ISPRS Journal of Photogrammetry & Remote Sensing 56 (2002) 236–245

PHOTOGRAMMETRY & REMOTE SENSING

www.elsevier.com/locate/isprsjprs

Estimating relative lidar accuracy information from overlapping flight lines

Damir Latypov

TerraPoint, 4800 Research Forest Drive, The Woodlands, TX 77381-4142, USA

Received 22 May 2001; accepted 15 March 2002

Abstract

A pure statistical method for estimating relative lidar accuracy by comparing overlapping lidar datasets is described. It has been designed for commercial, fast turnaround, high-volume data production environment and has the following features. It operates on raw data and no building of grid or TIN is required. The method performs accuracy computations on statistical samples that are orders of magnitude larger than ones used in other methods reported in the literature, thus allowing detailed error analysis. Finally, the method provides a lot of relative accuracy information even when no ground control points are available. Software implementation of the method has been written and extensively tested on almost 100 Gb worth of data. Robust and fully automated, it has become an important quality control tool for data processing at TerraPoint. The method begins with the definition of a relative accuracy estimate, which is generalized from being a measure of point-to-point closeness to a measure of surface-to-surface closeness. Thus, a problem of matching the individual points in the overlapping datasets is eliminated. The generalized accuracy estimate is a function of surface size and flatness. Sampling overlap area by surfaces of different size allows one to discriminate between the random, systematic and locally systematic errors. The relative accuracy in the traditional sense of using only some tie points is a limiting case of the generalized one and can also be computed. The concepts and ideas of the method are illustrated by examples from real-life lidar projects. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Lidar; Spatial data accuracy; Strip overlap; Statistical analysis

1. Introduction

Accuracy of the lidar data becomes an important topic as the number of lidar applications and commercial lidar data vendors increases (Krabill et al., 1995; Kraus and Pfeifer, 1998; Gutierrez et al., 1998; Shrestha et al., 1999; Airborne 1, 2001). From the end-user perspective, it often sets a limit on validity of a model derived from the data. From a vendor's point of view, assessing the accuracy is necessary in order to correctly set the specifications, monitor the hardware performance and to improve the technology.

Yet, there is no simple and cost-effective procedure for establishing the accuracy of the spatial data generated by lidar. In fact, the term "accuracy" itself is a source of debates and some confusion. For example, when derived from statistical samples, should the accuracy be defined as an average over

E-mail address: damir.latypov@transamerica.com (D. Latypov).

PII: S0924-2716(02)00047-3

the entire scan width or should it be stated separately for the scan edges where it is likely to be the worst? Confusion sometimes arises when lidar vendors quote accuracy specifications without specifying their confidence level (c.l.) and the procedure used to establish them. The National (USA) Standard for Spatial Data Accuracy (NSSDA) defines the accuracy at 95% c.l. and assumes errors to have a normal distribution. In real life, the assumption of normal distribution is often rather poorly realized, especially at the tails. The outliers make it a bit too optimistic to assume that NSSDA accuracy specification can be derived from 68% c.l., specification—quoted by most vendors—by simply doubling it.

The first and very important step in a system analysis is to make an error budget. It sets a theoretical limit on the performance that can be attained by a particular system. An excellent discussion of an error budget for lidar can be found in Baltsavias (1999). People however agree that the theoretical performance is generally not achievable in the field. Therefore, in order to back up their claims on accuracy some vendors conduct extensive ground GPS surveys and check the lidar data against the control points (Airborne 1, 2001).

While collecting ground control points could be the most logical and straightforward way to assess the accuracy, it is also an expensive and time consuming one. Therefore, such surveys are usually limited to a few dozen points and very extensive and rare ones to a few thousand points. This number represents typically a tiny fraction of the number of collected lidar points. Other approaches for checking lidar accuracy make use of qualitative visualisation tools (e.g. representation of the lidar DSM with shading), or of overlaying vector data on the lidar DSM or orthoimages produced by the lidar DSM. A limiting factor for the last case is that vector data may not exist, or may not have sufficient accuracy or density to allow a reliable and comprehensive evaluation of the lidar accuracy. Keeping in mind that lidar errors are not constant within a flight line but possibly even a single scan line, a realistic survey using few control points simply cannot provide a detailed error analysis.

In this paper, we propose to complement the ground surveying method of assessing the accuracy of spatial data by the method of overlap analysis. This analysis is performed on areas that were covered by two or more flight lines. Adjacent flight lines of a well-planned lidar project should always overlap so as to avoid slivers—areas not covered by any flight line. At TerraPoint we now plan our flight lines to overlap 30% on either side. Thus, a full 60% of all lidar points (except at the border lines) can be checked against each other for consistency.

