

## A THRESHOLD-FREE FILTERING ALGORITHM FOR AIRBORNE LIDAR POINT CLOUDS BASED ON EXPECTATION-MAXIMIZATION

Z. Hui<sup>1</sup>, P. Cheng<sup>1\*</sup>, Y. Y. Ziggah<sup>2</sup>, Y. Nie<sup>1</sup>

<sup>1</sup> Faculty of Geomatics, East China University of Technology, Nanchang, China – huizhenyang2008@163.com, pgcheng1964@163.com, nieyunju@126.com

<sup>2</sup> Faculty of Mineral Resources Technology, University of Mines and Technology, Tarkwa, Ghana - yziggah@umat.edu.gh

Commission III, WG III/5

**KEY WORDS:** Airborne LiDAR, Point clouds, Filtering, Ground points, Expectation-maximization, Intensity

### ABSTRACT:

Filtering is a key step for most applications of airborne LiDAR point clouds. Although lots of filtering algorithms have been put forward in recent years, most of them suffer from parameters setting or thresholds adjusting, which will be time-consuming and reduce the degree of automation of the algorithm. To overcome this problem, this paper proposed a threshold-free filtering algorithm based on expectation-maximization. The proposed algorithm is developed based on an assumption that point clouds are seen as a mixture of Gaussian models. The separation of ground points and non-ground points from point clouds can be replaced as a separation of a mixed Gaussian model. Expectation-maximization (EM) is applied for realizing the separation. EM is used to calculate maximum likelihood estimates of the mixture parameters. Using the estimated parameters, the likelihoods of each point belonging to ground or object can be computed. After several iterations, point clouds can be labelled as the component with a larger likelihood. Furthermore, intensity information was also utilized to optimize the filtering results acquired using the EM method. The proposed algorithm was tested using two different datasets used in practice. Experimental results showed that the proposed method can filter non-ground points effectively. To quantitatively evaluate the proposed method, this paper adopted the dataset provided by the ISPRS for the test. The proposed algorithm can obtain a 4.48% total error which is much lower than most of the eight classical filtering algorithms reported by the ISPRS.

### 1. INTRODUCTION

Airborne Light Detection and Ranging (LiDAR) has been developing very fast in the past decades. Comparing with traditional remote sensing means, airborne LiDAR owns the strength of high efficiency in collecting geographic information of large areas (Liu, 2008; Meng et al., 2010). Moreover, weather conditions have no influence on this technique when gathering point clouds (Vosselman and Maas, 2010). Thus, airborne LiDAR has been widely used in many areas, such as digital terrain model (DTM) extraction (Li et al., 2017; Ozcan and Unsalan, 2016), three-dimensional model generation (Yang et al., 2017; Zhou and Zhou, 2014), forest parameter estimation (Hyypä et al., 2012; Vega et al., 2016), etc.

Although airborne LiDAR is cost-effective when acquiring three-dimensional information, it takes up a lot of time in processing point cloud data, especially in manual classification. It has been established that the manual classification and quality control consume an estimated 60–80% of processing time (Flood, 2001). Consequently, it is urgent to develop algorithms to speed up point clouds post-processing efficiency.

To develop algorithms for many other applications, one fundamental step is to extract DTM information from point clouds that contain both ground points and non-ground points. This process is generally known as filtering. Aiming at realizing filtering effectively, lots of algorithms have been put forward in the past twenty years. These algorithms can be categorized into four classes, namely slope-based (Vosselman, 2000) (Sampath,

2005; Sithole, 2001; Susaki, 2012), morphology-based, interpolation-based (Chen et al., 2007; Hui et al., 2016; Hui et al., 2017; Li et al., 2013; Li et al., 2014; Mongus et al., 2014; Pingel et al., 2013), and segmentation-based (Tóvári and Pfeifer, 2005) (Lin and Zhang, 2014) (Chen et al., 2016). Although these algorithms in the literature can obtain a good performance, most of them suffer from parameters setting or thresholds adjusting. Obviously, it is inconvenient for inexperienced staff. Moreover, many parameters setting or thresholds adjusting is always time-consuming and reduces the degree of automation of the algorithm. To overcome this problem, this paper proposed a threshold-free filtering algorithm based on expectation-maximization.

