A APPROACH FROM POINT CLOUD PLUS FEATURE DATA TO GRID DEM

Zhikuan Shi[a,b] , Guoan Tang[a*]

[a]Key laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, Jiangsu, P.R.C., 210046
[b]The Research Center for China Land Policy(Nanjing Agricultural University) No. 6,Weigang Road, Nanjing, Jiangsu, P.R.C., 210095

Commission III, WG III/3

KEY WORDS: DEM, LIDAR, Point Cloud, Feature Data ,Pattern Recognition

ABSTRACT:

Data sets acquired by LIDAR consist of land surface information and non-surface information, through some filter methods can get rid of non-surface information, but leave behind non-data area, some importance terrain feature maybe disappear. Using those data to construct DEM can't keep precise. we present an PCPFD algorithm for constructing a grid DEM from point cloud and feature data. The empirical results show that the PCPFD algorithm is more availability within the interpolation based on Nearest neighbour finding phase and the filter phase, lots of computation time saved. The precision of the result Grid DEM satisfy national standard proved by GPS points.

1. INTRODUCTION

The airborne Light Detection and Ranging (LiDAR), or named Airborne Laser Scanning (ALS), is a relatively new technology designed for the acquisition of terrain information. The development of the airborne LiDAR system started in the 1970s, primarily in North America.

The airborne LiDAR system consists of three components: a laser scanner, a Global Positioning System (GPS) and an Inertial Navigation System (INS). The airborne LiDAR data (or ALS data) are aggregated as three dimensional (3-D) point clouds. The clouds include ground points (on the bare Earth) and non-ground object points (on vegetation and artificial objects, such as buildings, bridges, tow

ers, power lines, and so forth). A number of applications have been developed for LIDAR data, such as the generation of digital elevation models (DEMs), topographic mapping, environmental monitoring, forest resource management, and cyber city modeling. This technology has also been accepted as a critical monitoring tool by powerful geospatial information users, including mapping and disaster management agencies, oil and gas exploration companies, the telecommunications industry, pipeline companies, and environmental agencies.

Data sets acquired by LIDAR consist of hundreds of millions of high-resolution points and are too large to processing and extracting a Grid DEM. Many manual construction and vegetation points data hide in the mass LIDAR points, Thus, some essential step must be taken to move spikes and errors due to noise. The topographical maps consist of many Feature Data which can be used to distinguish terrain points and non-terrain points, the valley lines, the ridge lines and so on can be used to extract DEM. An new approach developed to combine the two sorts of data so that to generate huge amounts of high-resolution DEM data quickly and accurately.

Even if we can get usable ground points from the Point Cloud, the number of remain points is too large to load these points to memory at one time, some valuable computational methods such as interpolation methods often infeasible to use directly on even moderately large points sets. Therefore, many practical algorithms use a segmentation scheme that decomposes the ground points into a set of non-overlapping segments, each containing a small number of input points. One then interpolates the points in each segment independently. Numerous segmentation schemes have been proposed, including simple regular decompositions and decompositions based on quad trees. At the same time the Feature Data can be used to found the data boundary, which is very helpful to segment.

2. MATERIALS AND METHODS

Accurate coordinate especially elevation of topography is of very importance to construct Grid DEM. The data used in the experiment is LIDAR Point Cloud in Nanjing and Xuzhou city, Jiangsu province and the correspond Feature Data of the same region, Feature Data can be got through GPS. but the workload is too hard to accomplish in an greatness region with fixup map scale (such as 1:5000). A new method is collected 3D feature data through Inpho DTMaster. DTMaster is a graphical editing and measurement tool for LIDAR / DTM data. The user can measure and edit data in different views and is allowed to store vector data in a layer based structure.

We can get the road edge by DTMaster, the valley lines and so on.
LIDAR can measure distance from air to land surface, the point clouds's coordination must be calculated by GPS and INS in

* Guoan Tang: tangguoan@njnu.edu.cn
Supported by High-Tech Research and Development Program of China 863 (2006AA12Z212)

aerocraft, at the same we must compute the aerocraft position and pose, so we distributed 346 GPS control point in ground, which is in UTM projection system.

2.1 LIDAR Data Acquisition

Processing is supported by a field survey that also could serve to obtain specific information on hydraulic structures such as bridges and actual heights of buildings. Besides DEM extraction, LIDAR data could be used to derive base-line data on land use to derive hydraulic roughness data.