Various methods of analyzing residuals between the overlapping flight lines have already been reported in literature (Fritsch and Kilian, 1994; Kilian et al., 1996; Crombaghs et al., 2000; Behan, 2000; Maas, 2000). The main goal of these authors was to compensate for discrepancies between two datasets in the overlap area rather than create an overlap-based quality control tool. Most of the methods perform a least squares matching at certain points of the gridded data. A considerable effort has been put into understanding the effects of the grid size and different types of interpolation. It has been found that these methods often produced wrong results due to the rasterization of the data, and were affected by the amount of height variation present in the dataset (Behan, 2000). Maas (2000) has shown that an improvement in the quality of matching results is possible using least squares matching on a TIN of laser points. Further interesting work in this direction is underway at Delft University of Technology, the Netherlands and Dresden Technical University, Germany.

The method we describe here is different from all of the above methods in many respects. First of all, it operates directly on the lidar data points. No grids or TINs—neither of which brings any new information into what is already in the data points—are built. Thus, we gain greater efficiency. We also avoid degrading effects caused by interpolation in case of the grid-based methods (Behan, 2000). Secondly, our method is a purely statistical one and does not require a feature matching or prior segmentation of the datasets. It is based on a generalized definition of accuracy, which to a large extent eliminates the problem of matching individual points in the overlapping datasets. Finally, it utilizes the data in the entire overlap area, as compared to an analysis performed on selected patches of points (Maas, 2000). These features of our method are very important in a production environment where efficiency, reliability and total quality control are of primary concern.

In this paper, we only deal with the estimation of relative accuracy information. It is not our goal here to compensate for any mismatch that may occur between the overlapping datasets; rather we want to quantify it, i.e. find the mean and standard deviation of the local height differences in the overlap area. Therefore, in what follows we assume that the lidar system has been correctly calibrated. This assumption is not really necessary. In fact, the system can be calibrated by minimizing the relative accuracy estimate considered as a function of the calibration parameters. Such a calibration is however a topic in itself and will be discussed elsewhere (Latypov and Zosse, in press).

The task of an overlap analysis can be divided into three parts: (1) finding overlap areas, (2) estimating relative accuracy information and (3) error resolution (i.e. determination of its planimetric and height components). The first part is a relatively straightforward problem of computational geometry although the algorithms and implementation issues are of paramount importance here in view of the sheer volume of the data to be processed. We will not discuss these algorithms in the present paper but will give an example to illustrate their efficiency.

The second part of the analysis—estimating the relative accuracy information—is an interesting statistical problem. It can be formulated in several different ways: as a comparison of overlapping flight lines point by point (similar to the ground control points procedure) or surface by surface. The surface approach has many advantages. It circumvents the problem of matching the individual points in the overlapping datasets and offers a lot of flexibility as one can vary the surface size used for analysis and examine overlap on surfaces with specified flatness. By looking at surfaces of different sizes, one can to a certain degree discriminate between the random, systematic and locally systematic errors. Indeed, the contribution of the random errors into the average height of a sufficiently large surface becomes negligible. For larger surfaces yet the local systematic errors may average out.

Flatness of the surface is also an important parameter of our approach. It provides a link to the standard procedure of defining the vertical accuracy using ground control points over the flat surfaces (Krabill et al., 1995; Carabajal and Harding, 2001).

The surface formulation of the overlap analysis leads to a new definition of an accuracy estimate itself. Instead of asking how close is a lidar point to the "true" point, we can ask: "How close is a set of lidar points to the set of "true" points that they represent?" A user concerned with the accuracy of a surface representation by lidar rather than individual points may find this definition to be more suitable for his or her purpose.

It is not a goal of this paper to present a complete and final solution to all the problems described, especially to the error resolution problem. In this paper, we discuss and illustrate only the following aspects of the overlap accuracy analysis:

(a) How are the surface- and point-based relative accuracy estimates correlated?
(b) What can be said about the absolute accuracy based on the relative accuracy information?
(c) Does terrain type affect the accuracy?
(d) How does the data accuracy change with the altitude at which it has been collected?
(e) How stable is the performance of a lidar system over time?

We hope that the ideas presented in this paper will open a new avenue in research on spatial data analysis and would help the lidar community to converge to a standard unbiased procedure for establishing the accuracy of the data.

