LIDAR PRINCIPLES AND APPLICATIONS*

Robert Burtch
Surveying Engineering Department
Ferris State University
Big Rapids, MI 49307

Abstract

LIDAR is becoming an important mapping tool that is finding a growing audience in GIS. This presentation will discuss the principles of LIDAR and the capabilities that this new technology offers the mapping professional. LIDAR systems will be described and accuracy issues will be discussed. Examples of LIDAR projects will be presented to see how terrain detail can be extracted from LIDAR measurements. The goal of this presentation is to make the GIS professional aware of the role that LIDAR had in mapping.

Introduction

LIDAR is a relatively new technological tool that can be used to accurately georeference terrain features. LIDAR is an acronym for <u>LI</u>ght <u>D</u>etection <u>A</u>nd <u>R</u>anging and in some literature it is referred to as laser altimetry. While the system is new, the underlying technologies that comprise a system have been around for a number of years. A LIDAR system (see figure 1) is composed of a laser scanning system, global positioning system (GPS), and an inertial measuring unit (IMU).

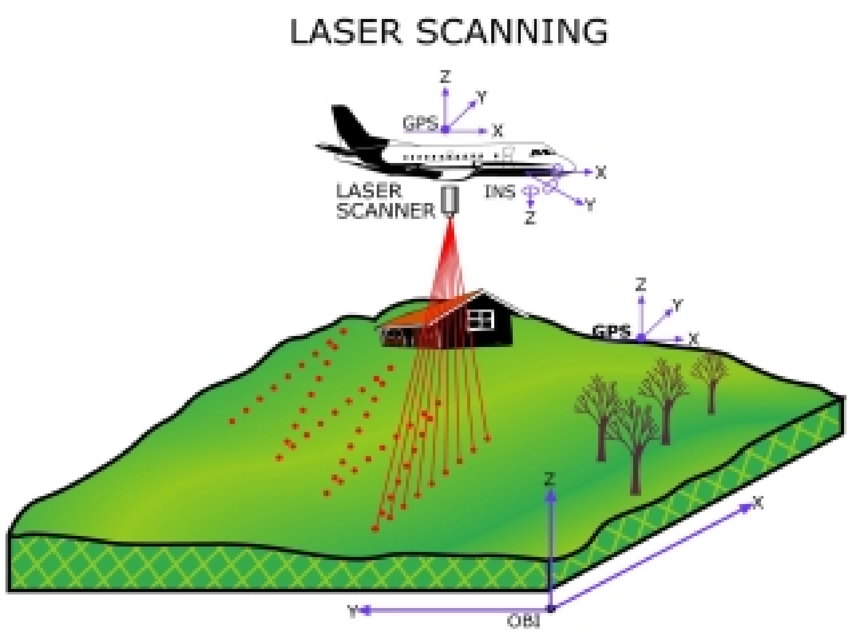

Figure 1. LIDAR system.

* Paper presented at the 2002 IMAGIN Conference, Traverse City, MI.

Lasers (Light Amplification by Stimulated Emission of Radiation) have been used for years in the geospatial sciences. The laser basically consists of an emitting diode that produces a light source at a very specific frequency. The signal is sent towards the earth where it is reflected off a feature back towards the aircraft. A receiver then captures the return pulse. Using accurate timing, the distance to the feature can be measured. By knowing the speed of light and the time the signal takes to travel from the aircraft to the object can back to the aircraft, the distance can be computed using the basic relationship

$$D = \frac{vt}{2}$$

where D is the distance from the aircraft to the object (this is one-half the total distance that the laser signal actually traveled), v is the velocity or speed of light, and t is the time between emitting and receiving a particular signal.

There are basically two kinds of lasers used in LIDAR [Ackermann, 1999]. These are the pulse lasers and the continuous wave (cw) lasers. Flood [2001] refers to these as small footprint, time-of-flight laser altimetry and large footprint waveform digitizing.

The pulse laser emits a narrow laser pulse in the near infrared region of the electromagnetic spectrum. Each discrete pulse is then reflected off a surface on the earth and returned to the receiver. This signal yields a small footprint on the surface of the earth. One of the problems with this method of LIDAR data collection is that acceptable results may be somewhat difficult to achieve in dense and complex canopies [Flood, 2001]. While the signal may penetrate to the ground through holes in the canopy, many returns have to be filtered for correct classification of the ground surface.

