Chapter Two

Laser Scanner

Chapter Two

Laser scanner

2-1 Introduction

Within a time frame of only two decades airborne and terrestrial laser scanning have become well established surveying techniques for the acquisition of geospatial information. A wide variety of instruments is commercially available, and a large number of companies operationally use airborne and terrestrial scanners, accompanied by many dedicated data acquisition, processing and visualisation software packages. The high quality 3D point clouds produced by laser scanners are nowadays routinely used for a diverse array of purposes including the production of digital terrain models and 3D city models, forest management and monitoring, corridor mapping, revamping of industrial installations and documentation of cultural heritage.[3]

Airborne and terrestrial laser scanning clearly differ in terms of data capture modes, typical project sizes, scanning mechanisms, and obtainable accuracy and resolution. Yet, they also share many features, especially those resulting from the laser ranging technology. In particular, when it comes to the processing of point clouds, often the same algorithms are applied to airborne as well as terrestrial laser scanning data. In this book we therefore present, as far as possible, an integral treatment of airborne and terrestrial laser scanning technology and processing. [3]

2-2 Laser Scanning Technology

In the last 50 years, many advances in the fields of solid-state electronics, photonics, and computer vision and graphics have made it possible to construct reliable, high resolution and accurate terrestrial and airborne laser scanners. Furthermore, the possibility of processing dense point clouds in an efficient and cost-effective way has opened up a multitude of applications in fields such as topographic, environmental, industrial and cultural heritage 3D data acquisition. [3]

Airborne and terrestrial laser scanners capture and record the geometry and sometimes textural information of visible surfaces of objects and sites. These systems are by their nature non-contact measurement instruments and produce a quantitative 3D digital representation (e.g. point cloud or range map) of a surface in a given field of view with a certain measurement uncertainty. Full-field optical 3D measurement systems in general can be divided into several categories (Figure 2.1). Airborne and terrestrial laser scanners are usually part of what are classified as time-of-flight based optical 3D measurement systems. These systems use a laser source to scan a surface in order to acquire dense range data. Triangulation systems using light sheet or strip projection techniques and passive systems exploiting surface texture (stereo image processing) will be covered only briefly here. [3]

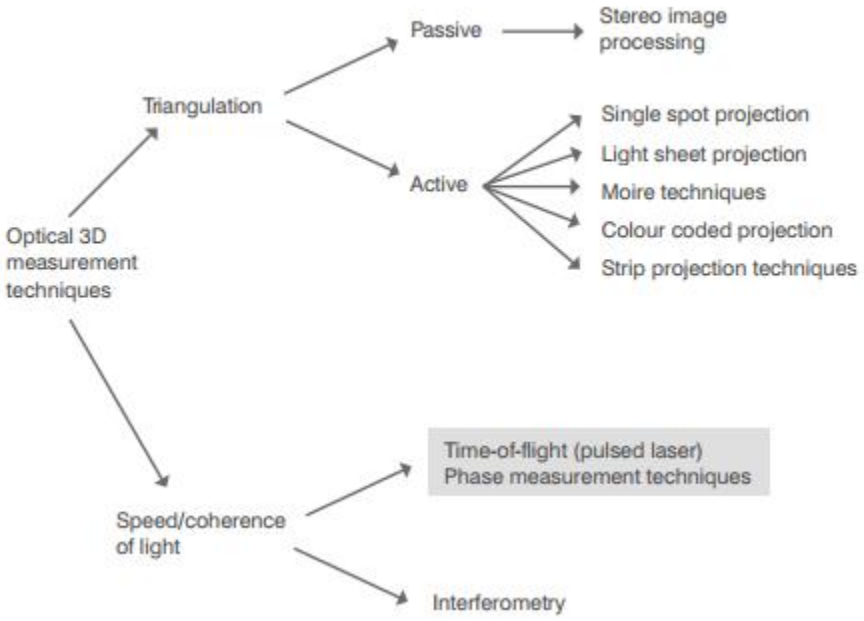
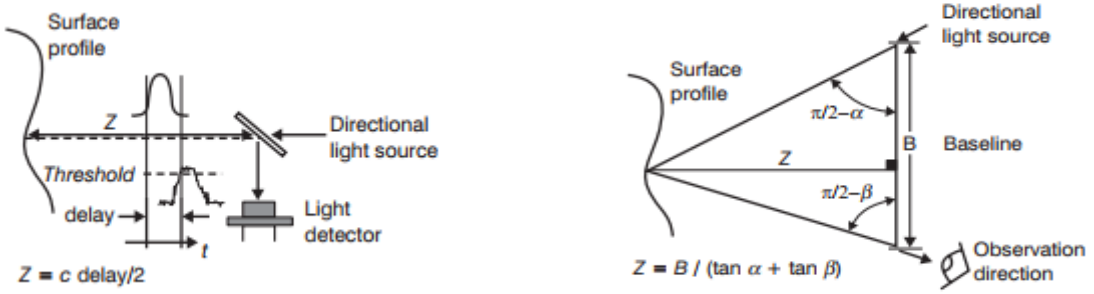