2. Overlap analysis

Let S be a surface with the true mean height h_S, and the standard deviation σ_S from this mean. If S is mapped by two lidar datasets, one can ask how close the two representations of the surface that these datasets provide to each other are. For each dataset, we compute the number of points N falling into the area A_S comprised by the surface, the average height of these points $\hat{h}_S$, the standard deviation from this average $\hat{\sigma}_S$ and normalized xz and yz moments, M_{xz}, M_{yz}

$$M_{xz} = \sum_{i=1}^{N} x_i z_i \Big/ \sum_{i=1}^{N} z_i,$$

$$M_{yz} = \sum_{i=1}^{N} y_i z_i \Big/ \sum_{i=1}^{N} z_i \tag{1}$$

Here xyz is a local coordinate system that differs from the geographic coordinate system by a xy-translation to the geometric center of the area A_S.

For a meaningful comparison of the datasets on S, the surface must be sampled uniformly by each dataset. For each dataset the number of points in it should be consistent with its size and the point density. Formulae for lidar data point density are given in Baltsavias (1999). If these requirements are not met the surface should be excluded from further analysis. As a measure of closeness of two datasets on S, we take a difference in calculated average heights of points on S

$$\delta h = \hat{h}_S^{(2)} - \hat{h}_S^{(1)}. \tag{2}$$

Occasionally, one might find a surface evenly sampled by two datasets and still with substantially different average heights. This can happen due to different systematic errors or due to blunders, which shift each average in opposite direction, or due to multiple-valued surfaces (e.g. trees). The blunders often occur at steep surfaces (e.g. mountains, building walls) and can be caused by secondary reflections or slight planimetric difference between the overlapping measurements. It would be interesting to develop an algorithm to automatically classify these situations. Such an algorithm can start with the comparison of the values of the standard deviation $\hat{\sigma_S}$ and the moments M_{xz}, M_{yz} for two datasets. In case of pure systematic height difference between the flight lines for example, both lines will still have close values for $\hat{\sigma_S}$. Blunders will likely result in very different values for the moments. Until such a classification algorithm is in place, surfaces with substantial different average heights may require a visual inspection. For the statistical overlap analysis, we ideally only want to keep single-valued surfaces that do not contain blunders.

Now we have a statistical sample of average height differences $\{\delta h_i, i=1, 2, \ldots, N_S\}$ for all N_S qualified surfaces. We can compute the average value for this sample $\overline{\delta h}$ and the standard deviation $\sigma_{\delta h}$. These values quantify the generalized accuracy.

We can also compute these values for a subset of surfaces with specified flatness (flatter than some threshold value). Let a surface S be described by an equation

$$z = g(x,y) \tag{3}$$

The flatness f of the surface S is best defined as an average (on S) of an absolute value of the gradient of the surface function g

$$f = \frac{1}{A}\int_{A_S} \sqrt{(\partial g/\partial x)^2 + (\partial g/\partial y)^2}\mathrm{d}x\mathrm{d}y \tag{4}$$

where A is the area (size) of A_S. However, it is time-consuming to evaluate Eq. (4) and a simpler though less precise measure of flatness may be considered as a compromise. Usually, we use $\hat{\sigma}_S$ to quantify the flatness. The drawback of such a choice is that the flatness might become size dependent: a large surface of a constant slope for example, will have a larger value of $\hat{\sigma}_S$ than a smaller surface with the same slope. For the purpose of the data quality control though we find this choice quite acceptable.

The two components of the generalized accuracy convey different information about the overlap. The first one, $\overline{\delta h}$, tells us whether the average height of the overlap area (or its subset with specified flatness) in one dataset appears higher or lower than in the other. For good datasets (no problems during the data acquisition and precise calibration) $\overline{\delta h}$ should be very close to zero. This alone, of course, does not guarantee a good quality of the data. It is equally if not more important to have a small value for $\sigma_{\delta h}$ which quantifies the spread of the height differences δh for all qualified surfaces. A value of $\overline{\delta h}$ can be corrected for by simply applying a height shift to one of the datasets. Such a shift does not change $\sigma_{\delta h}$. On the other hand, any transformation that changes the value of $\sigma_{\delta h}$ most likely will also change $\overline{\delta h}$. We will not be discussing such transformations here. They are part of a calibration process described in Latypov and Zosse (in press).

Let us now consider qualitatively the generalized accuracy as a function of sample surface size and its flatness. In general, the larger the size of a sample surface, the less important becomes the contribution of random errors (both planimetric and vertical) into the accuracy. Therefore, at a fixed flatness, one might expect that the $\sigma_{\delta h}$ as a function of the surface size should first decrease and then become a constant (when random errors average out to a negligible level). To quantify this behavior, one needs to know the data acquisition rate for different data streams-inertial, GPS and laser. At 1 Hz GPS and 20 kHz laser shot acquisition rate, a random GPS measurement error will affect in exactly the same way 20,000 lidar data points, thus appearing as a systematic error for a surface built upon these points. On the other hand, $\overline{\delta h}$ should be constant with respect to the sample surface size since a height difference in the overlap area does not depend on how the area is being sampled (by large or smaller surfaces). As function of flatness, we can expect $\sigma_{\delta h}$ to be a monotonically increasing function. Indeed, on rougher surfaces the horizontal shifts result on average in larger height differences.

