical classifier is implemented as classification tree to do this task by examining various properties. The tree is constructed by binary recursive splitting based on the *Gini* index. By viewing object as the unit for classification, a large set of features can be calculated providing diverse information about its spatial, topological and contextual properties for each point segment. Thus, the overall number of observations is dramatically reduced for a given dataset. Five object attributes were considered as potential sources of information for the classification.

Table 1 Object-based features

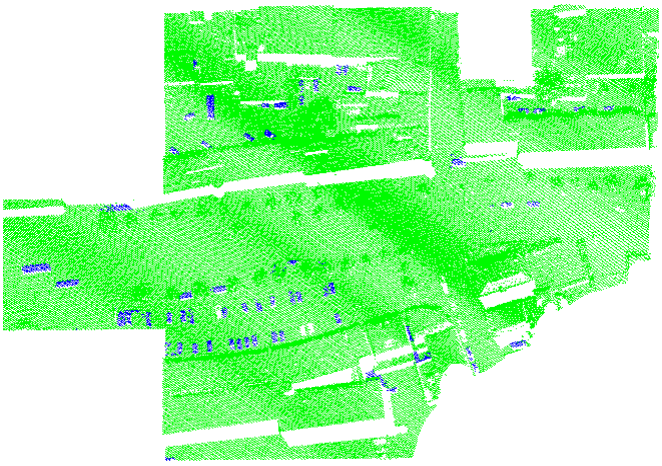
Feature	*Description*	*Formula*
Elongatedness	*Ratio of the object area and the square of its thickness d*	$E(i)=\frac{A(i)}{2d^2}$
Area	*It is represented as polygon of n vertices*	$A(i)=\frac{1}{2}\lvert\sum_{k=0}^{K=n}(x_k y_{k+1}-x_{k+1}y_k)\rvert$
Planarity	*Eigenvalues analysis on 3x3 covariance matrix of the object*	$P(i)=\frac{\lambda_2-\lambda_3}{\lambda_1}$
Vertical position	*Height of centroid of the object*	$Vp(i)=Z(centroid(i))$
Vertical range	*Extension of the object in the Z-axis*	$Vr(i)=\max(Z(i))-\min(Z(i))$

4. Experimental results

The aim of our experiments is to demonstrate the efficiency of our approach. We used two data sections from one ALS data acquisition campaign which was originally acquired for the purpose of city mapping. The data was acquired by the FLI-MAP sensor system of John Chance Land Surveys, Inc over the city of Enschede, the Netherlands. Both datasets cover an area of ca. 200x300 m^2; since the nominal point density of the data is known, a vehicle is supposed to consist of 4x6 laser points. The parameters and thresholds used in the experiments are empirically determined based on another dataset but acquired from the same source.



(a)



(b)

Figure.4 Results of vehicle extraction (blue) overlaid on the original point cloud (green), $NCut_{thres} = 0.37$

As depicted in Fig.4, for both datasets, most of vehicles are detected with promising shape accuracy, particularly including several vehicles beneath trees (within red ellipse of Fig.4a) for the first dataset, few ones are not which is caused by ambiguous laser reflection or deficient topological relation to the surroundings. For vehicles densely placed in parking lots, little poorer shape preservation can be observed due to the oversmoothed height discontinuity. Some parts of building footprints, small vegetations and bulged objects above ground are incorrectly detected as vehicle. However, the incomplete extraction of vehicles seems to be greatly alleviated in our experiments.

The object-based evaluation criterion for vehicle extraction - *AAOE* and *EAOE* [5] is used to assess the algorithmic performance in the sense of detection completeness and correctness, respectively. Additionally, the shape fidelity of extracted vehicles needs to be examined to guarantee that they can be delivered to the motion analysis to obtain the correct velocity based on moving object model [4]. To this end, the Hausdorff distance metric $H(E,R)$ is used to measure the shape similarity between extracted and reference point sets of vehicles. The quantitative results from the evaluation are summarized in Tab.2, which confirmed that most of vehicles were very well extracted in terms of detection rate as well as shape accuracy.

Table.2 Object-based quality evaluation

	AAOE	*EAOE*	*Mean*$\{H(E,R)\}$	*Std*$\{H(E,R)\}$
Data I	83.2%	82.7%	0.23m	0.15m
Data II	81.7%	84.5%	0.29m	0.18m

5. Conclusion

In this paper, an object-based classification approach is proposed to extract vehicles from airborne LiDAR date is presented. Starting from applying adaptive MS clustering to raw point cloud at local level followed by using *MNCut* at global level, point segments corresponding to significative object entities were generated, and then the classification tree technique has been used to assign them the semantic category. The experimental results have shown a success in extraction of most of vehicles from the complex background. The delineated vehicle shapes are also preserved as well and can consequently serve as reliable data foundation for vehicle motion determination. For future work, we will exploit the potential of the extracted vehicles towards motion analysis.

Acknowledgement

The authors would like to thank Prof. Vosselman from International Institute for Geo-Information Science and Earth Observation-ITC for providing the experimental dataset.

References

[1] S. Hinz, R. Bamler, U. Stilla, "Editorial Theme Issue: Airborne und Spaceborne Traffic Monitoring," *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4):135-136, 2006.

[2] S. Suchandt; H. Runge; H. Breit; U. Steinbrecher; A. Kotenkov, U. Balss, "Automatic Extraction of Traffic Flows Using TerraSAR-X Along-Track Interferometry," *IEEE Transactions on Geoscience and Remote Sensing,* , vol.48, no.2, pp.807-819, 2010.

[3] CK. Toth and D. Grejner-Brzezinska, "Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation," *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4): 137-148, 2006.

[4] W. Armbruster,. "Model-based object recognition in range imagery*," Proc. of SPIE* Vol. 7481 748102 (2009).

[5] W. Yao, S. Hinz, U. Stilla, "Automatic vehicle extraction from airborne LiDAR data of urban areas aided by geodesic morphology," *Pattern Recognition Letters*, 31(10), pp,1100-1108, 2010.

[6] W. Yao, S. Hinz, U. Stilla, "Object extraction based on 3D-segmentation of LiDAR data by combining mean shift with normalized cuts: two examples from urban areas," *Proceedings of 2009 Urban Remote Sensing Joint event* (URBAN2009 - URS2009)

[7] M. Rutzinger, B. Höfle, M. Hollaus, N. Pfeifer, "Object-Based point cloud analysis of full-waveform airborne laser scanning data for urban vegetation classification,"*Sensors*, 8, 4505– 4528, 2008.

[8] L. Zhang, X. Huang, B. Huang, P. Li,. "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery*," IEEE Trans. Geosci. Remote Sens*, 44(10), 2950–2961, 2006.

[9] J. Shi, J. Malik. "Normalized cuts and image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(8), 888–905, 2000.