SEGMENTATION-BASED CLASSIFICATION OF LASER SCANNING DATA

Zahra Lari, Ayman Habib
Department of Geomatics Engineering
University of Calgary, 2500 University Drive NW,
Calgary, Alberta, T2N 1N4, Canada
zlari@ucalgary.ca, ahabib@ucalgary.ca

ABSTRACT

Over the past few years, laser scanning has been established as a leading technology for the acquisition of high density 3D spatial information. Digital Terrain Models (DTMs), which can be used for different engineering applications, are obtained by classification of laser data and removing the points that do not belong to terrain surface. The commonly used methods for the classification of laser scanning data are point-based. The major drawback of these methods is focusing on the discontinuities between neighbouring points regardless of the nature of the objects they belong to, which might lead to unreliable classification results. A segmentation-based approach for the classification of both airborne and terrestrial point clouds is presented in this paper. This approach is designed to overcome the drawbacks of point-based classification methods. As the first step, the laser point cloud is segmented by clustering the points with common attributes. To compute precise attributes, an adaptive neighbourhood of each point is firstly defined while considering the proximity of the points in 3D space, surface trend, and noise level in datasets. Then, the coordinates of the origin's projection on the best fitted plane to each point's neighbourhood are computed and used as segmentation attributes. Finally, the laser points with similar attributes are aggregated in the attribute space using a new clustering approach. After segmentation, a heuristic approach is used to classify the segmentation results. The boundaries of segmented surfaces are utilized to determine the adjacency relationship among derived segments. Then, different measures such as the slope and area of each segment, the height difference, and planimetric distance between adjacent segments are checked to classify them into terrain and off-terrain surfaces. The classification of non-segmented points is carried out by comparing the height difference between them and their nearest classified terrain-segments. Experimental results from real data have demonstrated the feasibility of the proposed approach for the classification of airborne and terrestrial laser data.

KEYWORDS: Laser scanning, Digital Terrain Model (DTM), Segmentation, Neighbourhood definition, Minimum convex hull, Classification

INTRODUCTION

In recent years, laser scanning systems have been acknowledged as cost-effective and reliable tools for rapid collection of high density and accurate spatial data. These systems provide 3D point cloud over the scanned surfaces without semantic information about the nature of the associated surfaces. Therefore, the collected data should be processed and classified to extract pertinent information and generate helpful products. One of the most applicable products derived by laser scanning systems is Digital Terrain Model (DTM) that can be utilized in different applications such as 3D city modeling, transportation planning, coastal management, land use/ land cover classification, flood hazard evaluation, and forest mapping. In order to derive a Digital Terrain Model from a laser scanning point cloud, it is necessary to classify the points into terrain and off-terrain points and then remove the points belonging to off-terrain surfaces. This procedure, which is commonly called filtering (Kraus and Pfeifer, 1998), provides useful information for further processing activities such as feature extraction and data interpretation.

Different methodologies have been proposed for the classification and filtering of laser scanning data. The first group of these methods are slope-based approaches introduced by Vosselman (2000). In these methods, the topography of the terrain is approximated locally, using a structuring element. This structuring element defines the maximum acceptable height difference between two points with respect to their distance. All points below the structuring element are accepted as terrain points and all the points above it are identified as off-terrain points. The success of these methods is highly dependent on the defined structuring element and complexity of surfaces under investigation (Zhang et al., 2003). Therefore, a priori knowledge about the characteristics of topography in the study area should be considered in the structural element definition in order to improve the classification results (Sithole,

2001). The second group are surface-based methods, in which a parametric surface with a corresponding buffer is used to classify the laser points (Kraus and Pfeifer, 1998; Elmqvist, 2001). In these methods, a rough estimation of the terrain surface is initially approximated using all laser points. Then, all points are weighted based on their normal distances to the approximated surface. The process of surface fitting is repeated by modification of points' weights until the weights of individual points do not change significantly. The classification of the laser point is then carried out based on their normal distances to the best-fitted surface (Lee and Younan, 2003; Chen et al., 2007). Another method, following this concept, has been presented by Axelsson (2000) which is based on progressive densification of Triangulated Irregular Network (TIN). In these methods, a first approximation of the terrain surface is obtained using the lowest points in the local areas, and triangulated to a TIN. Further points are classified as terrain and added iteratively to the TIN if they fall within pre-defined description of length and angle criteria. The main disadvantage of surface-based methods is that the existing blunders in laser scanning data could affect the estimated surface and shift it upwards or downwards (Sithole and Vosselman, 2004). Another type of methods, which have been proposed for the laser data classification, are occlusion-based methods proposed by Habib et al. (2009). The concept of these approaches is based on occlusions caused by off-terrain points under perspective projection. In these methods, a digital surface model (DSM) is initially generated from the irregular laser points. The presence of occlusions is then discerned by sequentially checking the off-nadir angles to the lines of sight connecting the DSM cells and a pre-defined set of synthesized projection centers. Detected occlusions are then used for the identification of off-terrain points.