So far, our discussion has been concerned with the overlap or relative accuracy only. Our next task is to establish the relationship between the overlap accuracy and the absolute accuracy.

2.1. The overlap and the absolute accuracy

The question here is whether we can quantify the absolute accuracy based on relative accuracy computations. Generally, it seems that the answer is "no". Indeed, the datasets may overlap perfectly well but may as a whole be translated or rotated with respect to the true dataset. However, achieving a good overlap accuracy goes a long way toward assuring the quality of the data since these global transformations represent only 6 degrees of freedom. Besides, most of the measurement errors result in local deformations. It is only systematic GPS errors that introduce global translation errors into the lidar data. Global horizontal errors can still be detected by the overlap analysis, if we consider the overlap of several lines flown in parallel and in orthogonal directions. Therefore, at least theoretically, the only component of the error that cannot be detected by the overlap analysis is the systematic vertical error.

3. Algorithm and implementation details

Given a set of lidar flight lines, the data is first prepared for the overlap accuracy analysis as follows:

(1) The outlines of the flight lines are calculated.
(2) The overlapping outlines (and therefore the flight lines) are found.
(3) Points from overlapping flight lines falling into the overlap area are written into the overlap files.

For one of the case studies for this paper, overlap files for 16 flight lines with the total of 92,000,000 points have been created in just under 15 min of CPU time on an SGI Octane 250 MHz machine.

From this point, one can proceed with the overlap accuracy analysis in a number of different ways. Upon a due consideration, it has been decided that the following Monte Carlo type approach provides the most flexibility and efficiency. A random point is selected within the overlap area. A square of a user-defined size is created with the selected point at its center. Lidar points from both flights that fall within this square are found and statistic computations—finding average heights, standard deviations, *xz*, *yz* moments, etc.—are performed. This loop is repeated k times, where k is the ratio of the number of points in the entire overlap area to the number of lidar points within the sample square. The k sample surfaces may overlap, which is of little consequence for the algorithm. The results are written into a database for further queries (calculating statistics on a subset of surfaces with specified flatness, etc.).

4. Overlap accuracy analysis

In this section, we illustrate the concepts and ideas of the overlap analysis by examples from real-life lidar projects completed by TerraPoint. The projects vary in their accuracy specifications and data provided here are for illustration purposes only. For some examples, we intentionally chose datasets of degraded quality (e.g. due to GPS problems) in order to see what effect such errors have on our accuracy estimate.

We start with the discussion of overlap of two flight lines and examine in detail the overlap accuracy as a function of the surface size and its flatness. We quantify the relationship between the surface and point definitions of the overlap accuracy by looking at surfaces of decreasing size. Then, we zoom out to the mission level analysis where we look at the change in the overlap accuracy from one pair of overlapping flight lines to another, all flown in the same day. Finally, we comment on overlap accuracy of the data collected at different altitudes.

4.1. Flight line level analysis

For this case study, we picked two flight lines flown over a mountainous terrain with plateaus and valleys. In the overlap area the elevation was changing from 724 to 1074 m with an average of 887 m and a standard deviation from this average of 69 m. The lines were flown at 1800 m with the shot spacing of 1.5 m across the track and 1.8 m along the track. In overlap area fell 3.9 million points from the first flight line and 4.5 million points from the other one. Table 1 summarizes the results of overlap accuracy analysis performed on surfaces of various sizes with $\hat{\sigma}_S \leq 0.21$ m (for both overlapping datasets). Surface size is quantified by the average number of points (per flight line) on it (column 1) and by the area (column 2). The third column of the table shows the number of sample surfaces. The fourth column gives $\overline{\delta h}$, the difference in average surface heights averaged over the samples. The 68% (95%) interval is the interval that contain 68% (95%) of all the samples. For example, in case of sample surface size of 1102 m^2 (446.1 points per surface), 95% of the surfaces with $\hat{\sigma}_S \leq 0.21$ m, had a difference in height within the interval -0.02 ± 0.31 m.

The data has been analyzed "as is", though for more precise accuracy estimates the undersampled surfaces should have been excluded. By not excluding these surfaces, however, we can only underestimate the relative accuracy of the data.