### 2. METHODOLOGY

According to central limit theorem, naturally measured sample data will lead to a normal distribution. Conversely, due to the complex terrain environments point clouds can be assumed as a mixture of Gaussian models. Therefore, the separation of ground points and non-ground points from point clouds can be seen as a separation of a mixed Gaussian model.

Expectation-maximization (EM) is applied for realizing the separation. EM is an approach for fitting probability distributions and can calculate maximum likelihood estimates of parameters to probabilistic models being fit to data. When we do not know which component (ground or object) the point belonging to, EM can be used to calculate maximum likelihood estimates of the mixture parameters. Using the estimated

\* Corresponding author

parameters the likelihoods of each point belonging to ground or object can be computed. It is obvious that the point is labelled as the class corresponding to the maximum likelihood. The detailed procedure of the proposed algorithm comprises three steps as presented in Sections 2.1 to 2.3, respectively:

### 2.1 Point Clouds Denoising

Due to the influence of external environments or the laser rangefinder malfunction, the acquired point clouds always contain noisy points, including high outliers and low outliers. Both of these two outliers may disturb the assumed normal distribution; especially the low outliers may have a great influence on the final filtering results. Hence, the outliers should firstly be removed.

To realize denoising, the point clouds are first organized using  $k$ -dimension tree. A point is eliminated if its elevation value changes greatly before and after morphological opening operation among its  $k$  neighbors.

The morphological opening operation is achieved by performing an erosion of the dataset followed by a dilation given as Equation (1):

$$\begin{cases} E_p = \min_{(x_i, y_i) \in Neighbors} (z_i) \\ D_p = \max_{(x_i, y_i) \in Neighbors} (z_i) \\ O_p = D_p(E_p) \end{cases} \quad (1)$$

where  $x_i, y_i, z_i$  = coordinates of  $p$  point's neighbors  
 $E_p$  is morphological erosion operation  
 $D_p$  is morphological dilation operation  
 $O_p$  is morphological opening operation

Points are selected as outliers if their elevation changes greater than the threshold that can be calculated automatically according to Equation (2):

$$\begin{cases} Z_{th} = 3 * Z_{std} \\ Z_{std} = \sqrt{\frac{1}{k} \sum_{k=1}^k (Z_k - Z_{mean})^2} \\ Z_{mean} = \frac{1}{k} \sum_{k=1}^k Z_k \end{cases} \quad (2)$$

where  $Z_{th}$  = threshold for detecting outliers  
 $Z_{std}$  = standard deviation of the neighbors  
 $Z_{mean}$  = mean elevation of the neighbors  
 $Z_k$  = elevation of the  $k$ -th point

### 2.2 Ground Points Extraction Using EM

To realize the filtering, the posterior probability of a point  $p$  belonging to ground points ( $G$ ) should be calculated. It can be determined according to Equation (3):

$$P(G|p) = \frac{P(p|G)P(G)}{P(p)} \quad (3)$$

where  $P(p) = \sum P(p|G)P(G)$   
 $P(G)$  is the prior probability of ground points

It is obvious that  $p$  will be labelled as a ground point if  $P(G|p)$  is greater than 0.5. Since we assume that two classes namely ground points and object points form the mixed Gaussian models, it is easy to set  $P(G)$  to be 0.5. Thus, to obtain the posterior probability of a point belonging to ground points, we need to calculate the class-conditional density  $P(p|G)$  given as Equation (4), which can be estimated using EM algorithm.

$$P(p|G) = \sum_{j=1}^2 P_j G(p|\mu_j, \delta_j) \quad (4)$$

where  $P_j$  is the mixing probability  
 $G(\bullet)$  is the Gaussian equation with parameters  $\mu_j$  and  $\delta_j$ , which are mean and standard deviation of elevations.