All previously mentioned approaches are point-based methods in which individual points have been considered as the classification entities. The major drawback of these approaches is focusing on the discontinuities between the points regardless of the nature of the objects they belong to, which might lead to unreliable classification results. In order to overcome the problems of these approaches, a new group of methodologies have been suggested for the classification of laser data based on the segmentation results (Jacobson and Lohmann, 2003; Sithole, 2005; Sithole and Vosselman, 2005). In these methods, the laser point cloud should be firstly segmented into homogenous regions by aggregating the points with similar attributes. Then, discontinuity measures between neighboring segments (height, distance…) are utilized as the decision rules for the classification of distinct segments into terrain or off-terrain surfaces. The disadvantage of the approaches in this category is that they do not consider the physical properties of individual segments (area, slope…) in the classification process.

In this paper, an alternative approach is presented for the classification of laser scanning data based on the segmentation results from a newly developed clustering approach. This approach considers the physical properties of segmented clusters together with discontinuities between neighboring segments to improve segmentation-based classification results. Figure 1 illustrates the outline of the proposed methodology.

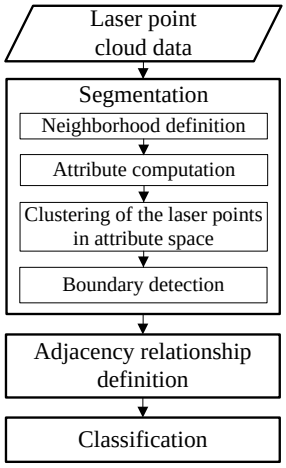


Figure 1. The outline of the proposed methodology.

The next section describes the implemented technique for the segmentation of the laser point cloud using an attribute space. This discussion is followed by an introduction of the proposed method for the classification of the segmentation results. Afterwards, experimental results with real data are presented to verify the feasibility of the proposed approach for terrain and off-terrain classification of airborne and terrestrial laser scanning data. Finally, concluding remarks and recommendations for future work are summarized.

SEGMENTATION

The main purpose of the segmentation process is to abstract the laser point cloud into distinct regions enclosing spatially connected points with similar attributes. These regions will later be used as primitive entities for efficient laser data classification. In this paper, a clustering-based segmentation technique is proposed in which the laser points are aggregated into homogenous regions based on the attributes defined by an adaptive neighborhood of individual points. The main steps of this segmentation procedure will be briefly described in the following sections.

Neighborhood Definition

Neighborhood definition is the primary step of any 3D laser data processing activity. This definition is a rule that determines the neighbors of each point, and as a result has a great impact on the computed attributes for laser data segmentation. In this paper, an adaptive cylinder neighborhood definition is introduced and employed which takes the proximity of the points in 3D space, physical shapes of associated surfaces (Filin and Pfeier, 2005), and expected noise level in the dataset into account.

The adaptive cylinder neighborhood of each point in question is defined as follows: At first, a spherical neighborhood of the point in question is determined. This neighborhood encloses *n* nearest neighbors of the point in question, where *n* is the number of the points needed for reliable definition of a plane while considering the existence of outliers. Afterwards, the best fitted plane for the points in the spherical neighborhood is computed using an iterative plane fitting process. Since, both airborne and terrestrial laser data are dealt with in this research and these datasets comprise planar surfaces with different slopes, a slope-intercept representation of 3D plane is used to select the suitable form of the plane for each point. The selected plane's parameters are then refined using the least square adjustment process. Finally, the adaptive neighborhood is defined by determining the points whose normal distances to the best fitted plane are less than half of the cylinder height. The cylinder height is an adaptive parameter which is defined according to the expected noise level (σ) in different datasets. The schematic concept of the adaptive cylinder neighborhood definition is illustrated in Figure 2.