Figure (2-1): Classification of optical 3D measurement systems [2]

2-2-1 Basic measurement principles of laser scanners

There are two basic active methods for optically measuring a 3D surface: light transit time estimation and triangulation. As illustrated in Figure 2.2(a), light waves travel with a known velocity in a given medium. Thus, the measurement of the time delay created by light travelling from a source to a reflective target surface and back to a light detector offers a very convenient method of evaluating distance. Such systems are also known as time-of-flight or lidar (light detection and ranging) systems. Time-of- flight measurement may also be realised indirectly via phase measurement in continuous wave (CW) lasers. Triangulation exploits the cosine law by constructing a triangle using an illumination direction (angle) aimed at a reflective surface and an observation direction (angle) at a known distance (base distance or baseline) from the illumination source . Interferometry (which is not covered here) can be classified separately as a third method or included with time-of-flight methods depending on how the metric used to measure shape is seen. [14]



a) light transit time

b) triangulation

Figure (2-2): Active methods for optically measuring a 3D [14]

2-2-2 Time-of-flight measurement

Early work on time-of-flight ranging systems used radio waves in what became known as radar (radio detection and ranging). The fundamental work can be traced back to experiments conducted by Hertz in the late 1880s. With the advent of lasers in the late 1950s it became possible to image a surface with angular and range resolutions much higher than that obtained with radio waves. The theoretical principles are the same for all systems that use radiated electromagnetic energy for ranging except for their implementations, the differences in their performance and obviously their use. [14]

A fundamental property of a light wave is its propagation velocity. In a given medium, light waves travel with a finite and constant velocity. Thus, the measurement of time delays (also known as time-of-flight) created by light travelling in a medium from a source to a reflective target surface and back to the source (round trip, τ) offers a very convenient way to evaluate the range [14]

$$\rho = \frac{c}{n}\frac{\tau}{2} \qquad (2.1)$$

The current accepted value for the speed of light in a vacuum is c = 299 792 458 m/s. If the light waves travel in air then a correction factor equal to the refractive index, which depends on the air temperature, pressure and humidity, must be applied to c, n ≈ 1.00025. Let us assume c = 3 × 10^8 m/s and n = 1 in the rest of the discussion. More than one pulse echo can be measured due to multiple returns that are caused by the site characteristics, especially when vegetation (the canopy) is scanned. [7]

Airborne systems capture at least the first returned pulse or echo and also the last echo for each emitted pulse. Most airborne systems are capable of capturing four to five separate echoes. Multiple echo measurements are also becoming available for terrestrial time-of-flight scanners.It is important to have a more detailed look at pulse shape and pulse repetition time. The characteristics of a transmitted pulse are the pulse width t_p and the pulse rise time tr (Figure 1.3). A typical pulse width of tp 5 ns corresponds to a length of about 1.5 m at the speed of light, while a pulse rise time tr 1 ns corresponds to a length of 0.3 m. [14]