EM algorithm is a general method for fitting probability distributions that mainly includes the following four steps (Dempster et al., 1977):

- i Initializing mixture parameters to random values.
- ii 'E' step: computing membership probabilities.
- iii 'M' step: updating mixture component parameters.
- iv Checking for convergence. If not, repeating steps ii and iii.

EM algorithm keep iterating until it reaches convergence. Meanwhile, the posterior probability of a point  $p$  belonging to ground points or object points is obtained. Obviously,  $p$  is classified as the class with larger posterior probability. According to the binary classification results filtering outcome can be achieved.

### 2.3 Optimization Using Intensity Information

Nowadays, most airborne LiDAR systems provide intensity information which can help to optimize the filtering results obtained at Section 2.2. Owing to the characteristic that same materials have similar reflection intensities, the error-classifying points obtained by the EM method can be revised according to intensity information. Since point clouds are organized using  $k$ -dimension tree at Section 2.1, it is easy to find the  $k$  nearest neighbors of one point. If one point's intensity value is greatly different from that of its neighbors, we can infer that this point is wrongly classified. Meanwhile, the corresponding label should be modified.

## 3. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed algorithm was tested using two datasets used in practice. The first dataset (Sample1) was acquired using an

Optech ALTM scanner from the city of Jingmen in China with an area of 1.20 km<sup>2</sup> and 984,998 points cloud data in total. The second dataset (Sample2) was located within the city of Luoding, China, characterized by modern architecture with low and high-storey buildings. Both of these two datasets contain X, Y, Z coordinates and intensity information. Figures 1 and 2 are intensity images of the two areas. It can be found that these two areas contain different terrain features.

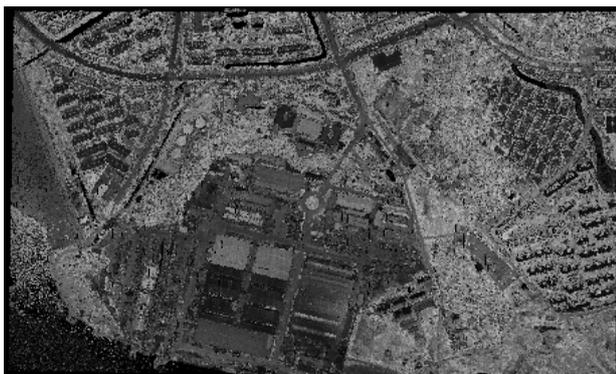


Figure 1. Intensity image of the first dataset



Figure 2. Intensity image of the second dataset

Three indicators namely Type I error, Type II error and Total error are applied for accessing the performance of the proposed method. Type I error is referred to as the percentage of rejecting ground points as object points, while Type II error denotes the percentage of accepting object points as ground points. Total error is the percentage of all the misclassifying points. The three indicators are defined in Equation (5):

$$\begin{cases} Type1 = \frac{b}{a+b} \\ Type2 = \frac{c}{c+d} \\ Total = \frac{b+c}{a+b+c+d} \end{cases} \quad (5)$$

where  $a$  = the count of correctly classified ground points  
 $b$  = the count of incorrectly classified ground points  
 $c$  = the count of incorrectly classified object points  
 $d$  = the count of correctly classified object points

To test the effect of intensity information on the filtering, this paper compared the performance before and after using the intensity data. The comparison results are shown in Figure 3. It can be found that both the performances of Sample1 and Sample2 are improved by applying the intensity information to

the filtering results in terms of total error. It is because the intensity data commonly provides more semantic information which will help revise the filtering results. Conversely, the Type I error of Sample1 turns a little larger after using the intensity information. This can be explained that there are many different natural objects and water areas in Sample1. These complex objects or areas are easy to generate intensity noise, which will affect the filtering results.