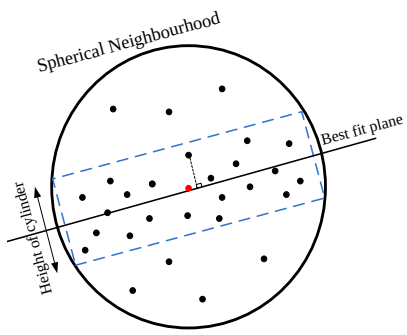


Figure 2. 2D representation of the adaptive cylinder neighborhood definition.

The adaptive cylindrical neighborhood can be utilized to classify individual laser points before further processing. The point in question is deemed to be part of a planar surface if the majority of the points in its spherical neighborhood belong to the cylindrical neighborhood. Otherwise, that point is considered to belong to a rough surface. In order to speed up the segmentation process, a region growing algorithm is employed to aggregate the classified points in planar or rough groups based on their 3D distances. The segmentation is performed for the individual groups rather than the entire point cloud.

Attribute Computation

In this research, the segmentation attributes are computed based on the adaptive cylinder neighborhood defined for individual points. These attributes are the coordinates of the origin's normal projection (X_0, Y_0, Z_0) on the best fitted plane to the neighboring points of the point in question (determined through adaptive cylinder neighborhood definition) as it is shown in Figure 3.

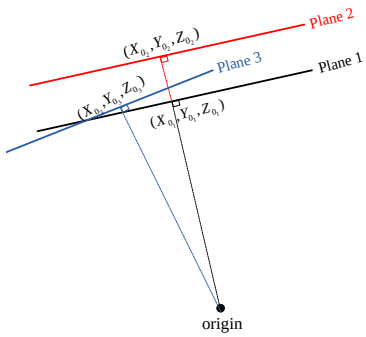


Figure 3. Computed attributes for the segmentation of planes.

Clustering of the Laser Points in Attribute Space

Once the segmentation attributes are computed, a clustering approach should be utilized for the detection and extraction of the accumulated peaks in the attribute space. These peaks will represent clusters of the points with similar attributes in the object space. In this research, the computed attributes in the previous step are organized in a kd-tree structure, which is manipulated for peak detection. A novel peak detection approach is then introduced based on the constructed kd-tree; for which, an appropriate extent of the cluster that considers acceptable deviation among the attributes of coplanar points is estimated. In this approach, all points in the attribute space are checked to count the number of neighboring points within the established appropriate extent for coplanar points. The attribute point with the highest count is chosen as the peak (the most frequent attribute). The points in the spatial domain whose attributes are the constituents of the detected peak are considered as a segmented planar patch. Then, all the points in the detected peak are removed from the attribute kd-tree. After their removal, a search for the second highest peak is carried out. This process is repeated until the number of detected attributes in the identified peak is less than a pre-specified count that corresponds to the smallest region to be segmented.

Boundary Detection

The disadvantage of the proposed segmentation technique is that the points belonging to the coplanar but spatially disconnected planes will be segmented into the same cluster. In order to resolve such an ambiguity, a neighborhood analysis is conducted through boundary detection of the clustered points. In this research, the modified convex hull algorithm (Jarvis, 1977) is adopted to determine the boundaries of individual clusters. The detected boundaries will then be utilized for the definition of adjacency relationships among clusters in the classification process.

CLASSIFICATION

In this paper, a heuristic technique is proposed for the classification of segmented point cloud into terrain and off-terrain surfaces. This approach is developed based on physical properties of the derived clusters (such as area and slope), planimetric distances, and vertical discontinuities between adjacent clusters. For this purpose, the adjacency relationship between derived clusters is defined using their detected boundaries.

The classification process starts with the identification of off-terrain segments. In the first step, the physical properties of individual segments (slope and area) are checked for the discrimination of off-terrain surfaces. A segmented cluster with steep slope is considered to belong to an off-terrain object if its area is less than a predefined threshold (Figure. 4).

Figure 4. Classification of the clusters with steep slope.