According to Equation (2.1), the relationship between S (the distance between the laser scanner and the illuminated spot) and time-of-flight τ is given by [7]

$$\tau = n\frac{2\rho}{c}$$

With n= 1, the time-of-flight is 6.7 μs at a distance (the survey height in an airborne system) of 1000 m. When assuming that from a distance S two or more echoes are generated from only one pulse P1 , different ranges can only be discriminated if the echoes E11 and E12 are separated, i.e. do not have an overlap. [7]

$$\tau^{12} \geq \tau^{11} + l_p$$

With the above relation between time-of-flight and distance and the pulse length lp , which is given by $t_p = \frac{n}{c} l_p$, this leads to

$$2\frac{n}{c}\rho^{12} - 2\frac{n}{c}\rho^{11} \geq \frac{n}{c}l_p$$

$$\rho^{12} - \frac{n}{c} \geq \frac{1}{2}l_p \qquad (2.2)$$

Two echoes (for example related to objects heights) can only be discriminated if their distance is larger than half of the pulse length lp . This means that for a pulse width of 5 ns objects can be detected as separate objects if their distance is at least larger than 0.75 m. [14]

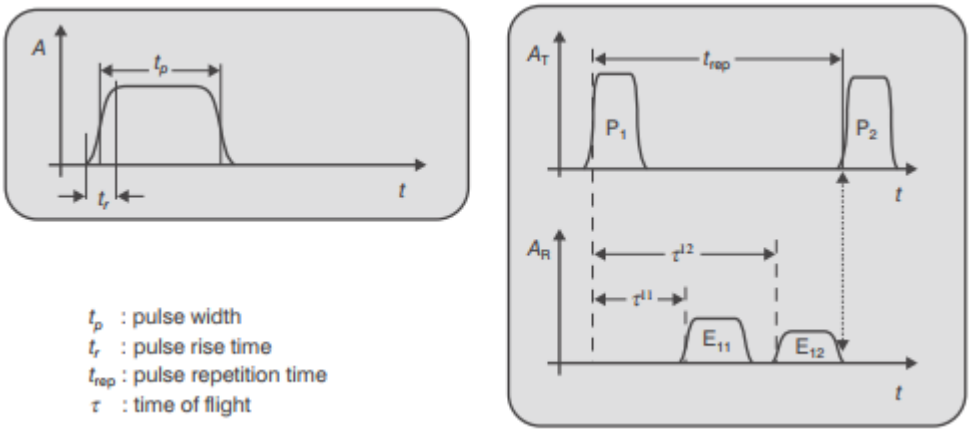


Figure (2-3): Pulse characteristics and measurement principle. [14]

An essential part of the time-of-flight measurement is the detection method for determining the time-of-flight (and thus the range). The detector will generate a time tagged trigger pulse depending on the implemented criterion. Some detection methods take characteristic points of the path of the pulse as the decisive factor. Due to the rapidity of data capture and the ability to obtain point clouds instantaneously, laser scanning has become an essential tool along with image-based documentation methods. Total station surveys, on the other hand, require more time on site and usually do not deliver the same level of surface detail. [14]

Peak detection: The detector generates a trigger pulse at the maximum (amplitude) of the echo. Time-of-flight is the time delay given by the time span from the maximum of the emitted pulse to the maximum of the echo. Correct detection can become problematic if the echo provides more than one peak. [14]

Threshold or leading edge detection: Here the trigger pulse is actuated when the rising edge of the echo exceeds a predefined threshold. The disadvantage of this method is that the time-of-flight strongly depends on the echo's amplitude. [14]

Constant fraction detection: This method produces a trigger pulse at the time an echo reaches a preset fraction (typically 50%) of its maximum amplitude. The advantage of this method is that it is relatively independent of an echo's amplitude. [14]

Each detector has its pros and cons, however constant fraction has proven to be a good compromise. [14]

The range uncertainty for a single pulse buried in additive white noise is approximately given by equation (2.3): [14]

$$\delta_{r-p} = \frac{c}{2}\frac{t_r}{\sqrt{SNR}} \quad (2.3)$$

where tr is the rise time of the laser pulse leading edge and the SNR (signal-to-noise ratio) is the power ratio of signal over noise. Assuming SNR =100, tr=1 ns, and a time interval counter with adequate resolution then the range uncertainty will be about 15 mm. Most commercial time-of-flight based laser scanner systems provide a range uncertainty in the order of 5–10 mm (with a few at 50 mm) as long as a high signal-to-noise ratio is maintained. In the case of uncorrelated 3D samples, averaging N independent pulse measurements will reduce Er-p by a factor proportional to the square root of N. Obviously this technique reduces the data rate by N and is of limited applicability in scanning modes. [14]

