A PROGRESSIVE QUALITY CONTROL TO IMPROVE THE ACCURACY OF LIDAR DATA PROCESSING

Xiaodong Zhi [a]*, Liang Zhong [a]

[a] School of Remote Sensing Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

WG I/2 - SAR and LiDAR Systems

KEY WORDS: LiDAR, Quality control, Accuracy analysis, Strips Adjustment

ABSTRACT:

Quick data acquisition using LiDAR, is a way to generate dense accurate DEM. The difference with an image recorded by a standard digital camera is that the LiDAR recorded points do not have a regular distribution. The quality and the true resolution could be variable to traditional photogrammetry. So in application of airborne LiDAR, the accuracy of LiDAR data is ambiguous. Actually LiDAR quality control is a vital post-process used to verify the quality of the data. The objectives of this research are, through analyzing the LiDAR data characteristic and the data processing flow of Airborne LiDAR, research the way to evaluate the LiDAR data accuracy, make clear the quality criteria, study the methods of accuracy evaluating, and bring out a working flow on LiDAR data processing.

1. INTRODUCTION

1.1 About LiDAR

The fields of application for LiDAR are very diverse and include generation of digital elevation models, 3D-city modelling, forestry management, coastline protection, disaster management, erosion studies, archaeology, monitoring of corridors such as powerlines, pipelines, railways and roads. The technology offers short data acquisition time, highly detailed detection of the earth surface and the accuracy fits the needs of many applications.

But from an end user perspective, they need the quality criteria, the methods of evaluating to examine performance of LiDAR and evaluate the accuracy of point clouds. These problems are studied in following chapters.

1.2 Coordinates calculated mathematical model

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{WGS-84} = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}_{WGS-84} + R_{WGS-84(p,o,k)} R_{\mathrm{Re}f(IMU)} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix}_{Laser} \quad (1)$$

Where x, y, z *is* Ground coordinates of the object point

X_0, Y_0, Z_0 is Ground coordinates of the GPS antenna phase center

X_L, Y_L, Z_L is Laser Scanner coordinates of the object point with calibration parameters

$R_{\mathrm{Re}f(IMU)}$ is Rotation matrix that needs to be applied to the laser unit coordinate system to make it parallel to the reference coordinate system

$R_{WGS-84(p,o,k)}$ is Rotation matrix that needs to be applied to the reference coordinate system to make it parallel to the WGS-84 coordinate system

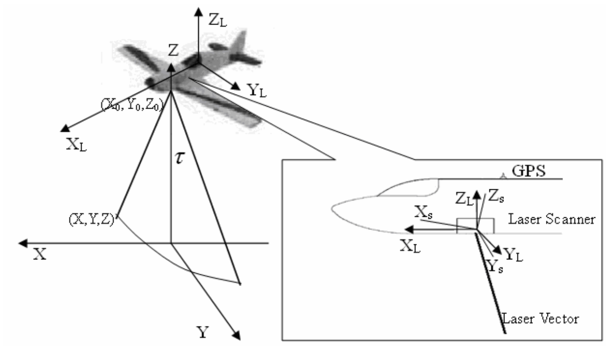

Figure 1. The coordinates of the LiDAR system

1.3 LiDAR quality

LiDAR quality control is an important post-process used to verify the quality of the produced data. The objectives of this research are to develop new methods and tools for LiDAR quality control.

Before talking about quality, quality criteria, we have a look at the definition of the term quality. In ISO9000 (ISO, 2000) quality is defined as "degree to which a set of inherent characteristics fulfils requirements".

In geodesy, the term quality is mostly used synonymous to accuracy.

Accuracy is the degree to which information on a map or in a digital database matches true or accepted values. Accuracy is an issue pertaining to the quality of data and the number of errors contained in a dataset or map.

Generally, the comparison analysis among the various datasets (e.g., LiDAR and photogrammetric DTM, LIDAR and GPS check points, LiDAR and GIS vector data) was performed using different approaches as point to point comparison, point vs surface, profile vs profile, surface (shape) vs shape, to evaluate the quality of Airborne LiDAR.