The cw laser emits a continuous signal stream where the receiver captures the full return wave. Distances are determined from phase measurements. The return signal covers a wider footprint and contains the entire structure of the return signal.

There are two distinct types of LIDAR systems based on the environment in which they are being used. A topographic system, which is the topic of this paper, is used over land and operates in the infrared portion of the electromagnetic spectrum. Over water, the infrared signal is partially absorbed by the water resulting in almost no return signal. A bathymetric system is used over water and it utilizes the blue-green portion of the electromagnetic spectrum, thereby allowing penetration and a return signal though the water.

While the speed of light is well known in a vacuum, one would expect that it would vary in the actual atmosphere. Thus, the raw distance, or sometimes called the range, is influenced by the variation in the actual speed of light. This variation can be modeled and corrected for in the processing of the raw laser signal.

The laser scanner is mounted in an aircraft just like an aerial camera. It can emit upwards to 50,000 pulses per second[1]. The laser scan data is collected using a scanning mirror that rotates transverse to the direction of flight. The scan angle is generally less than 20^{o} in both directions from the nadir line, although some system may scan up to 30^{o}. The laser scan signal forms a footprint on the ground, which is referred to as the instantaneous field of view (IFOV). If the aircraft is completely level and if the laser scan is in the vertical position, then the IFOV will be a circle. As the laser scan signal moves off the vertical, the IFOV will become elongated, forming an ellipse, along the scan direction thereby enlarging the footprint (figure 2).

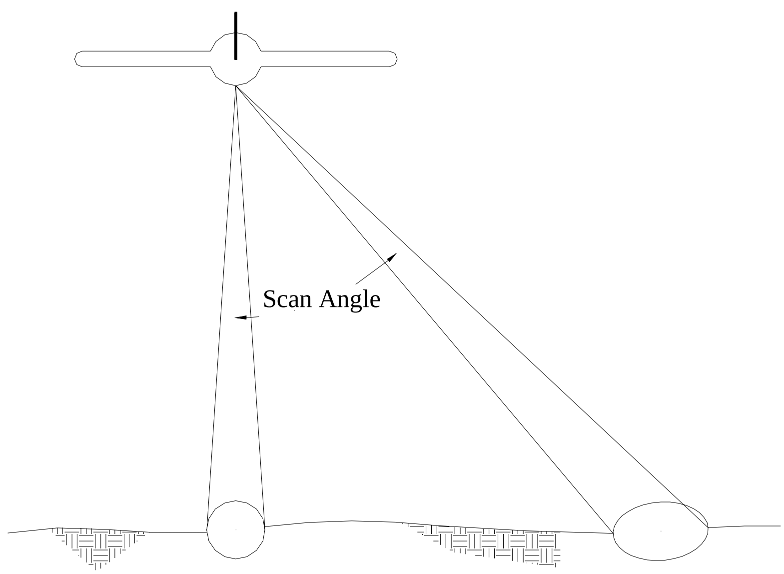

Figure 2. Example of IFOV.

One of the disadvantages of the moving mirror is that the rate of movement is not a constant. As the mirror nears the end of the scan it slows down, then stops, reverses direction, and finally speeds up. This type of movement, besides adding strain to the mechanics of the system, affects the positional accuracy of the system.

One way around this problem is to use a rotating prism, which moves only in one direction thereby fixing the speed of the prism movement at a constant rate. The disadvantage of the rotating prism is timing when the data are to be collected since there are no fixed stop positions to indicate the extent of the swath. Additionally, since all the data are collected in one direction there might be a bias in the measurement that would not be distinguishable unless there was a field check.

Recalling the principles presented in the last lesson, we should recognize that the LIDAR signal is not a point but rather is an area. One of the advantages of a laser signal is that the beam is very narrow, but it does get larger as it farther away from the source.

[1] This figure of 50,000 pulses is based on expected sampling rates from systems in development. Current rates found in the commercial sector are well below this value.