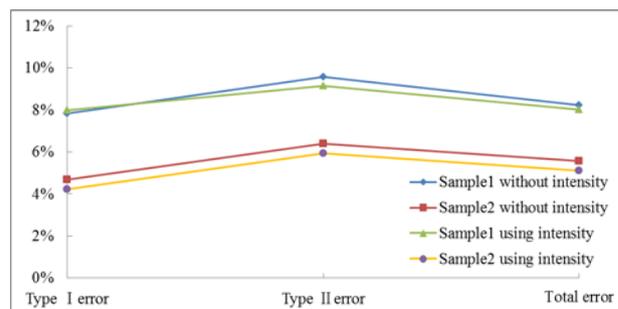


Figure 3. Performance comparison results

In order to more objectively evaluate the filtering accuracy of the proposed method, this paper selected one sample data (sample\_21) provided by the ISPRS and compared its total error with the one of other eight classical filtering algorithms tested by the ISPRS (Sithole and Vosselman, 2004) as shown in Figure 4. Sample\_21 was selected since it contains some special features, including roads with bridges, data gaps, vegetation, etc., which constitute filtering challenges. From Figure 4, it can be found that the total error of the proposed method is only slightly higher than that of the methods proposed by Axelsson and Pfeifer. Considering that both of these two methods involve complex parameters setting, the proposed method owns great advantages in automation.

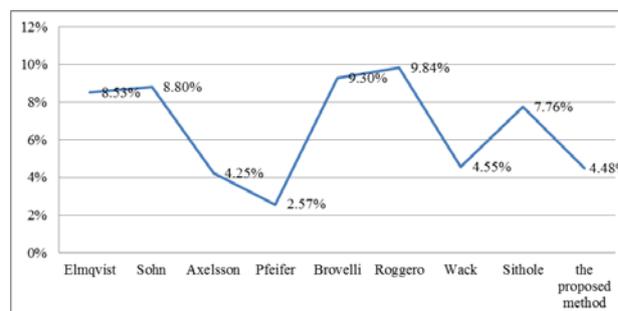


Figure 4. Total error comparison

#### 4. CONCLUSION

Point cloud filtering is a necessary step in the point cloud processing, analysis and applications. To break through the limitation of complex parameter settings for the existing filtering algorithms, this paper proposed a threshold-free filtering algorithm based on expectation-maximization. In this paper, the filtering is seen as a separation of mixed Gaussian models. By applying EM algorithm to elevations of point clouds, ground points can be extracted automatically. Intensity information is also adopted for refining the filtering outcomes. Experimental results show that the proposed method can achieve a good performance without complex parameters setting or thresholds adjusting. Furthermore, the filtering

performance is little affected by the format and resolution of the airborne LiDAR data. Therefore, the proposed method can provide a good foundation for the post-processing of airborne LiDAR point clouds.

#### ACKNOWLEDGEMENTS

The authors would like to thank the Fund from the Education Department of Jiangxi Province (Project No: GJJ170449) and East China University of Technology Ph. D. Project (Project No: DHBK2017155) for their financial support.