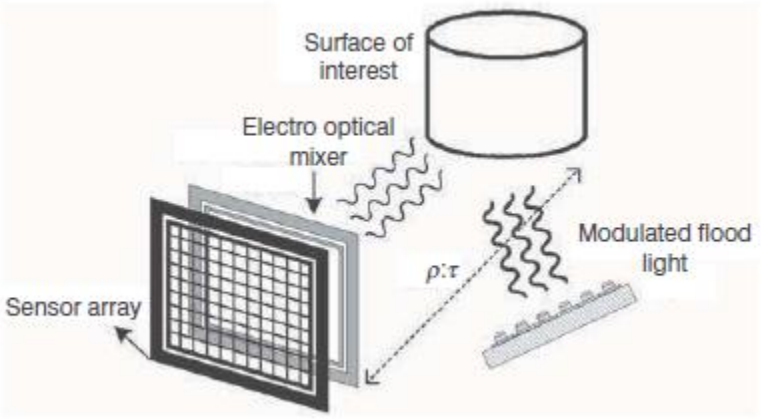
2-3 Airborne laser scanning

In the early 1970s, it was shown that airborne lidar systems were able to measure distances between aircraft and ground targets to a precision of less than 1 m. However, laser altimeter systems did not come into widespread use for precise topographic mapping mainly for two reasons. First, for precise mapping, the vertical position of the aircraft had to be known to a level of accuracy comparable to the measurement capability of the lidar system. Second, the horizontal position of the illuminated spot on the ground (laser footprint) had also to be known. Although the second requirement is less stringent, the means to determine both aspects for larger areas with sufficient quality were not available at that time. Trials were done to determine aircraft altitude by recording pressure data with precise aircraft aneroid barometric altimeters and vertical accelerometers. Horizontal control was a tedious process as it was done post-flight by means of time-tagged photographs and rarely by IMU. [14]

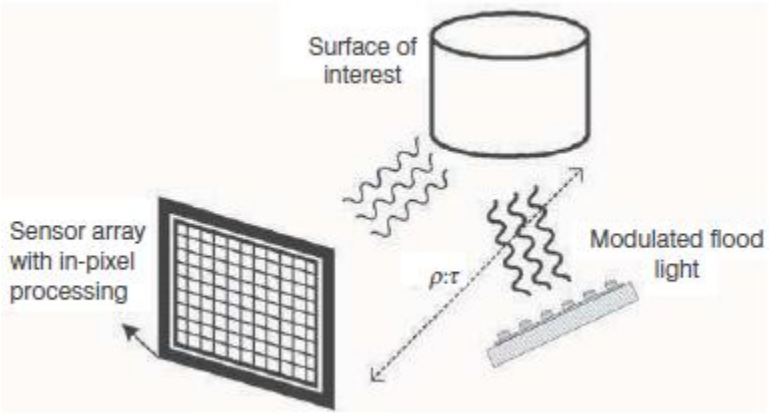
At the end of the 1980s with the availability of GPS, a method was developed allowing precise registration of position and orientation over larger areas. With the introduction of differential GPS (DGPS) the scanner position became known in horizontal and vertical coordinates in the sub-decimetre range. Enhancements in DGPS technology along with Kalman filtering and use of IMUs provided sufficient accuracy from the beginning of the 1990s. The standard accuracy of elevation data became $\mp$10 cm in height, and $\mp$50 cm in position. [14]

Until the end of the 1980s range measurements were done by laser profilers providing laser pulses but no scanning mechanism. In the early 1990s profilers were replaced by scanning devices which generated 5000 to 10 000 laser pulses per second at that time. Nowadays, laser

pulse rates reach 300 kHz, however, depending on the kind of scanning mechanism 100% of the instrument's pulse rate may not really be available on the ground. [14]



a) detection using an external mixer and correlator



b) detection on a specially designed sensor or focal plane array.