Moreover, it also becomes distorted, taking on an ellipsoidal shape, as it travels along the scan. Thus, the density of a scan will relate to the size of the signal on the ground. This is typically 2 to 10 feet. The collected data will consist of a "herringbone" pattern of spot elements (figure 3). The scan rate must be sufficiently fast to prevent any unwanted gaps in the data. This allows for a good uniform distribution of data over the project site.

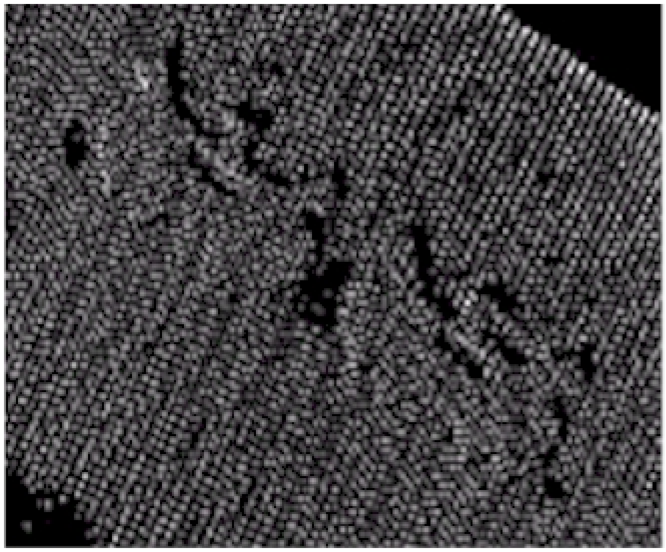

Figure 3. Sample LIDAR scan [from http://www.airborne1.com/Pages/technology.htm accessed 5/31/2001].

Many vendors feel that the mirror-type laser scanning system yields more accurate results. To compensate for the slowing down effect, software is used to correct for this effect and to disregard data at the end of the swath. For example, most data providers may advertise that they have a swath width of 3,000' but use only the middle 2,500'.

Different lasers emit signals at different rates, which are designated as kHz rates. This gives the number of individual pulses of light emitted by the system per second, in thousands of pulses. For example, a 10 kHz system will emit 10,000 individual laser pulses in one second. A 50 kHz system will generate 50,000 pulses during this same one-second interval.

GPS is a very common technology used for a myriad of purposes. It consists of a series of satellites that emit signals in the radio frequencies that are captured by a receiver. Conceptually, the distance from the satellite to the receiver can be calculated using the same basic relationships used to describe how the laser measures the range. The satellite position is considered as a known variable and the range (called pseudorange) is measured by the receiver. With one satellite, one can basically determine a location on a sphere whose radius is equal to the pseudorange. If we use three satellites, all of whom their position is known, then by a process of resection, a unique location can be determined for the receiver. In reality, a minimum of four satellites are required because of another unknown variable – clock bias[2].

GPS is used to precisely locate the position of the scanner during the measurement. The GPS system will be comprised of at least two receivers: one located on a know point on the ground and the second located on the aircraft. It is desirable to position the on-board receiver on the fuselage directly above the laser scanner. The relative position of the GPS receiver to the scanner will be accurately measured so that the offsets can be transferred to the scanner position. The two-receiver configuration is used to apply the differential correction to the roving receiver on-board the aircraft. It is also very important in that it will help resolve any datum uncertainty that might exist in the GPS measurement.

The last component of a LIDAR system is the IMU. As the aircraft moves along its flight line, it is susceptible to small angular variations due to pitch, yaw, and roll. The IMU measures these angular changes thereby allowing the analyst to determine the orientation of the scanner. Without this information, there is no way to fix the location of the laser scan footprint on the ground. For example, if the scan angle was 10^o to the right of the flight direction when the laser signal was emitted, the flying height was 1,000 m, and the return distance was 1015.43m (one-way distance, see figure 4), then the location of the center of the IFOV is