2. FACTORS AFFECTING ACCURACY OF LIDAR DATA

The quality of laser scanning has been studied during the last few years (e.g. Crombaghs et al., 2002, Ahokas et al., 2003). It has been shown that the terrain height can be typically collected within 15 cm. But, in fact, there are large numbers of factors affecting the quality and accuracy obtained, e.g. the surface material, flight altitude of sensor and platform, GPS/INS and observation angle, and so on.

2.1 The criteria of positional accuracy

The NSSDA (National Standard for Spatial Data Accuracy) uses root-mean-square error (RMSE) to estimate positional accuracy. RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points.

$$RMSE_x = sqrt\left[\sum (x_{data,i} - x_{check,i})^2 / n\right] \quad (2)$$

$$RMSE_y = sqrt\left[\sum (y_{data,i} - y_{check,i})^2 / n\right] \quad (3)$$

$$RMSE_z = sqrt\left[\sum (z_{data,i} - z_{check,i})^2 / n\right] \quad (4)$$

Where:

$x_{data,i}$, $y_{data,i}$, $z_{data,i}$ are the coordinates of the i th check point in the dataset

$x_{check,i}$, $y_{check,i}$, $z_{check,i}$ are the coordinates of the i th check point in the independent source of higher accuracy

n is the number of check points tested

i is an integer ranging from 1 to n

2.2 The errors in LiDAR

Concerning data accuracy, the major error sources in LiDAR are the following:

- Positioning of the platform (Satellite has a major role in GPS positioning reliability. It is quantified by Positional Dilution Of Precision (PDOP). Poor satellite, in other words, a high PDOP, generates inaccurate GPS coordinates.)
- Orientation determination from IMU (The error from the IMU is as systematic differences on strips depend strongly on the error from the IMU. The IMU is one of the main causes of horizontal error in scanned data points, and errors very often increase or decrease with consistency in a flight’s direction.)
- Offsets between the laser sensor, INS/POS equipment and an aircraft platform
- Errors in the electro-optical parts of the laser sensor
- Wrong laser and INS/POS data processing
- Integration and interpolation of the INS and GPS data
- Erroneous data from the reference ground GPS base stations, or, there are no enough base stations
- Data coordinate transformation
- Time latency (In such a system as a whole, the time synchronization of the important role. e.g., GPS and laser finder system synchronization problems. GPS of measuring the rate of general in dozens of Hz, but the laser canner of measuring the rate of in 10 to 100 kHz)

3. WORKING FLOW AND METHODOLOGY

As discussed by the LiDAR equation (Equation 1), there is no redundancy in LiDAR measurements. This is because, due to the random nature of LiDAR points, one cannot measure the exact same point in different strips. Therefore, unlike with photogrammetric data, one cannot use explicit measures to assess the quality of LiDAR derived positional information. Therefore, alternative quality assessment methods are necessary for this type of data.

According to author's data processing experience and other literature, a working flow of data processing based on progressive quality control, is put forward as Figure 2.