2.2 Model Approach

How can we handle mass point clouds data sets consider feature line? May be we have a powerful enough computers with large memories to load all points in at one time. But the amount of LIDAR point too large to do it. To make large data sets useful to the wider audience that have commodity processors, however, we need algorithms that can use a small amount of memory to manage mass point clouds.

Several types of algorithms are used to process large geometric data sets: divide-and-conquer algorithms, which cut a problem into small sub-problems which can be solved independently ; cache efficient algorithms, which cooperate with the hardware's memory hierarchy (caches and virtual memory); external memory algorithms, which exercise control over where, when, and how data structures are stored on disk (rather than trusting the virtual memory); and streaming algorithms, which sequentially read a stream of data (usually once, perhaps in several passes) and retain only a small portion of the information in memory. All of these algorithms try to exploit or create spatial coherence.

The LIDAR point clouds are collected point by point, the data have some spatial sequence in it, so divide-and-conquer algorithms can solve this problem.

In order to get high precision DEM, terrain character must be taken into account seriously.In this paper we describe a PCPFD (Point Cloud Plus Feature Data) algorithm for constructing a grid DEM from LIDAR points based on quad-tree segmentation. The PCPFD algorithm consist of three separate phases; First: the filter phase, where the topographical maps be used to filtrate the non-terrain points and get the terrain points; Second: the segmentation phase, where the decomposition is computed based on Point Cloud; Third: the interpolation based on neighbor finding phase, where a null grid will be created ,then for each grid cell elevation , the relevant neighboring Point Cloud will be find to compute it, the restrict condition is the neighbor relation can't span the Feature Data. In this paper, We will focus on the interpolation based on neighbor finding phase and the filter phase.

2.2.1 Filter Phase: The Optech ALTM LIDAR instrument can afford range file, which is process through ZinView quick processing software plus ground GPS information and IMU data to generated LAS format raw point clouds data.

The airborne LiDAR contains error datas or Outliers, before filter phase ,these Outliers must be found out and got rid of raw data. These task can be done recur to the software SCOP DTMaster in an visual LIDAR edit environment by man-machine conversation manner.
The raw LIDAR point clouds are covered by building, vegetation and other offground points, some filter steps must be taken to get real ground point. Firstly the filtering of buildings, Secondly a hierarchical filtering separates ground points and offground points. When using Strategy strong in scop++, buildings include small buildings are correctly removed , vegetation can be removed applying vegetation height information from the field and using SCOP++ filtering techniques. The vegetation height information is created from the real vegetation height data collected by GPS, these height information is removed from the LIDAR point clouds by using a DEM filtering package SCOP++ to remove vegetation cover.

2.2.2 Segmentation Phase

The LIDAR data sets are often larger than the main memory of commodity computer, and overwhelm the algorithms and data formats used to manage and analyze them. Not like general spatial data process software which can load data to main memory at one time, LIDAR process software have to be designed to handle large amounts point clouds stored in main memory and large hard disk, it is CPU

The goal of Segmentation is to reduce the high frequency accessing disk, five buffer will be used to perform Segmentation, We first construct the a bottom buffer, when buffer is full , neighboring Points will be putted in the other buffer, when four buffers are full all, the information of four buffer will write to disk, and description information will store in the fifth buffer. When query the Point Cloud one can query the data in memory buffer, if no found ,the disk points will re-load to the memory buffer.

Using Segmentation, Point clouds is organized as one or several cells, each of which contains one or more sub-cells. cells are considered elementary coherent survey areas. Data is acquired in flights collecting one or more tracks of data that may cover a single or multiple cells. While flights and tracks result from sequential data acquisition and thus represent the temporal structure of a project, regions and cells reflect its spatial organization. A flight will usually be done with a single lidar sensor and one set of calibration values. A cells may, however, contain data from several flights with different sensors and calibration sets.

Cells are subdivided into square sub-cells for manageable main memory sizes which themselves are divided into small cells for rapid location-based access. Sub-cells hold in basic lidar data coordinates. This organization also makes it easy to include additional attributes without having to change the basic file structure and access mechanisms.

Segmentation methods partition point clouds into index cells, and include interview flags in the cells that indicate when no more points in the collected data will fall in specified cells. When the interview flag is true show the cell it indicated can be treated independently.