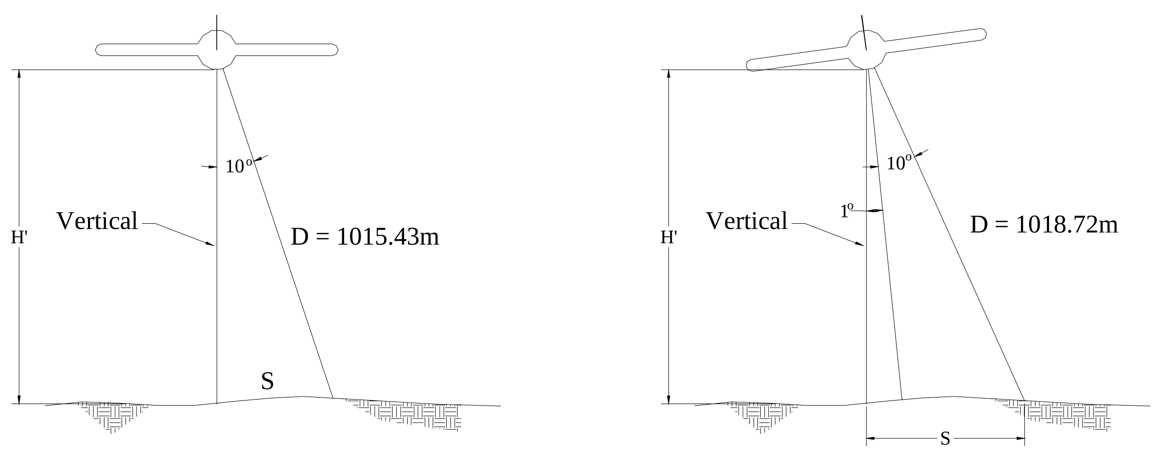

Figure 4. Example of the effects of tilt on a LIDAR signal.

[2] This description of the GPS satellite system is a gross simplification of the real system. It is used just to illustrate how one could locate a receiver in space. Like the laser, GPS is one of the triad of enabling technologies that make LIDAR work.

$$S = D \sin 10^{o} = 176.33\,\mathrm{m}$$

where S is the ground distance from the vertical to the center of the IFOV.

Lets further assume that the aircraft was tilted 1^{o} to the right during the same time and that the return distance was found to be 1018.72 m. Then, using the same 10^{o} angle, the distance from the flight line to the center of the IFOV is 176.90 m. But, if the tilt of the aircraft is taken into account, then the total angle from the vertical is 11^{o} and the distance is now

$$S = D \sin 11^{o} = 194.38\,\mathrm{m}$$

This is a difference of over 17 m. Therefore, it is very important that the IMU accurately measure the angular change that occurs in all three directions and make the necessary corrections to the georeferencing algorithm.

Each component is sampling data at a set interval. Hence, timing becomes an important issue. For example, if the GPS is sampling at a one-second interval and the laser is sending out 20,000 pulses per second, and the aircraft is traveling at a predefined velocity, the location of the aircraft, and thus the location of the laser scanner, needs to be interpolated between the GPS sample times. This same situation exists for the IMU. Therefore, the timing system is frequently referred to as the fourth component of the LIDAR system.

Processing

Post-processing of LIDAR data is required for two main reasons [Maune, 2001]. First, it is necessary that the data be related to the particular frame of reference required by the user. This is accomplished using GPS. It is recommended that two ground receivers be set up over known points in the vicinity of the project. Using differential GPS, the location of the receiver can then be very accurately determined and since it is based on vectors from two bases, the results can be easily checked for accuracy. Since the GPS measures in the World Geodetic System 1984 (WGS 84), a transformation of the GPS coordinates to the user coordinate framework is required. With the GPS position known, georeferencing of the return pulses can be undertaken.

The second part of post-processing is the elimination of the irrelevant data collected during the survey. Very often it is the “bald” earth that is required by the user, generally for DEM purposes. Yet, the laser scanner does not know if the last return is the ground, the side of a building, the top of a building, or other type of signal reflected from the surface of the earth. Thus, sophisticated algorithms are used to separate this data and throw out measurements that are clearly not ground features. Much of this can be done automatically. This may remove 90% of the irrelevant data, but still artifacts exist. These are removed manually and this may represent 90% of the time required for post-processing.

Raw LIDAR data are post-processed after the initial aerial flight is completed. The slant distance, as described above, is calculated for each returned signal. This data is then corrected for atmospheric effects. The roll, pitch and yaw are determined from the IMU and these angles are then applied to the slant distances to correct for the orientation of the scanner during data collection. The GPS data are processed separately and are then imported into the LIDAR processing system. Using the position of the sensor and the swath angle during the individual scan, the elevation of the ground point can be easily computed. For example, look at the geometry in figure 5.