Figure (2-4): Flash 3D system with floodlight [14]

Airborne laser scanning is now a common technique for generating high quality 3D presentations (digital elevation models) of the landscape .The following will outline the principle and system components of airborne laser scanning. Characteristics of this measurement technique as well as accuracies and limitations will be discussed. The specifics of calibration, processing of gathered data and product generation will not be discussed as these are covered in other chapters. Several airborne laser scanners are commercially available and in operational service.

2-3-1 Principle of airborne laser scanning

Airborne laser scanning is done either from a fixed wing aircraft or a helicopter. The technique is based on two main components: a laser scanner system which measures the distance to a spot on the ground illuminated by the laser and a GPS/IMU combination to measure exactly the position and orientation of the system. Active systems based on laser scanning are relatively independent of sunlight. They may be operated during the day or at night. This characteristic is a considerable advantage of airborne laser scanning compared to other methods of surveying landscapes. [14]

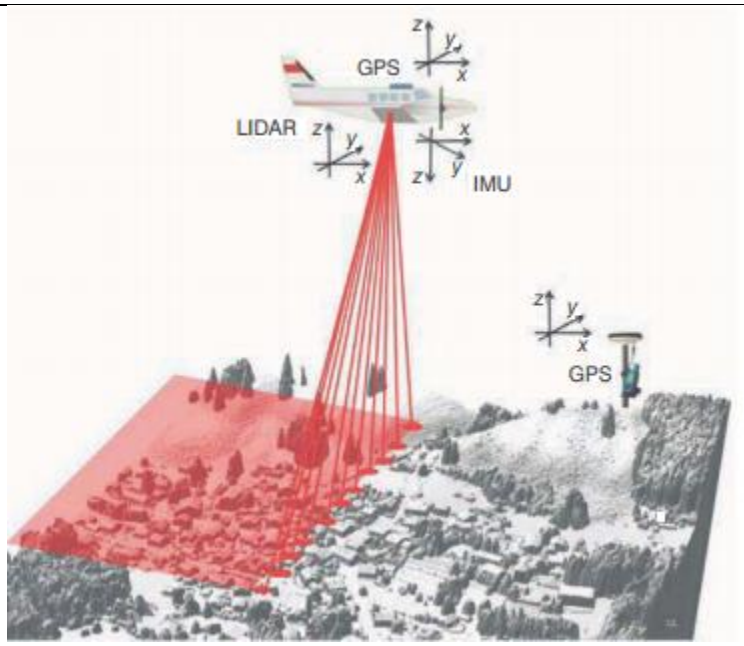


Figure (2-5): Airborne laser scanning principle [14]

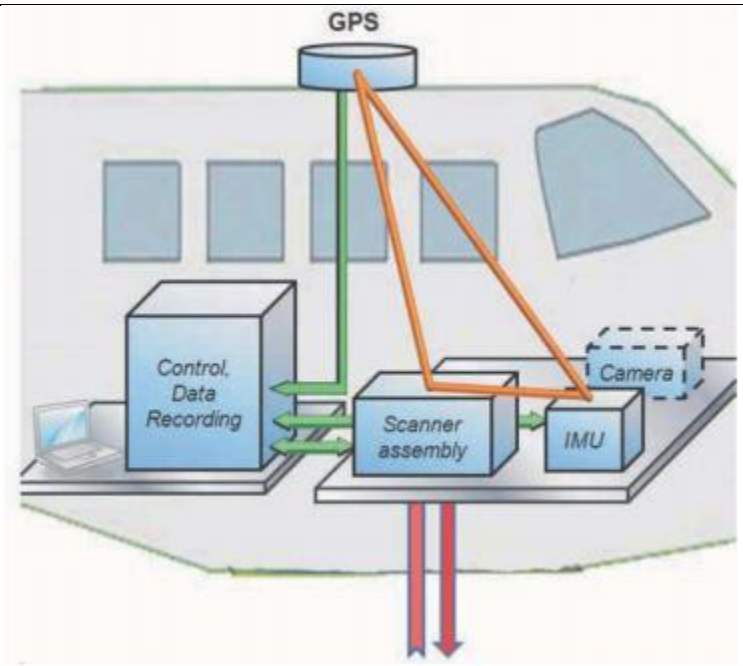


Figure (2-6): On-board components of an airborne laser scanner [14]