The next group of off-terrain segments are the ones which are significantly higher than their surrounding adjacent segments. The height difference and planimetric distance between adjacent clusters are checked as the discontinuity measures for the detection of this group of segments. In this case, the non-classified segments are firstly sorted based on the average height of their boundary points. Starting from the highest segment, the segments

which are significantly higher than their 2D adjacent segments within predefined planimetric distance are classified as off-terrain surfaces (Figure. 5).

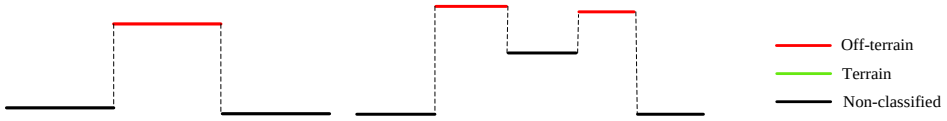


Figure 5. Classification of the clusters based on the height difference measure.

Following on, the segments which are at the same height or slightly higher than their 2D off-terrain classified adjacent segments are classified as off-terrain objects (Figure. 6).

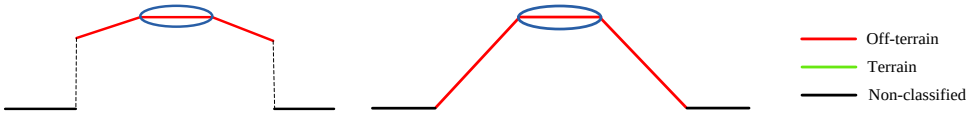


Figure 6. Classification of the segments which are slightly at the same height of their off-terrain classified adjacent segments.

Once the off-terrain segments are classified through the aforementioned steps, the process of terrain and off-terrain classification of the segmentation results is followed by determining the clusters belong to the terrain surface. A non-classified segment is considered to be part of terrain surface when it is relatively large and implicitly lower than its adjacent segments. In order to detect these segments, the remaining non-classified segments are sorted based on the average height of their boundary points. The segment with minimum average height whose area is larger than a predefined threshold is considered to be part of the terrain surface. The rest of non-classified segments are checked in an ascending order based on their average height. For each segment in question, if its average height is not significantly more than the average height of its 2D nearest terrain-classified segment, while considering their planimetric distance, it is classified to be part of the terrain surface; otherwise, it is classified as an off-terrain object (Figure. 7).

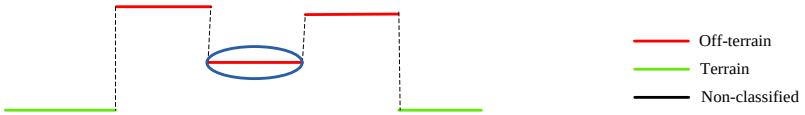


Figure 7. Classification of the remaining non-classified segments.

At last, the classification of the non-segmented (rough) points which are assorted in distinct groups is carried out by comparing the height difference between them and their nearest terrain-classified segment. If the majority of the points in each group are significantly higher than their 2D nearest terrain-classified segment, that group will be classified as an off-terrain object; otherwise it is classified to be a part of terrain surface.

EXPERIMENTAL RESULTS

The performance of proposed methodology for the classification of laser scanning data will be evaluated by conducting a set of experiments using real datasets. The test datasets which have been selected for these experiments include an airborne laser dataset over the British Columbia Institute of Technology (BCIT) campus in Vancouver, Canada and a terrestrial laser data from a building façade in University of Calgary campus. The airborne laser data has been collected using Leica ALS50 and the terrestrial laser data has been obtained using a Trimble GS200 3D laser scanner. These datasets comprise different features such as complex buildings, vegetation and terrain surfaces.

In order to derive accurate results, the processing parameters should be adjusted for each laser dataset based on the scanning system properties. Table 1 lists the processing parameters used for these datasets.

Table 1. Processing parameters of provided dataset

	Airborne laser dataset	Terrestrial laser dataset
σ (expected noise level)	80 cm	4cm
Normal distance threshold (Δd)	80 cm	5cm
Anglular deviation threshold (Δα)	10°	12°
Size of the minimum detectable cluster	10	50
height difference threshold	1m	0.5m
Planimetric distance threshold	2m	1m

Airborne Laser Dataset

Figure 8(a) illustrates an orthophoto and figure 8(b) shows the airborne laser dataset over the BCIT campus. The results of segmentation (figure 8(c)) and classification (figure 8(d)) processes of this dataset have been projected onto the orthophoto to visually verify the quality of the derived results. In figure 8(d), the planar segments which have been classified as terrain are shown in green, and the ones that have been classified as off-terrain objects are shown in red. The groups of rough points that have been classified as a part of terrain surface are shown in blue, and the ones which have been categorized as off-terrain classes are shown in magenta. Visual analysis of the presented results verifies the feasibility of the proposed approach for the classification of airborne laser data.