Lets assume that the laser signal was sent out at a 10^o angle from the nadir along the swath width ($\alpha = 10^o$). Further, assume that the sensor orientation is perfect (no pitch, roll, or yaw effects) and that the distance measured by the laser scanner was found to be 1,387.50 m. Then, using simple trigonometry, the vertical distance from the sensor to the ground at the elevation of A (V_A) is

$$
\begin{aligned}
V_A &= D_A \cos\alpha_A \\
&= 1{,}387.50 \text{ m} \cos 10^o \\
&= 1{,}366.42 \text{ m}
\end{aligned}
$$

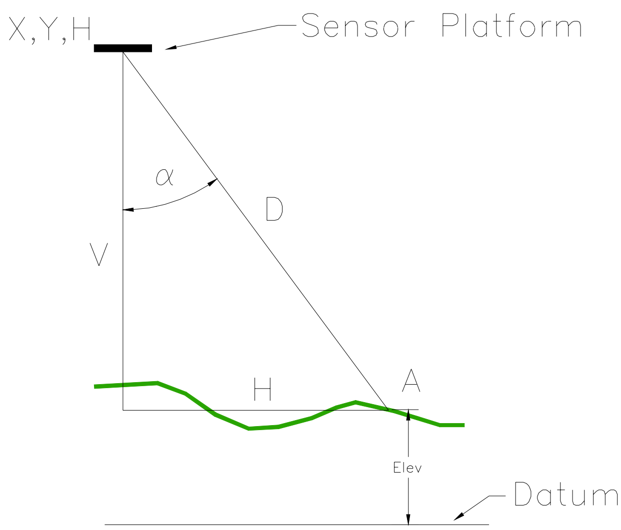

Figure 5. Geometry of the laser scan pulse.

If the GPS on board the aircraft determined the location of the sensor at that instant the signal was sent with state plane coordinates and orthometric height as:

$$
\begin{aligned}
X_{Sensor} &= 1{,}268{,}471.13 \text{ m} \\
Y_{Sensor} &= 588{,}614.47 \text{ m}
\end{aligned}
$$

$$Z_{Sensor} = \quad 1{,}806.59 \text{ m}$$

then the ground elevation of point A is

$$\begin{aligned} Elev_A &= H_{Sensor} - V_A \\ &= 1{,}806.59 \text{ m} - 1{,}366.42 \text{ m} \\ &= 440.17 \text{ m} \end{aligned}$$

In a similar fashion, the X and Y state plane coordinates can also be determined. For example, the horizontal distance, H_A, from the vertical line to the ground point can be computed using basic trigonometry.

$$H_A = D_A \sin \alpha_A$$

which for our example becomes,

$$\begin{aligned} H_A &= 1{,}387.50 \text{ m} \sin 10^{\circ} \\ &= 240.94 \text{ m} \end{aligned}$$

If we assume that the aircraft is flying due north (along the Y-axis) and the scan angle is to the right (to the east of the vertical line), then the Y-coordinate would remain the same and the X-coordinate would become

$$\begin{aligned} X_A &= X_{Sensor} + H_A \\ &= 1{,}268{,}471.13 \text{ m} + 240.94 \text{ m} \\ &= 1{,}268{,}712.07 \text{ m} \end{aligned}$$

The X, Y, H coordinates of ground point A are then 1,268,471.13 m, 588,614.47 m, and 240.94 m respectively. Hence, each return has been georeferenced. While conceptually the georeferencing is very easy, the realization of the geometry that existed at the instant the laser scan was taken is a little more complex. But, mathematically, the calculations are easily handled within the processing system.

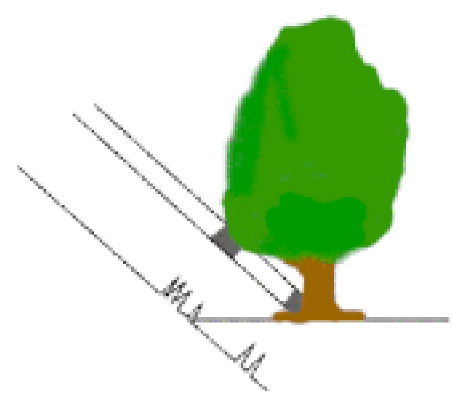

Figure 6. Laser signal hitting multiple objects during its travel.