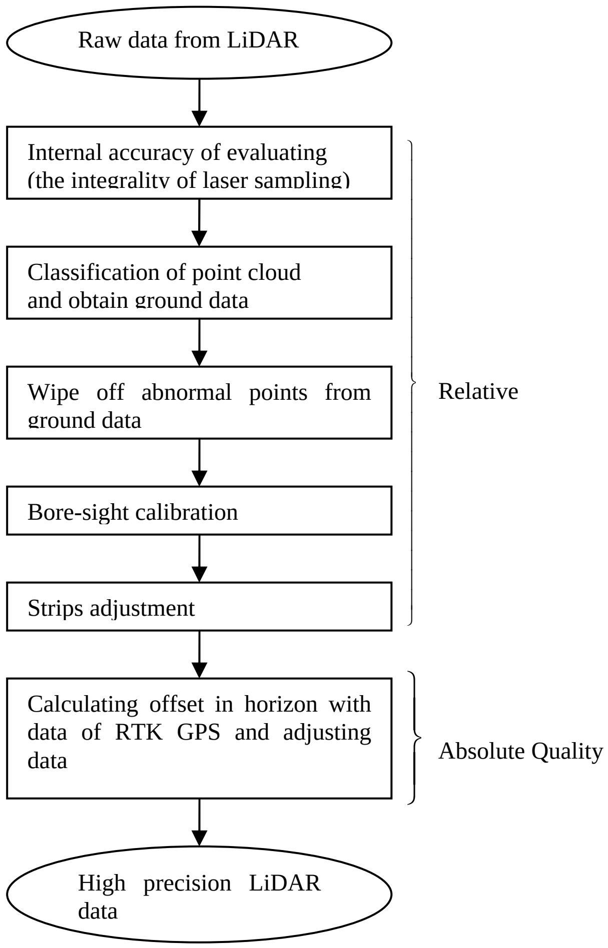

Figure 2. Working flow of LiDAR data processing

Against every step in above figure, its methodology could be discussed following.

3.1 Internal accuracy of evaluating

In data extraction and processing, internal accuracy of the estimates used to evaluate the quality of the data. Although this criteria is not within the very tight precision, but to a certain extent, we can see that the accuracy of the data and the processing of the data quality. Such as the integrality of laser sampling and data packet.

3.2 Classification of point cloud

Above ground objects are problematic for tie point observations. Due to the angle of incidence, foliage and multiple echoes, vegetation is rarely observed in the same way from two different strips. To avoid poor tie points around vegetation or man made objects, the data should be filtered to extract the ground.

Although it is usually quite simple for a human operator to identify what is ground, it is not practical to manually edit large amounts of LiDAR data. There are several methods of automated filtering to choose from including morphological, slope based filters and least square estimators. The results of these systems vary by terrain types (urban, steppe, mountainous) and the density of the LiDAR data. The systems can be compared by effectiveness of building/vegetation removal, speed of algorithm and smoothness of the derived surface.

3.3 Wiping off abnormal points from ground data

In order to wipe off abnormal points from ground data, the filter based in entropy of range image is put forward by author.

The Airborne LiDAR data is a high-density point sampling of the terrain. These data are processed either as TIN elevation surface model or interpolated into a regular elevation surface model (DSM). The LiDAR DEM data can be also converted into a raster image. In addition LiDAR system captures also the intensity of the response signal corresponding to the LiDAR ground point. The intensity is a function of the reflectivity of the ground material and the intensity changes form a georeferenced grey level image like output. Figure 3 shows examples of range (left) and intensity (right) images generated from a LiDAR point cloud.

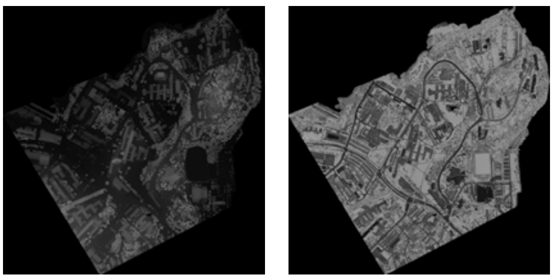

Figure 3. Range image (left) and intensity image (right) generated from a LiDAR point cloud

3.3.1 Entropy: Entropy is the quantitative measure of disorder in a system. The concept comes out of thermodynamics, which deals with the transfer of heat energy within a system.
In image, the entropy is the measurement of information. So in a digital image, there are n gray values, g1, g2, …, g_n. The probability of g_i is p_i. The entropy of this digital image is defined as below:

$$H\left[P\right] = H\left[p1, p2, ..., p_n\right] = -k\sum_{i=1}^{n} p_i \log p_i \qquad (5)$$

Where The probability p_i is approximately equal to:

$$p_i = \frac{f_i}{N} \qquad (6)$$

Where f_i is frequency of g_i
N is the sum of pixels in image

As proved, the most entropy exists in image where distributing homogeneously of gray value.

The entropy of part of image is measurement of information locally. It denotes whether existing feature in part of image.

3.3.2 Procedure:

a) Calculating the entropy of entire image as E.
b) Exporting raster image of elevation from ground LiDAR data after classification.
c) In unit area (the size is defined by experience), calculating the entropy as e. If e > E, considering there are abnormal points.
d) In the area where e > E, calculating average AVR, standard deviation S of elevation. For per point, calculating $\Delta H = abs(H_i - AVR)$, where H_i is the elevation of this point. If $\Delta H > S$, considering this point is abnormal and wiping off from ground data.