The order of index cells may be random, but we can compute GRID cells in one index cells one by one, the points number of one index cell is finity so we can load the points in one index cell to main memory buffer.
A Segmentation methods makes a small number of sequential passes over a data file, and processes the data using a memory buffer whose size is a fraction of the points clouds length. First pass make certain cells size which can be determined by manual or by automatism, the relation ship between cells must be

maintenanced in this pass. Second pass then we implement index cells and non-leaf cells and leaf cells that read the mass point clouds as input, and produce leaf cells in sequence mesh formats. The memory footprint of the GRID cells is typically less than 1% of the GRID cells in hard disk .

2.2.3 Interpolation Based on Nearest Neighbour Query Methods:

There are different interpolation methods, such as Method of geometrical nearness; Statistical methods based on weighted average; Methods using basis functions; Method of artificial neural networks. etc. There are about eight interpolation methods are always used to generate DEM. Inverse distance weighted averaging (IDWA) is a deterministic estimation method where values at unsampled points are determined by a linear combination of values at known sampled points. Optimal inverse distance weighting is a form of IDWA where the power parameter is chosen on the basis of minimum mean absolute error. Splining may be thought of as the mathematical equivalent of fitting a long flexible ruler to a series of data points. Polynomial regression is a stochastic, global technique which fits the variable of interest to some linear combination of regressor variables. Trend surface analysis (TSA) can be thought of as a subset of polynomial regression. TSA is a stochastic technique which separates the data into regional trends and local variations. The lapse rate method uses the relationship between temperature and elevation for a region to estimate temperatures at unsampled sites. Kriging is a stochastic technique similar to inverse distance weighted averaging in that it uses a linear combination of weights at known points to estimate the value at an unknown point. Cokriging is similar to kriging except it uses additional covariates, usually more intensely sampled, to assist in prediction. Cokriging is most effective when the covariates are highly correlated. All eight spatial interpolation methods investigated accept irregularly scattered data and can create a regular grid of interpolated points amenable to contouring. The spatial interpolation methods differ in their assumptions, local or global perspective, and deterministic or stochastic nature.

We will calculate the GRID cell elevation from the neighboring Point Cloud, but the number of points can be ascertain according to the Interpolation method, once the number of neighboring Points be confirmed, we can search Points near by the GRID cell, the elevation of these points will be push into a stack, a validate algorithm can be implemented to proven the line which between the neighboring point and GRID cell span the feature data. The capability of stack is limited, when the stack is full, a Interpolation method will be executed bases on the points in the stack.

The restrict is the points in stack and GRID cell point can't through breakline.

PCPFD algorithm description: we can use Point Cloud Plus Feature Data to create GRID DEM elevation.(Figure 1)

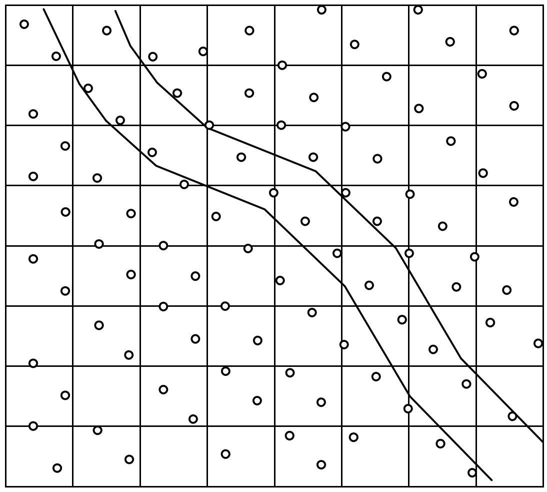


Figure 1. PCPFD algorithm

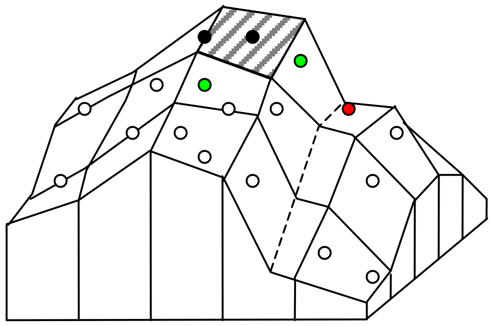


Figure 2. PCPFD Interpolation

When some points are on some vegetation or non-terrain at it happens, these points must be filtrated out in the filter phase to generate the true ground elevation in order to construct digital elevation model. so some GRID cell must Interpolate its elevation from points which is not within it, in this condition, the stack may contain some points which across Break Lines or Form Lines. In figure 2 ,the point (black colour)in the the GRID cell which have diagonal line is happen on the vegetion which be filtrated in filter phase, the interpolation can use point in green colour but can't use the red colour point because a feature line appear between in the red colour point and the GRID cell. Restrict search point can create GRID DEM, but the GRID cell size always be confined by state criterion .