#### REFERENCES

- Chen, C. et al., 2016. An improved multi-resolution hierarchical classification method based on robust segmentation for filtering ALS point clouds. *International Journal of Remote Sensing*, 37(4): pp. 950-968.
- Chen, Q., Gong, P., Baldocchi, D. and Xie, G., 2007. Filtering Airborne Laser Scanning Data with Morphological Methods. *Photogrammetric Engineering & Remote Sensing*, 73(2): pp. 175-185.
- Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1): pp. 1-38.
- Flood, M., 2001. LIDAR activities and research priorities in the commercial sector. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W4): pp. 3-7.
- Hui, Z., Hu, Y., Yevenyo, Y.Z. and Yu, X., 2016. An Improved Morphological Algorithm for Filtering Airborne LiDAR Point Cloud Based on Multi-Level Kriging Interpolation. *Remote Sensing*, 8(1): pp. 35.
- Hui, Z., Wu, B., Hu, Y. and Ziggah, Y.Y., 2017. Improved progressive morphological filter for digital terrain model generation from airborne lidar data. *Applied Optics*, 56(34): pp. 9359-9367.
- Hyppä, J., Holopainen, M. and Olsson, H.K., 2012. Laser Scanning in Forests. *Remote Sensing*, 4(10): pp. 2919-2922.
- Li, Y. et al., 2013. A gradient-constrained morphological filtering algorithm for airborne LiDAR. *Optics & Laser Technology*, 54: pp. 288-296.
- Li, Y. et al., 2014. Filtering Airborne Lidar Data by Modified White Top-Hat Transform with Directional Edge Constraints. *Photogrammetric Engineering & Remote Sensing*, 80(2): pp. 133-141.
- Li, Y. et al., 2017. Airborne LiDAR Data Filtering Based on Geodesic Transformations of Mathematical Morphology. *Remote Sensing*, 9(11): pp. 1104.
- Lin, X. and Zhang, J., 2014. Segmentation-Based Filtering of Airborne LiDAR Point Clouds by Progressive Densification of Terrain Segments. *Remote Sensing*, 6(2): pp. 1294-1326.
- Liu, X., 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography*, 32(1): pp. 31-49.
- Meng, X., Currit, N. and Zhao, K., 2010. Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sensing*, 2(3): pp. 833-860.
- Mongus, D., Lukač, N. and Žalik, B., 2014. Ground and building extraction from LiDAR data based on differential morphological profiles and locally fitted surfaces. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93: pp. 145-156.
- Ozcan, A.H. and Unsalan, C., 2016. LiDAR Data Filtering and DTM Generation Using Empirical Mode Decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1): pp. 360-371.
- Pingel, T.J., Clarke, K.C. and McBride, W.A., 2013. An improved simple morphological filter for the terrain classification of airborne LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 77: pp. 21-30.
- Sampath, A., 2005. Urban DEM Generation from Raw Lidar Data: A Labeling Algorithm and its Performance. *Photogrammetric Engineering & Remote Sensing*, 71(2): pp. 217-226.
- Sithole, G., 2001. Filtering of laser altimetry data using slope adaptive filter. In: *International Archives of Photogrammetry & Remote Sensing*, 34(3/W4): pp. 203-210.
- 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): pp. 85-101.
- Susaki, J., 2012. Adaptive Slope Filtering of Airborne LiDAR Data in Urban Areas for Digital Terrain Model (DTM) Generation. *Remote Sensing*, 4(12): pp. 1804-1819.
- Tóvári, D. and Pfeifer, N., 2005. Segmentation based robust interpolation-a new approach to laser data filtering, In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 36: pp. 79-84.
- Vega, C., Renaud, J., Durrieu, S. and Bouvier, M., 2016. On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. *Remote Sensing of Environment*, 175: pp. 32-42.
- Vosselman, G., 2000. Slope based filtering of laser altimetry data. In: *International Archives of Photogrammetry & Remote Sensing*, xxxiii: pp. 935-942.
- Vosselman, G. and Maas, H.G., 2010. *Airborne and Terrestrial Laser Scanning*. CRC, Boca Raton, pp. 34-36.
- Yang, B. et al., 2017. Automated Reconstruction of Building LoDs from Airborne LiDAR Point Clouds Using an Improved Morphological Scale Space. *Remote Sensing*, 9(1): pp. 14.
- Zhou, G. and Zhou, X., 2014. Seamless Fusion of LiDAR and Aerial Imagery for Building Extraction. *IEEE Transactions on Geoscience & Remote Sensing*, 52(11): pp. 7393-7407.