3.4 Bore-sight calibration

Bore-sight error is the angular misalignment between the laser sensor unit and IMU. Unlike a photographic image, a bore-sight error affects each observation and cannot be removed by applying a simple affine transformation to the entire strip.

The systematic errors of bore-sight misalignment and scanner torsion error can be expressed in a parametric form. To solve for the unknown values, standard least-squares techniques can be used to determine the parameters in redundant data.

In procedure of adjustment, a method of patch matching is needed to improve accuracy.

3.4.1 Patch Matching: Conjugate planar patches in overlapping strips are supposed to be coplanar, regardless of the flying direction or any other parameters, unless there are biases affecting the data. Using this fact, planar patches can be used to qualify or detect biases in the LiDAR datasets.

In a common area for two overlapping strips, that is represented the same physical surface for conjugate patches. So the angle of normal is the least for conjugate patches. This could be accomplished through an automatic process as described below:

e) Generating a Triangulated Irregular Network (TIN) for overlapping areas.
f) Choose one patch in left strip, matching the conjugate patch in right strip. The condition of marching successfully is that the angle between the first patch and the potential

conjugate patch are compared to determine the degree of similarity between these patches.

g) Once the patch with the best matching parameters is found, both patches are visualized to ensure the correctness of the chosen patch, and to determine the number of points constituting these patches; patches containing more points are preferred.

Then the angle derived from patch matching can be calculated average, standard deviations, RMS to evaluate the quality of Airborne LiDAR data.

In addition, through matching patch, also generating a lot of corresponding points for adjustment.

3.5 Strips adjustment

A three-parameter mathematical model of adjusting is put forward to improve the accuracy of LiDAR data.

If only considering error in height, the error parameters are generalized to three parameters. As follows:

a : elevation constant

b : linear variable along strip

c : linear variable across strip

c

Where b and c are relative to the direction of strip, so firstly transforming coordinates in overlapping strip to one set of coordinates. As follows:

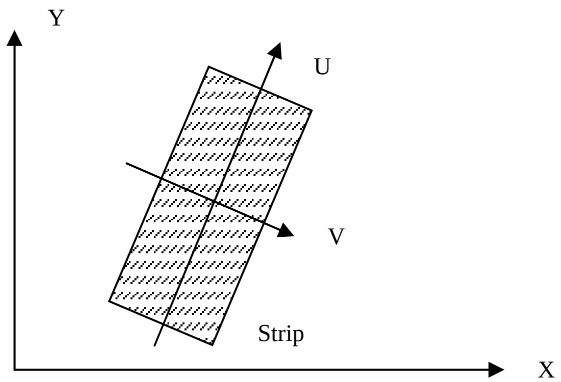

Figure 4. A local coordinate system

In elevation, ΔH can be shown as following formula:

$$\Delta H(U,V) = a + bU + cV \qquad (7)$$

Then the corresponding points' coordinates in overlapping area of different strips are supposed to be coincidence in theory. As following equation:

$$H_{k,i}^{laser} + \Delta H_{k,i}(U,V) = H_{j,i}^{laser} + \Delta H_{j,i}(U,V) \qquad (8)$$

Where: k, j is strip numbers

i is the number of corresponding points

Then in fact the equation of difference in elevation is:

$$H_{k,i}^{laser} - H_{j,i}^{laser} = a_j + b_j U_{j,i} + c_j V_{j,i} - (a_k + b_k U_{k,i} + c_k V_{k,i}) \qquad (9)$$

Tie points were extracted from the overlap area between LiDAR strips. Since a point to point correspondence is not available between LiDAR strips, the points had to be interpolated in order to achieve a match.

Through large amounts of redundancy data, a , b , c can be resolved based on a least squares adjustment procedure.

3.6 Calculating offset in horizon with checkpoints

Elevations of LiDAR derived points were compared with RTK-derived reference points. A circle with a radius of 2 m using a reference point as a centre point of the circle was created for every reference point. Statistics of the LiDAR points were calculated inside the circles if there were more than 5 laser points included. Mean value, median, minimum, maximum and standard deviation, nearest laser point to the reference point and an interpolated height value from the laser points were calculated. A 10 cm by 10 cm grid and a cubic method was used in the height interpolation calculations. The above mentioned statistical values were calculated to find out if there was a difference between mean value, nearest laser point to the reference point and an interpolated height value in the comparison process.