The results conform to the general discussion of Section 2. The standard deviation $\sigma_{\delta h}$ is at first a decreasing function of the sample surface size A, and then it becomes a constant. The average height difference $\overline{\delta h}$ on the other hand does not depend on the size of the sample surface, again in accordance with the previous discussion. The point-based relative accuracy is equivalent to the case of very small surface size, albeit usually with much less samples. Very small $\overline{\delta h}$ (≤ 0.02 m) for all sample surface sizes indicates a good attitude determination and lack of constant systematic errors within each flight line, or no different systematic errors between the lines.

Table 1
Case study 1: relative accuracy as function of the surface size

Surface size (points)	Surface size A (m^2)	Number of surfaces with $\hat{\sigma}_S \leq 0.21$ m	Average height difference $\overline{\delta h}$ (m)	68% Interval (m)	95% Interval (m)
1.8	N/A	867,807	0.00	0.22	0.51
8.2	9	372,021	− 0.01	0.19	0.42
18.6	30	124,524	− 0.02	0.17	0.38
51.5	104	27,754	− 0.02	0.16	0.34
149.4	343	2804	− 0.02	0.16	0.37
446.1	1102	589	− 0.02	0.17	0.31
760.5	1923	240	0.00	0.16	0.31

Table 2
Case study 1: relative accuracy as a function of the surface flatness

Surface flatness $\hat{\sigma}_S$ (m)	Number of surfaces	Average height difference $\overline{\delta h}$ (m)	68% Interval (m)	95% Interval (m)
0.06	687	0.03	0.16	0.27
0.09	6326	− 0.04	0.17	0.38
0.12	12,948	− 0.06	0.16	0.38
0.15	18,448	− 0.05	0.16	0.33
0.18	23,214	− 0.06	0.17	0.35
0.30	40,864	− 0.06	0.17	0.37
0.61	65,024	− 0.01	0.18	0.45
1.52	78,069	0.04	0.21	0.99
3.05	81,662	0.04	0.23	1.19
15.24	89,582	0.08	0.27	6.63

Surface size: 101 m^2 (50 points).

Next, we performed the relative accuracy calculations for surfaces of different flatness. Results are summarized in Table 2.

The surface size for this study has been fixed at 101 m^2 (about 50 points on a surface patch per flight line) and the number of sample surfaces has been set to a maximum of 90,000. The second column of the Table 2 gives a number of sample surfaces, with flatness less than corresponding value of $\hat{\sigma}_S$. The third, fourth and fifth columns of Table 2 have the same meaning as in the Table 1.

The table shows that the average height difference and 68% interval are slowly varying functions of the flatness while the 95% interval increases drastically for large $\hat{\sigma}_S$. This is a good illustration to the note made in the introduction on relationship between 68% c.l. and 95% c.l. specifications. Visual inspection proved that surfaces with very different heights in two overlapping datasets were indeed examples of the multiple-valued and undersampled surface situations discussed in Section 2. This is not surprising when flying over mountains.

4.2. Mission level analysis

The purpose of the first study in this section was to check whether a certain intermittent degradation of GPS signal could be detected through the overlap analysis. The study is based on 6 lines flown at altitude of 1000 m over a relatively flat terrain with average elevation of about 21 m and standard deviation from this average of 3 m. The relative geometry of the lines is shown in Fig. 1.

Lines 1 and 2 were orthogonal to the lines 3, 4, 5 and 6 and there were seven pairs of overlapping flight lines. Their overlaps were analyzed on surfaces with $\hat{\sigma}_S \leq 0.21$ m. The size of a sample surface was 407 m^2 and it contained on average 188 points from each of the two overlapping flight lines. Since the lines 1 and 2 were orthogonal to the other ones, their overlaps have much smaller area and fewer number of sample surfaces. Table 3 provides the results.

One can see a significant inconsistency in results from one overlap to another. The average height difference of the surfaces built upon the datasets from flight lines 1 and 4 is a whopping 21 cm. On the other hand, we see good overlaps of the lines 1 and 3, 2 and 3, 5 and 6. As it turns out, the answer to this puzzle is an intermittent problem with GPS (high PDOP number) that occurred during the flight line 4. The line was not affected in the area of overlap with the flight line 2 and the statistics for this overlap is in line with the expectations. This is a good example of using overlap analysis as a quality control tool.

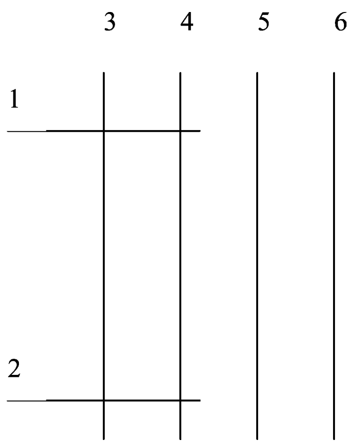

Fig. 1. The schematic arrangement of the six flight lines of the first study (Section 4.2).