When a laser signal is sent to the earth it can easily hit more than one object. For example, figure 6 shows a laser pulse heading towards the ground. A part of the signal first encounters a part of the foliage while the rest of the signal hits the ground. Depending on how the system is set up, the sensor can, as an example, collect both of these data pulses simultaneously. This is commonly called the first pulse or first return (the portion of the signal striking the foliage) and the last pulse or last return (the portion of the signal hitting the ground). In some systems it is possible to collect up to 5 different returns. For topographic mapping purposes, it is generally the last return that is provided to the client.

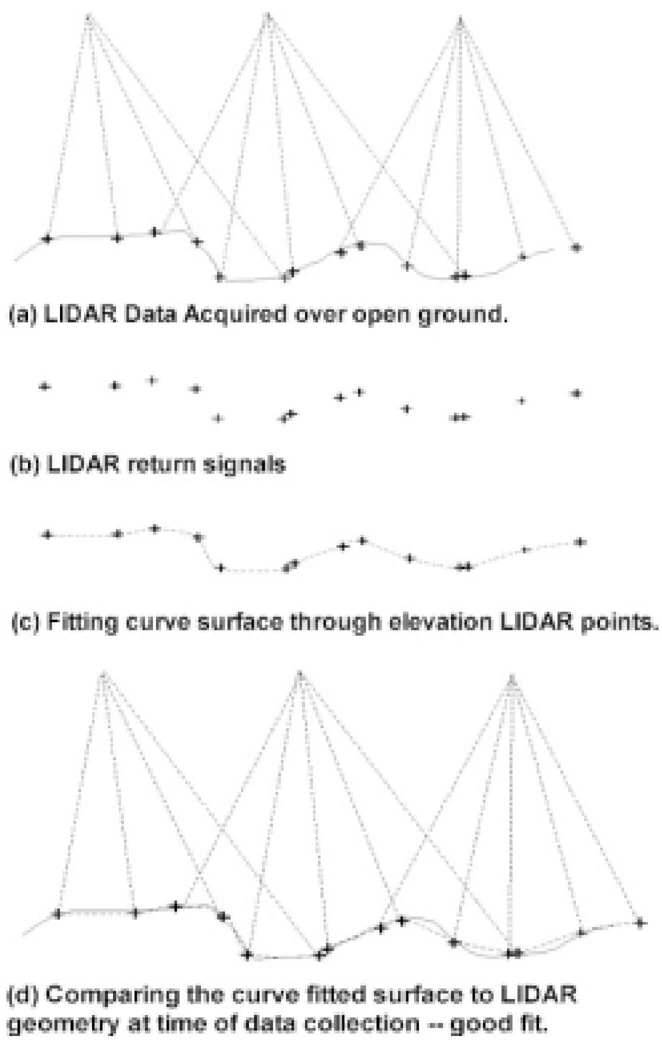

Figure 7. LIDAR data collected over bare ground.[3]

Collecting just the first return can lead to problems. For example, figure 7.a shows the situation where laser data are collected over bare ground. The corresponding elevation data is depicted in figure 7.b and a surface is fit through the data points using a fitting algorithm as shown in figure 7.c. Finally, figure 7.d shows the comparison of this fitted surface to the original geometry. Now assume that there is a forest canopy where some

[3] This excellent depiction and that shown in figure 8, at least from my perspective, comes from http://depts.washington.edu/feprojec/lidar/intro.htm.

of the returns are from the canopy and other obstructions. The elevation data are shown in figure 8.b and the surface is fit to the points as indicted in figure 8.c. The comparison (figure 8.d) shows that the fitted surface does not match the original geometry.