Then how to store the DEM? We bring forward a new digital elevation model. This model contains two layer like RS Multispectral data, but half cell size exists in the 2 layer west south coordinate. Like the Figure 3.

These two GRID DEM is interpolated separately, and store together.

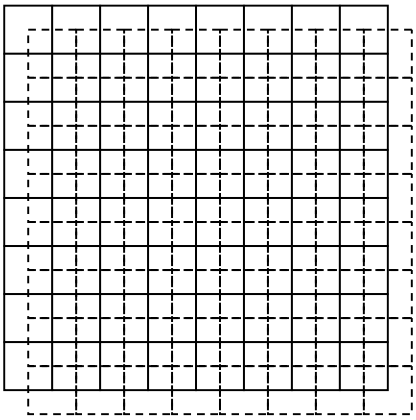


Figure 3 PCPFD storage model

3. RESULTS

We select XuZhou city in JiangSu province experimental point clouds data to validate our algorithm.

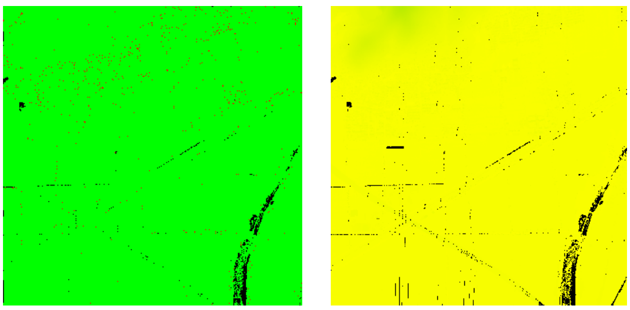
The experiment data include only one flying line, the lidar point clouds instrument is Optech which collect 55977576 points.

Optech ALTM Zinview software is used to calculate the point clouds coordination.

Inpho DTMaster software is used to collect the breakline in the sample area.

One 2×2km sample area is selected 687826 points in it.(in Figure 4.1)

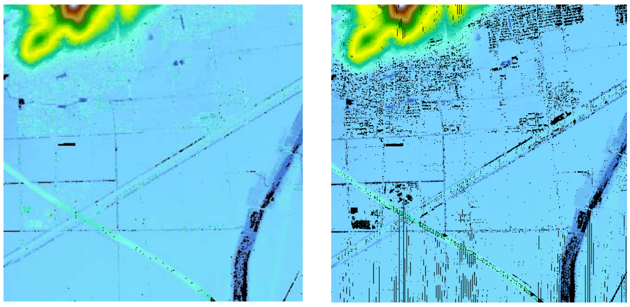
In order to get high precision, some overlap area must be appeared when Optech LIDAR scan point data , in filter phase these points be disposed to get rid of redundancy or non-consistent, the result is appeared in Figure 4.2.



1 2

Figure 4 pretreatment in filter phase

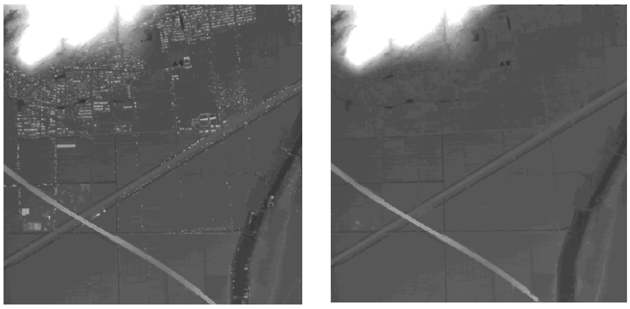
Some Outliers points appears in the sample data, point clouds are cleared up according to the elevations of sample area , 677311points keep down, the result is appeared in Figure 5.1. Then filtrating building and vegetation , the filtrated data contains 624385 points the result is appeared in Figure 5.2.