Table 3
Case study 2: mission level analysis

Flight lines	Number of surfaces with $\hat{\sigma}_S \leq 0.21$ m	Average height difference $\overline{\delta h}$ (m)	68% Interval (m)	95% Interval (m)
1, 3	83	0.06	0.13	0.27
1, 4	112	0.21	0.20	0.38
2, 3	141	0.01	0.14	0.28
2, 4	180	0.02	0.17	0.30
3, 4	1997	− 0.17	0.23	0.34
4, 5	2609	− 0.11	0.16	0.38
5, 6	13,924	0.07	0.06	0.14

Flat terrain, altitude flown: 1000 m, sample surface size: 407 m^2.

The second study has been conducted in order to see how stable is the performance of our lidar system during a complete 2-h test mission. The mission consisted of 16 flight lines flown at altitude of 2000 m over a rough terrain with elevation changing from 0 to 250 m. The spacing of lidar points was 3.0 m along track and 3.4 m across the track.

The coverage of a continuous area was not a main goal of this test mission and there were only 10 pairs of overlapping flight lines. Overlap areas containing more than 20,000,000 points were analyzed by sampling surfaces with the area of 1250 m^2 and flatness $\hat{\sigma}_S \leq 0.5$ m. The fact that here 0.5 m were used instead of 0.21 m as in Tables 1 and 3 does not influence the accuracy analysis, which is quite insensitive to the selection of this threshold, as long as it is not chosen too large. Results of the study are summarized in Table 4.

Table 4
Case study 3: mission level analysis

Flight lines	Number of surfaces with $\hat{\sigma}_S \leq 0.5$ m	Average height difference $\overline{\delta h}$ (m)	68% Interval (m)	95% Interval (m)
1, 2	1153	0.03	0.16	0.51
2, 3	1626	− 0.07	0.13	0.27
3, 4	1157	0.07	0.17	0.45
4, 5	1716	− 0.11	0.13	0.40
7, 9	894	− 0.02	0.08	0.30
9, 10	618	− 0.04	0.16	0.35
10, 11	1378	0.06	0.13	0.34
11, 12	309	− 0.33	2.36	6.08
13, 14	1468	0.07	0.16	0.34
14, 15	1386	0.00	0.13	0.31

Rough terrain, altitude flown: 2000 m, sample surface size: 1250 m^2.

The overlap of the flight lines 11 and 12 is way off. Here, visual inspection again confirmed the double and undersampled surface situation and when offending surfaces were removed, the relative accuracy estimators were similar to the ones from the other overlapping strips. The average height difference of sample surfaces for lines 4 and 5 is larger than for the other pairs, though the 68% c.l. and 95% c.l. intervals are consistent. A closer look at the navigation data revealed that during the flight line 5, the GPS/INS integrated solution for INS position and attitude provided by the Honeywell inertial navigation system (INS) H-764 G was of lower quality than usual (H-764 G gives a figure of merit for this solution). The rest of the overlaps are of good quality. We conclude that the performance of the system was consistent throughout the mission with the exception of INS problem during the flight line 5.

Let us now cross compare results from all three case studies discussed here. The studies were based on data collected with different but similarly built lidar units. Terrain type and height at which the data have been collected differ as well. Despite these differences, the overall relative accuracy results (computed using sample surfaces of the same flatness and size) turned out to be reasonably close. Table 5 provides the summary.

The last column—68% interval for the average height differences—is the measure of accuracy level. Interestingly enough, the data from a higher flown mission (case study 3) is slightly better than the data from the case study 1 flown lower. Without more data, we can only speculate whether this is to be attributed to a rougher terrain or to a better performance of one unit over the other. As one would expect, the mission flown over a flat area at low altitude resulted in better overall relative accuracy.

Table 5
Cross study relative accuracy comparison

Case study	Flight height above ground (m)	Terrain	68% Interval (m)
1	913	Very rough	0.17
2	1000	Flat	0.11
3	2000	Rough	0.13

5. Further results and plans for future

While the paper was being reviewed, we have discovered yet another striking example of the power of the overlap analysis. It has been noticed that data collected during a sharp aircraft maneuver had degraded overlap accuracy. This is not surprising since the attitude information is acquired at a fixed rate (33 Hz in our case) and then interpolated to the time of a laser shot. In general, the higher the rate of attitude change, the larger is the error of interpolation. Quantitatively, however, the degradation of the overlap accuracy was far worse than one could have expected from an interpolation error alone. Naturally, the attitude data interpolation routines were carefully examined but neither the linear nor the various spline interpolations were the cause of the observed discrepancies. During these exercises however, we noticed that if a time shift were introduced into the attitude data, it would bring the quality of the data in line with the expectations (previous results). Overlap analysis helped to quickly establish an optimum value for this shift. Later Honeywell (Willis, 2001) confirmed that a particular version of the firmware that we had been using at that time, indeed had an error resulting in the time shift that we had calculated.