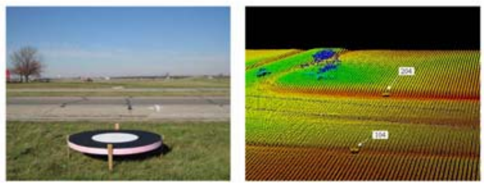

Figure 5. Checkpoints in the wild (left) and point clouds (right)

3.7 Density of point cloud based on TIN

In addition, building digital elevation model uses ground data after classification. Therefore, the precision of digital terrain model acquisition from Airborne LiDAR data mainly by the following factors:

- Accuracy of per point
- Density of points

When the raw point clouds have high density, we can better represent terrain and surface features in survey area. So, the density of point clouds from LiDAR is also an important quality indicator.

The method of calculate density of point clouds is:

- Generating a Triangulated Irregular Network (TIN) using point clouds.
- Calculating area of each triangle in TIN. Then sorting by area and generating histogram.
- From this histogram, we can be aware of data-density distribution of point clouds.

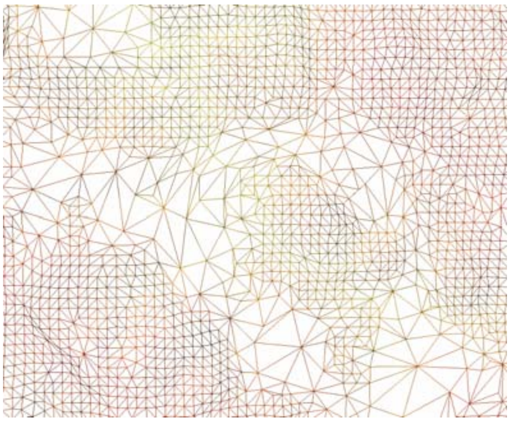

Figure 6. TIN generated from point clouds

3.8 Experimental Results

For this case study, four flight strips are picked. As follows:

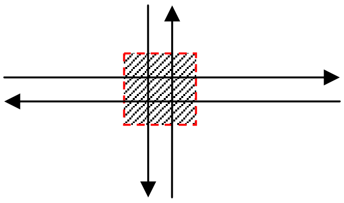

Figure 7. Distribution of four strips

According to the above working flow of LiDAR data processing (Figure 2), the results are shown in the following table:

roll shift	-0.20289257
pitch shift	0.26699907
heading shift	0.05376117
roll std	0.0006
pitch std	0.0027
heading std	0.0015
scale	-0.00001605
start dz	0.7668
final dz	0.0465
z shift	-0.14213259
z std	0.0024

Table 1. Results of data processing

4. CONCLUSION

Generally, processing of LiDAR data is expected to be fast and accurate. However, due to the large amount of data (LiDAR data) that comes out of a single mission, the process of progressive quality control should be planned carefully, with the highest degree of optimization to improve LiDAR data accuracy.

REFERENCES

Goodwin, N. R., Coops, N. C., and Culvenor, D. S., 2006. Assessment of forest structure with airborne lidar and the effects of plat form altitude. Remote Sensing of Environment, 103, pp. 319-330.

Ackermann, F., 1999. Airborne laser scanning – present status and future expectations. ISPRS JPRS, 54, 64-67.

Crombaghs, M.J.E., Brugelman, R., de Min, E.J., 2000. On the adjustment of overlapping strips of laser altimeter height data. International Archives of Photogrammetry and Remote Sensing, 33 (B3/1), 230-237.

ASPRS LiDAR Committee (PAD) ASPRS Guidelines, Vertical Accuracy Reporting for Lidar Data [Online] = ASPRS Guidelines - Vertical Accuracy Reporting for LIDAR Data and LAS Specifications. Ed. Flood, M.. - May 2004. - Augest 2006. http://www.asprs.org/society/divisions/ppd/standards/Lidar%20guidelines.pdf.

Behan, A., 2000. On matching accuracy of rasterized scanning laseraltimeter data. International Archives of Photogrammetry and Remote Sensing 33 (B2), 75–82.

Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. ISPRS JPRS, 54: 199-214.