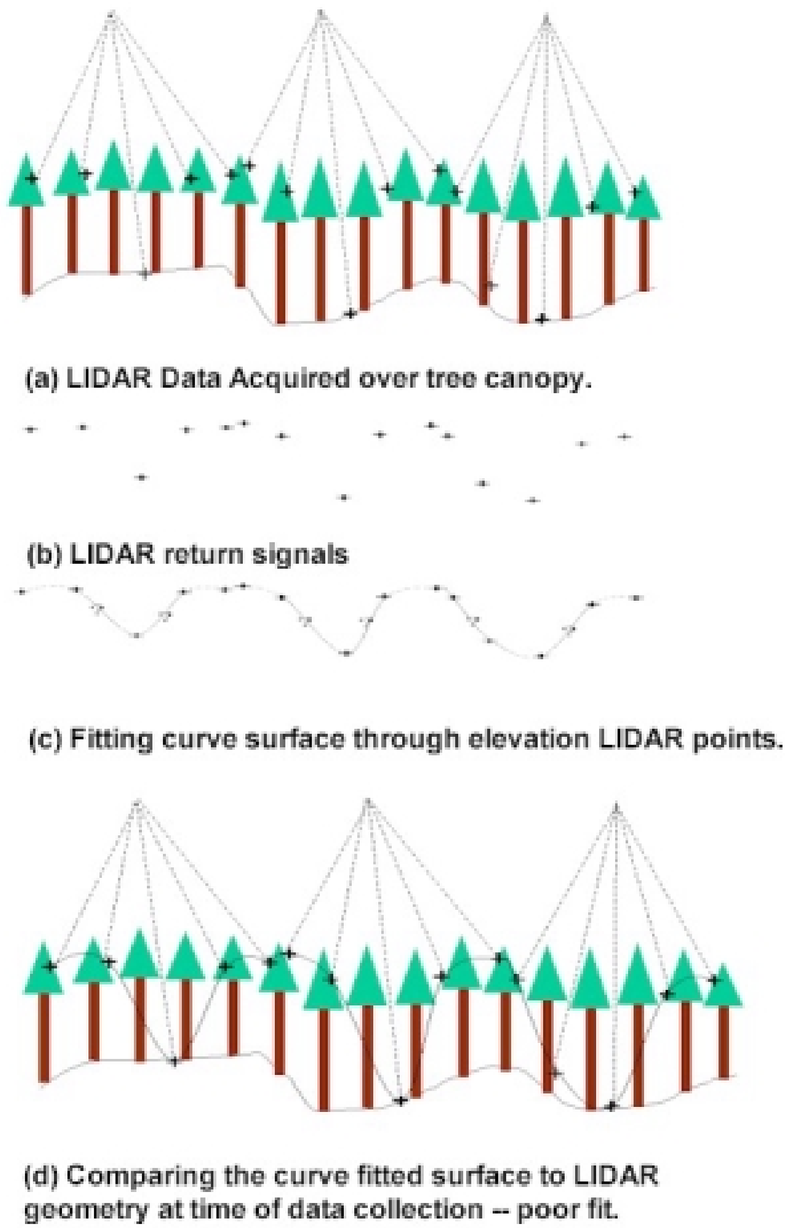

Figure 8. LIDAR data collected over tree canopy.[4]

[4] This excellent depiction and that shown in figure 7, at least from my perspective, comes from http://depts.washington.edu/feprojec/lidar/intro.htm.

For DEM/DTM data products, a “bare earth” ground surface is desired. To arrive at this surface, it is necessary to remove the effects of vegetation and some of the man-made objects from the data.

Advantages and Disadvantages of LIDAR

There are several advantages of LIDAR data. First, it is a very versatile technology that has been used for atmospheric studies, bathymetric surveys, glacial ice investigations, and numerous other applications. It is finding a lot of use in terrain mapping. Here we see that this technology is a very cost effective method of terrain data collection. It offers high precision and high point density data for DTM modeling. Moreover, it has been shown to accelerate the project schedule, upwards to 30% because the DTM data processing can begin almost immediately [Brinkman and O’Neill, 2000]. It is, theoretically, not restricted to daylight nor cloud cover like aerial photography, although if aerial imagery is being collected simultaneously, as it is commonly done, then those limitation will affect the particular project. In coastal zones and forest areas, LIDAR is considered as a superior data collection tool over conventional photogrammetric techniques where it is extremely difficult to locate terrain points in the imagery. LIDAR requires only one opening through a tree canopy to “see” the ground whereas photogrammetry requires that the same ground point be visible from two exposure stations. This would cut down on the amount of area identified as “obscured terrain” on a contour map.