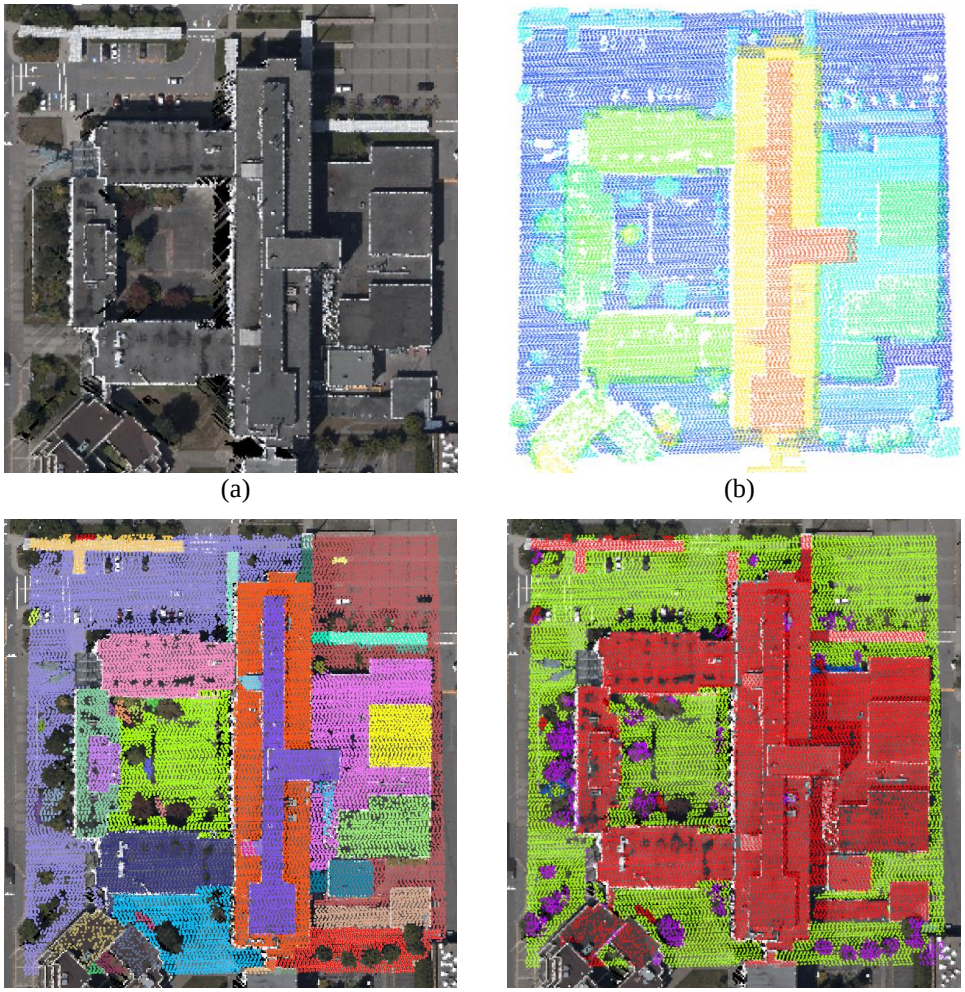