In another development of the overlap analysis, we found a very compact and informative presentation of the overlap quality for an entire lidar mission. It is a scatter-plot that shows the average height differences of all sampled surfaces of a given size and flatness. From this plot one can read both the height offset between the overlapping flight lines and the noise level around it. An example of such a plot is shown in Fig. 2.

The plot shows that on average the ground was seen 4 cm lower during the flight line 3 as compared to the line 2 with the RMS error of 1.5 cm. The lines 3, 4 and 5 on the other hand had very consistent height measurements but the noise level was twice as high. More detailed information can be obtained by sorting the sample surfaces by time, scan angle, the rate of attitude change, etc. In view of time limitations, these aspects will be not treated here.

Currently, we are testing and transferring into the production mode an overlap-based calibration procedure. The procedure takes data from the flight lines overlapping at different ranges of scan angle and

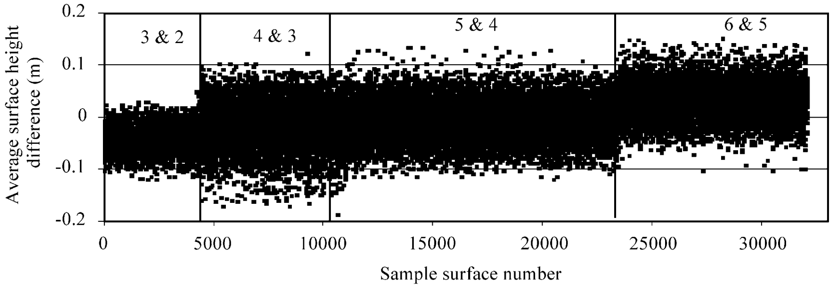

Fig. 2. Scatter plot of the average height differences of the sampled surfaces for four pairs of overlapping flight lines. Vertical lines separate the overlapping pairs of the flight lines. The flight line numbers are shown at the top of the plot (3 through 6). Overlapping flight lines were flown in opposite direction at an altitude of 1000 m. Sample surface size is 360 m^2 and flatness is -0.24 m.

adjusts the calibration parameters to optimize the overall overlap accuracy. As has been shown in Section 2.1, a calibration procedure without any ground control information cannot establish the value of the systematic vertical error. Addition of a few ground control points eliminates this problem. We have found that with the conventional calibration model (attitude and range calibration only) the overlap-based procedure gives at least as good results as the calibration techniques reported in the literature (Krabill et al., 2000). This, of course, is not surprising at all: it is natural to expect higher accuracy when one matches hundreds or thousands of surfaces, as in the overlap-based procedure, compared to only a few surveyed points or flat surfaces.

The overlap-based calibration proves itself by far superior to the calibration based on ground control information only, when one wants also to account for the distortion of the scan geometry caused by misalignment in the optics (e.g. mirror misalignment with respect to the laser beam). Unless compensated for, this distortion is the major contributor to the height "steps" between overlapping flight lines. Accurate determination of the misalignment parameters from the collected data (and possibly other errors, such as errors in timing of various acquired data streams) is a non-trivial inverse problem that rarely can be solved based on the ground control information alone. Vosselman and Maas (2001) have proposed a model to smooth out these steps by means of nine parameter model that includes three translations, three rotations and angular drifts. The major drawback of their model is that it does not use the actual scanning geometry and is therefore non-physical. The author agrees with Filin (2001) that the correction of laser points by means of a linear transformation focuses on the effect of the systematic errors but not on their causes. According to Filin, an analysis of the similarity transformation reveals that not all error effects can be modeled, and thus removed, by this transformation. Our calibration procedure uses the true scanning equations (these are specific to a lidar system) and compensates for the actual physical errors: attitude and range measurement errors and optics misalignment. We discuss our calibration procedure and some of the results in (Latypov and Zosse, in press).

6. Conclusion

The main result of this paper is a new method for estimating relative lidar accuracy by comparing surfaces of overlapping lidar datasets. The method is based on a generalized relative accuracy estimate, the estimate being not just a single number but a function of two model parameters-size and flatness of surfaces used in its estimation. The point-wise relative accuracy estimate is obtained as a limiting value of this function when surface size goes to zero.