There are several disadvantages as well. While the data collection appears to be cost competitive, the upfront cost of equipment acquisition is very significant, on the order of $1 million. This could be a hard sell since amortization would have to be spread over a very short period since the technology, like that of computers, will probably experience a lot of change over the next two to three years. That is a lot of imagery to collect over a short period of time. While LIDAR is an active system that can be, theoretically, used 24 hours a day, it cannot be used above cloud cover or when fog, smoke, mist, rain, or snow storms are present. Additionally, high winds and turbulence will cause problems with the inertial system.

Photogrammetry is a mature science that is still undergoing technological advances. The products derived from this mapping system are well received and the limitations are understood. While LIDAR appears to be an excellent alternative to photogrammetric mapping, there are several disadvantages to LIDAR when the two technologies are compared. These are [Maume, 2001]:

- There are problems with data collected over water, which leads to suspect delineation of water boundaries using LIDAR by itself.
- LIDAR systems are not capable of determining break lines. Laser scan data are collected in a more or less regular spacing pattern. In other words, it cannot be pointed on a specific feature. For example, a 2 meter-wide ditch may not be

shown on a LIDAR dataset with a spacing of 5 meters. Thus, LIDAR data are often augmented with break line data compiled from photogrammetric methods.

- Being a relatively new technology, standards have not been established that could help guide the user as to the quality of the results. There are a number of efforts underway to alleviate this problem.
- When elevation data are compiled from photogrammetric processes, the operator has a "cartographic license" when selecting points for measurement. Contour lines are generally smoothed to reflect the actual representation of the terrain. A large boulder, as an example, may be a ground surface point captured by LIDAR and the resulting data may depict this as a high point in the terrain. The photogrammetric operator would not use this point in the data collection process as a terrain point for DEM generation.

Accuracies

There are a number of very optimistic claims as to the accuracy of LIDAR data. To fully assess the accuracy one must consider the errors inherent in the three components of the system (laser scan, GPS, and IMU). It is conservatively estimated that the accuracy of LIDAR, as determined from error propagation, is about 15 centimeters in elevation and horizontal position. This can be thought of as typical results from LIDAR surveys. This does assume that the system is properly calibrated and functioning correctly and that the surface terrain conditions are ideal. This latter assumption is almost never correct. As a rule of thumb, horizontal accuracy is often claimed to be $1/2{,}000^{th}$ of the flying height. Vertical accuracies of better than 15 cm are obtainable when the sensor altitude is below 1,200 m and up to 25 cm when the operating altitude is between 1,200 m and 2,500 m [Brinkman and O'Neill, 2000].

There are some rules of thumb that pertain to the accuracy of LIDAR [Brinkman and O'Neill, 2000]:

- The spot spacing is much denser for slower aircraft.
- A more reliable or accurate DTM is available through a denser spot spacing since there is more data collected by the system.
- The highest accuracy heights occur at the nadir and decrease as the swath angle increases.
- A smaller width will yield a denser spot spacing problem.

Conclusion

It is clearly evident that many within the GIS industry are looking at LIDAR as an economical and accurate means of collecting both feature and terrain data. Indeed, this technology is growing. Like any new technological tool, there are times when the technology is misused. Just like GPS has not made conventional terrestrial surveying obsolete, LIDAR will not soon supplant photogrammetric mapping as an economical and accurate method of collecting data about features on the earth. As the technology matures, as new data processing techniques are developed, and as standards are developed, it is safe to say that LIDAR will become an important data collection methodology available to the user community.

References

Ackermann, F., 1999. "Airborne Laser Scanning – Present Status and Future Expectations", ISPRS Journal of Photogrammetry and Remote Sensing, 54:64-67.

Brinkman, R. and C. O'Neill, 2000. "LIDAR and Photogrammetric Mapping", The Military Engineer, No. 605.

Fowler, R., 2001. "The Thorny Problem of LIDAR Specifications", EOM, 10(4): 25-28.

Maune, D., editor, 2001. Digital Elevation Model Technologies and Applications: The DEM Users Manual, American Society of Photogrammetry and Remote Sensing, Bethesda, MD.