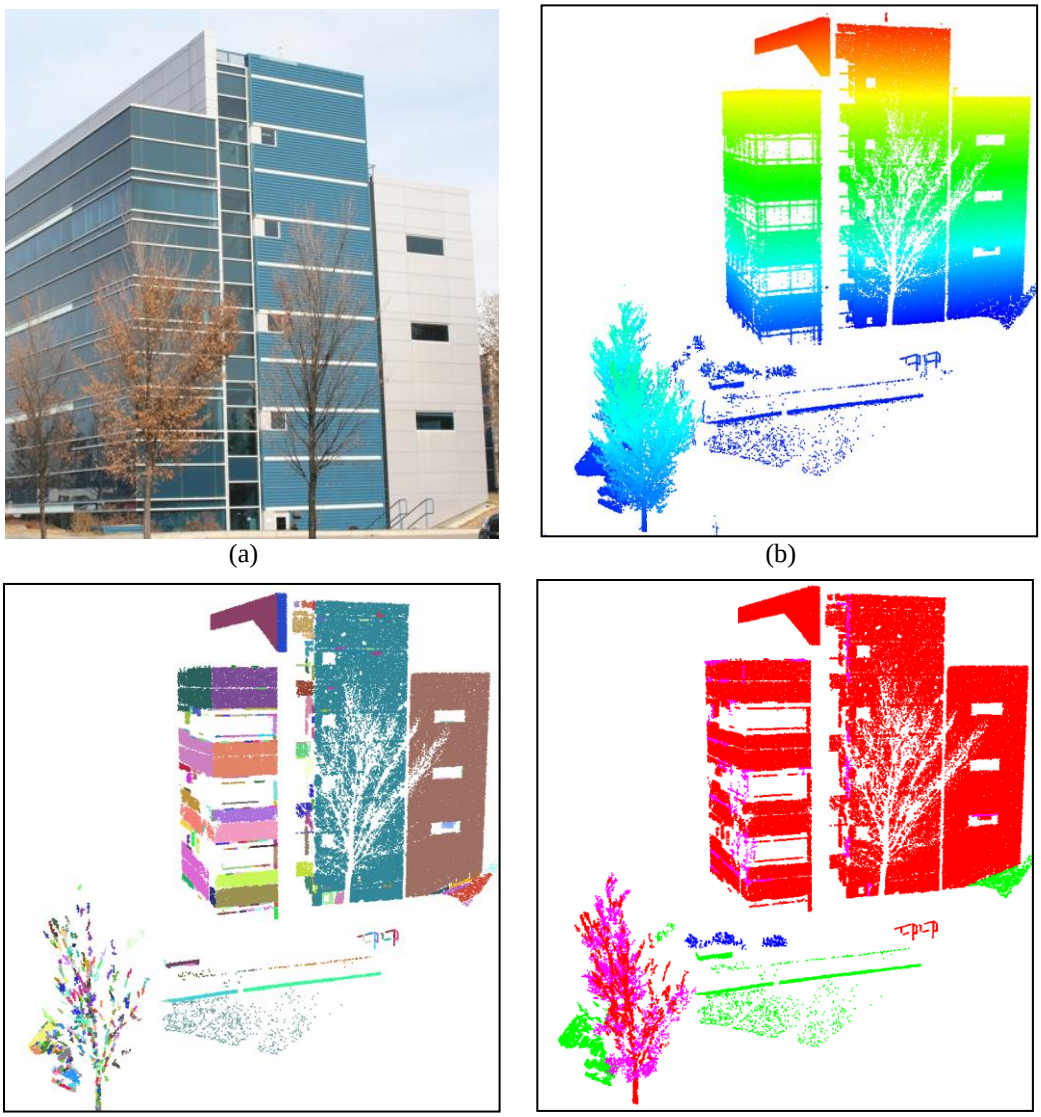