The advantages of the new method are the following:

(1) A straightforward, fully automated software implementation is possible.

(2) The relative accuracy computations are performed on the entire overlap area rather than on the selected, very small patches of the data as it is often done. Thus, significant amount of accuracy information can be obtained even if there are only very few (or none at all) ground control points.

(3) Instead of a single number, the relative accuracy can be provided as a function of various operating parameters, such as scan angle and rate of airplane attitude change. It is computed locally and maps of accuracy over a project area can be created.

A very nice example of relative accuracy information the method is capable of providing is the scatter plot described in Section 5. The plot quantifies the systematic height offset between the flight lines and the standard deviation of the height differences from this offset. This gives a very good idea on the quality of the calibration, GPS quality and the noise level of the hardware. A drawback of the method is that it calculates not an absolute but a relative error for the two overlapping flight lines. However, any constant offset common to the both datasets can be easily found with the help of a few ground control points.

Acknowledgements

The author would like to thank all TerraPoint personnel who one way or the other made this paper possible. The author is grateful to Prof. Amir Dembo of Stanford University for valuable comments on the problem and Dr. Yuri Rostovtsev of Texas A&M University for proof-reading the manuscript. Special thanks go to Dr. Manos Baltsavias of Institute of Geodesy and Photogrammetry, Zurich, Switzerland, for constructive criticism and numerous suggestions that made this paper better.

References

Airborne 1, 2001. Lidar accuracy. Airborne 1 Briefing Note BN#01 01.15.01.

Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas. ISPRS Journal of Photogrammetry and Remote Sensing 54 (2–3), 199–214.

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

Carabajal, C.C., Harding, D.J., 2001. Evaluation of geoscience laser altimeter system (GLAS) waveforms for vegetated landscapes using airborne laser altimeter scanning data. International Archives of Photogrammetry and Remote Sensing 33 (3/W4), 125–128.

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.

Filin, S., 2001. Recovery of systematic biases in laser altimeters using natural surfaces. International Archives of Photogrammetry and Remote Sensing 34 (3/W4), 85–91.

Fritsch, D., Kilian, J., 1994. Filtering and calibration of laser scanner measurements. International Archives of Photogrammetry and Remote Sensing 30 (3), 227–234.

Gutierrez, R., Gibeaut, J.C., Mahoney, M.P., Crawford, M.M., Smith, S., Gutelius, W., Carswell, D., Macpherson, E., 1998. Airborne laser swath mapping of Galveston island and Bolivar peninsula, Texas. Proc. of Fifth Int. Conf. on Remote Sensing for Marine and Coastal Environments, San Diego, 5–7 October, USA, pp. 236–243.

Kilian, J., Haala, N., Englich, M., 1996. Capture and evaluation of airborne laser scanner data. International Archives of Photogrammetry and Remote Sensing 31 (B3), 383–388.

Krabill, W.B., Thomas, R.H., Martin, C.F., Swift, R.N., Frederick, E.B., 1995. Accuracy of airborne laser altimetry over the Greenlad ice sheet. Geophysical Research Letters 16 (7), 1211–1222.

Krabill, W.B., Wright, C.W., Swift, R.N., Frederick, E.B., Manizade, S.S., Yungel, J.K., Martin, C.F., Sonntag, J.G., Duffy, M., Hulslander, W., Brock, J.C., 2000. Airborne laser mapping of Assateague National seashore beach. Photogrammetric Engineering and Remote Sensing 66 (1), 65–71.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 53 (4), 193–203.

Latypov, D., Zosse, E., 2002. Lidar data quality control and system calibration using overlapping flight lines in commercial environment. Proceedings of ACSM-ASPRS Annual Conference, 19–26, April, Washington, DC, in press.

Maas, H.-G., 2000. Least-squares matching with airborne laser scanning data in TIN structure. International Archives of Photogrammetry and Remote Sensing 33 (B3/1), 548–555.

Shrestha, R.L., Carter, W.E., Lee, M., Finer, P., Sartori, M., 1999. Airborne laser swath mapping: accuracy assessment for surveying and mapping applications. Journal of American Congress on Surveying and Mapping 59 (2), 83–94.

Vosselman, G., Maas, H.-G., 2001. Adjustment and filtering of raw laser altimetry data. Proc. OEEPE Workshop on Airborne Laser scanning and Interferometric SAR for Detailed Digital Elevation Models, 1–3 March, Stockholm, paper 5 (11 p.). Available at http://www.tu-dresden.de/fghgipf/forschung/material/publ2001/OEEPE_Stockholm.pdf (accessed 15 March 2002).

Willis, J., 2001. Honeywell. Personal communication.