1 2

Figure 5 filter phase

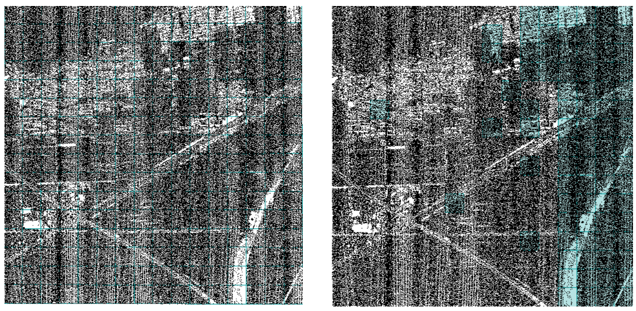
For the sake of the result of filter phase, the DSM and DEM are created which is appeared in Figure 6.



1 2

Figure 6 Filter result contrast

In the segmentation phase 256 index grid cells and 4094 leaf cells are created, need 6.9 seconds.



1 2

Figure 7 Segmentation phase

In the figure 7-2 The points on the left have not been put into segmentation index cells, The points on the right have been put into segmentation index cells and their leaf cells and GRID cells written out. At this moment, the GRID cells elevation are interpolated according to points in leaf cells.

In the Interpolation phase, some breaklines which is collected by DTMaster are participated in these phase, when searching points to interpolate the DEM cell elevation.

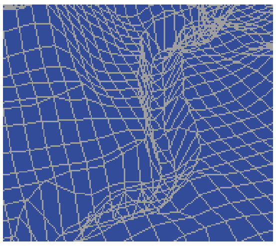


Figure 8 Interpolation result

4. DISCUSSION

The pretreatment point clouds is very importance and pivotal to our method. When some mistake data presence in raw data, we can't compute the correct grid cell elevation through interpolation based on neighboring search. How to get rid of these mistake data or outlier? Although there are many successful method in literature, but when the collection data scalar is larger then the main memory, all these method must faced one fact that is data not access directly. So finding spatial outlier detect method based on large data amount will be always a challenge and significative.

In the filter phase, we must distinguish ground point and off ground point, the off ground point commonly belong to a surface of building or something, recognise these points need a territorial method, but we must dispose the input data one point by one point.

In the Segmentation phase, the LIDAR point clouds can be divided into rectangle cells and we can dispose these point cell by cell, this divide-dispose method we have discussed here need one hypothesis that is the point clouds have some spatial characteristic, there are existence high spatial coherence among the sub data set in cell. For data sets with no spatial coherence at all, we too advocate sorting. But in our experience, large, real-world data sets have plenty of spatial coherence especially in LIDAR point clouds data sets.

Through the Segmentation phase, we can use the data's spatial coherence characteristic to partite point clouds into index cells and control or direct the data process with small efforts, rather than fight it head on. The breaklines can also be dived into these index cells as early as by any possibility. The breaklines have two attribute set, one is spatial the other is flag attribute. The spatial attribute include the cells which across by, generally this information is registerd by the rows and cols of the index cells or by the spatial order of space fill curve (such as Z-curve or Hibernate-Curve). The flag attribute include the spatial morphology which can show the breakline's spatial characteristic(such as furbish information or roughness).This information can be uses in Interpolation phase. How to organize this information more efficiency?

A depth-k quadtree/octree is choosed to organized the index cells because we can describe it succinctly with a bounding box and integer k, and it is relatively simple to determine which cells a sphere intersects. We believe it is possible to eliminate the first pass of Segmentation by computing a quadtree/octree partitioning adaptively during the second pass. Binary space partitions, k-d trees, and many other spatial subdivisions would work too. If a point stream is sorted along a space-filling curve like a Hilbert or z-order curve, the stream is chunked, and Segmentation can be implicit.

In the Interpolation phase which PCPFD algorithm is operated, we need to query Nearest Neighbour point of the GRID cell, at the time, the path between searched point and GRID cell can't through feature line. We have do many experiment to reduce the time invested in Nearest Neighbour query. But the result is not satisfied.

By using R-tree or R+-tree ,the query time can be cut down markedly, but constructing and maintenance the R-tree or R+-tree need abundant CPU spending, and the stability must be treated in real earnest.

By using Z-curve or Hibernate-Curve can be used to reduce the query time when only use GRID DEM cell also, but the complexity will be increased when take the feature line data into account.

The effort spent on these Nearest Neighbour query algorithm and new data structure will be worthwhile and valuable.

Some index technology in large point clouds set can also be used to quicken the Nearest Neighbour query, because the point number of LIDAR point clouds's data sets is too large to load it to main memory at one time, so the index technology must consider not only the memory and CPU but also the I/O spending.