Figure 8. Airborne dataset, orthophoto (a), original point cloud (b), segmentation result (c) and classification result (d).

Terrestrial Laser Dataset

Figure 9(a) and 9(b) show a digital image and a terrestrial laser scan from a building façade within the University of Calgary campus. The results of the segmentation and classification processes of this dataset have been illustrated in figures 9(c) and 9(d), respectively. In figure 9(d), the planar segments which have been classified as terrain are shown in green, and the ones that have been classified as off-terrain objects are shown in red. The groups of rough points that have been classified as a part of terrain surface are shown in blue, and the ones which have been categorized as off-terrain classes are shown in magenta. The inspection of derived results and available imagery data verifies the robustness of this approach for the classification of terrestrial laser data.



(a) (b)

(c) (d)

Figure 9. Terrestrial dataset, orthophoto (a), original point cloud (b), segmentation result (c) and classification result (d).

CONCLUSIONS AND RECOMMENDATIONS FUTURE WORK

In this paper, a new approach was presented for terrain and off-terrain classification of the laser scanning data based on the segmentation results. In order to provide the classification entities, a segmentation approach is introduced and implemented through three main procedures: neighborhood definition, attribute computation, and clustering the points with similar attributes. In this approach, the adaptive cylindrical method is initially applied to

define the neighborhood for each of the points. This neighborhood definition considers the local surface trend and expected level of the noise in each dataset to increase the flexibility of the proposed methodology for handling point clouds from different sources (airborne and terrestrial). Afterwards, the segmentation attributes are computed using the parameters of the best fitted plane to each point's neighborhood. A novel peak detection procedure is then employed for the identification and extraction of the clusters in the attribute space. The boundaries of the clusters are detected to resolve the segmentation ambiguities and define adjacency relationships among the derived segments. Once the laser point cloud is segmented and the adjacency relationship between derived segments is defined, the terrain and off-terrain classification procedure is carried out based on the physical properties of individual segments (slope and area), the vertical discontinuity, and horizontal separation between adjacent segments. The experimental results prove that the introduced technique can provide a reliable solution for the classification of the laser datasets from multiple sources (airborne and terrestrial).

Future research work will focus on the quality control and improvement of the segmentation results to provide more accurate information for classification and object extraction procedures. Furthermore, intensity values of the laser points and other external sources such as images will be applied in further processing steps for better interpretation of the derived surfaces.

ACKNOWLEDGEMENT

The authors would like to thank the GEOIDE (GEOmatics for Informed DEcisions) Network of Centers of Excellence (project: PIV_SII72), Natural Sciences and Engineering Research Council of Canada Discovery and Strategic Project Grants, and Tecterra for their financial support. They are also grateful to McElhanney Consulting Services Ltd, BC, Canada for providing the airborne dataset.

REFERENCES

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. In *International Archives of Photogrammetry and Remote Sensing*, XXXIII (B4): 111-118.

Chen, Q., Gong, P., Baldocchi, D. and Xin, G. 2007. Filtering airborne laser scanning data with morphological methods, *Photogrammetric Engineering and Remote Sensing*, 73(2): 175-185.

Elmqvist, M., 2001. Ground estimation of laser radar data using adaptive shape models. In *Proceedings of OEEPE workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models*, Stockholm, Sweden: 73-80.

Filin, S., Pfeifer, N., 2005. Neighborhood systems for airborne laser data, *Photogrammetric Engineering & Remote Sensing*, 71 (6): 743–755.

Habib, A., Kim, E., Kim, C., 2009. Occlusion-based methodology for the classification of the LiDAR data, *Photogrammetric Engineering and Remote Sensing*, 75(6): 703-712.

Jacobsen, K. and Lohmann, P., 2003. Segmented filtering of laser scanner DSMs, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIV (3/W13).

Jarvis, R.A., 1977. Computing the shape hull of pointsin the plane, In *Proceedings of IEEE Computer Society Conference on Pattern Recognition and Image Processing*: 231-261.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data, *ISPRS Journal of Photogrammetry & Remote Sensing*, 53 (4): 193-203.

Lee, H. S. and Younan, N. H., 2003. DTM extraction of LiDAR returns via adaptive processing, *IEEE Transactions on Geoscience and Remote Sensing*, 41(9): 2063-2069.

Sithole, G., 2001. Filtering of laser altimetry data using a slope adaptive filter, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34 (3/W4): 203-210.

Sithole, G., 2005. Segmentation and Classi_cation of Airborne Laser Scanner Data, Ph.D.thesis, TU Delft, Netherlands.

Sithole, G. and Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds, *ISPRS Journal of Photogrammetry and Remote Sensing*, 59 (1-2): 85-101.

Sithole, G. and Vosselman, G., 2005. Filtering of airborne laser scanner data based on segmented point clouds. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36 (3/W19): 66-71.

Vosselman, G., 2000. Slope based filltering of laser altimetry data, In *International Archives of Photogrammetry and Remote Sensing*, XXXIII (B3): 935-942.

Zhang, K. Q., Chen, S. C., Whitman, D., Shyu, M. L., Yan, J. H. and Zhang, C. C., 2003. A progressive morphological filter for removing non-ground measurements from airborne LiDAR data. *IEEE Transactions on Geoscience and RemoteSensing*, 41 (4): 872-882.