The new DEM store model is designed to improve the DEM precision in a relative little store space, but how to use this model to carry out spatial analysis? How to apply this model in real work? Whether exist a new better store model? Survey aircraft equipped with a LIDAR range detection system produce maps of terrain consisting of millions of non-uniformly sampled points in 3D. Whether these Point Cloud can express DEM and how to use it .To resolve this problem need through more investigation deeply.

The point cloud could be converted to an GRID DEM and transmitted in real time. Currently however, techniques for producing viewable images from point cloud data sets are neither time or memory efficient.

In future work we plan to determine whether the feature detection and recovery stages can be combined to make the algorithm even more robust. Some datasets contain isolated peeks without any creases. For example, the well known cow dataset has two peaks at the horns. These peaks are singleton points in the crease pattern and not detected by our algorithm. Their neighborhood is also quite difficult to mesh with polygonal faces. We want to investigate how to fit cones to the peaks and how to incorporate them into polygonal mesh representations.

5. CONCLUSIONS

In this paper we presented some I/O and CPU efficient solutions for constructing a GRID DEM from a large amount LIDAR point clouds. We implemented our algorithm and, using LIDAR data, experimentally compared it to other existing algorithms.

The experimental results indicate that the LIDAR point clouds can be used to generate high precision GRID DEM. The filter phase and Segmentation phase and the interpolation phase based on neighbor finding methods is importance .

REFERENCES FROM JOURNALS

Ackermann, F., "Airborne laser scanning — present status and future expectations", ISPRS Journal of Photogrammetry & Remote Sensing, Vol. 54, pp. 64–67, 1999.

Alemseged Tamiru Hailea, Effects of LIDAR DEM Resolution in Flood Modelling: A model sentitivity study for the city of Tegucigalpa,Honduras, ISPRS WG III/3, Enschede, the Netherlands, September 12-14, 2005.

Pankaj K. Agarwal, From Point Cloud to Grid DEM: A Scalable Approach

Suyoung Seo, Refinement of Filted LIDAR Data using Local Surface Properties, MAPPS/ASPRS 2006 Fall Conference, November 6 – 10, 2006.

Tim Warriner, Generating a New High Resolution DTM Product from various Data Sources, Photogrammetric Week 05,pp. 197-206, 2005

M. E. Romano, "Innovation in Lidar Processing Technology," Photogrammetric Eng. and Remote Sensing, vol. 70, no. 11, pp. 1201-1206, 2004.

Katzenbeisser, R. and Kurz, S. Airborne Laser-Scanning, ein Vergleich mit terrestrischer Vermessung und Photogrammetrie. In Photogrammetrie Fernerkundung Geoinformation, Heft 8 2004, pp. 179-187.

Mundt, J. T., Streutker, D. R., Glenn, N. F.(2006): Mapping Sagebrush Distribution Using Fusion of Hyperspectral and Lidar Classifications. In Photogrammetric Engineering & Remote Sensing, Vol. 72, No.1, January 2006, pp. 47 – 54.

Sithole, G., Vosselman, G. Filtering of airborne laser scanner data based on segmented point clouds. In International Archives of Photogrammetry and Remote Sensing, Vol. XXXVI, 3/W19, Enschede, Netherlands, 2005,pp. 66 – 71.

T. T. Vu and M. Tokunaga. Filtering airborne laser scanner data: A wavelet-based clustering method. Photogrammetric Engineering & Remote Sensing, 2004 70(11):1267 – 1274.

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

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

Zhang, K., S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, 2003. A Progressive morphological filter for removing nonground measurements from airborne LIDAR data, IEEE Trans. on Geoscience and Remote Sensing, vol. 41(4):872-882.

Wack, R. and Wimmer, A., 2002: "Digital Terrain Models From Airborne Laser Scanner Data – A Grid Based Approach". IAPRS, Vol. XXXIV Part 3B. ISPRS Commission III, Symposium, September 9 - 13, 2002, Graz, Austria. pp.293-296.

Axelsson, P., 2001, "Ground estimation of laser data using adaptive TIN-models", Proceedings of OEEPE workshop on airborne laserscanning and interferometric SAR for detailed digital elevation models, 1-3 March 2001. Official Publication No.40, CD-ROM. pp 185-208.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. LIU Yongqiang for providing all the necessary data for this study.