the distance to a spot on the ground illuminated by the laser and a GPS/IMU combination to measure exactly the position and orientation of the system. Active systems based on laser scanning are relatively independent of sunlight. They may be operated during the day or at night. This characteristic is a considerable advantage of airborne laser scanning compared to other methods of surveying landscapes The basic components of an airborne laser scanner [4]

Scanner assembly comprising laser, scanning mechanics and optics. The laser system (mostly a pulsed time-of-flight measurement system), mounted over a hole in the aircraft's fuselage, continuously sends laser pulses towards the terrain as the aircraft flies. Depending on aircraft velocity and survey height, current technology allows measurement densities between 0.2 and about 50 points/m2 . Modern scanner assemblies provide a roll compensation to compensate for the roll of the aircraft. This feature helps to avoid gaps in coverage which might occur between adjacent swaths due to roll. Roll compensation allows the overlap between flight lines to be planned to be smaller and therefore gives an economic advantage. [4]

Airborne GPS antenna: the standard is a dual frequency antenna recording GPS signals at a sampling rate of 2 Hz. The antenna is mounted at an exposed position on top of the aircraft, providing an undisturbed view to GPS satellites. [4]

Inertial measurement unit (IMU): the IMU is either fixed directly to the laser scanner or close to it on a stable survey platform. Typically it records acceleration data and rotation rates at a sampling rate of 200 Hz. Acceleration data can be used to support the interpolation of the platform position on the GPS trajectory, while rotation rates are used to determine platform

orientation. The combination of GPS and IMU data allows one to reconstruct the flight path (air trajectory) to an accuracy of better than 10 cm. [4]

Control and data recording unit: this device is responsible for time synchronisation and control of the whole system. It stores ranging and positioning data gathered by the scanner, IMU and GPS. Modern laser scanners, which generate up to 300 000 laser pulses per second, produce about 20 Gbyte of ranging data per hour, while GPS and IMU data only sum up to about 0.1 Gbyte per hour. [4]

Operator laptop: this serves as a means of communications with the control and data recording unit, to set up mission parameters, and to monitor the system's performance during the survey. [4]

Flight management system: this is a means for the pilot to display the preplanned flight lines, which provides support for him in completing the mission. [4]

An airborne laser scanner is completed by a GPS ground station. The ground station serves as a reference station for off-line differential GPS (DGPS) calculation. Differential GPS is crucial for compensating atmospheric effects disturbing precise position determination and for achieving decimetre accuracy. In order to cope with varying atmospheric conditions, the distance between the aircraft and the GPS ground station should not exceed 30 km (although sometimes adequate accuracy is reached for longer distances). Nowadays, several countries operate a network of permanent GPS stations, so there is often no need to set up one's own station. [6]

Airborne laser scanners are often complemented by a medium-sized digital camera system. Image data taken simultaneously with range data may support data interpretation significantly in cases where it is difficult to recognize objects only from range data. Image data will usually offer a better spatial resolution and form a basis for integrated 3D point cloud and image data processing. The camera head and the recording unit are separate. The optimum location for the camera is on the scanner assembly's ground plate, because then the existing IMU registrations can be shared for georeferencing. A separate IMU will be needed if the dimensions of the scanner assembly do not allow accommodation of a camera and a scanner assembly on the same survey platform. [2]

2-3-2 Laser scanner properties

Commercial airborne laser systems for land applications operate at wavelengths between 800 nm and 1550 nm. The spectral width is typically between 0.1 and 0.5 nm. As the reflectivity of an object depends on the wavelength, different laser systems show different pros and cons when scanning the Earth's surface: at wavelengths close to the visible part of the spectrum, the absorption of water is high. Therefore, water surfaces will rarely be seen by laser scanners operating in that part of the spectrum. At about 1550 nm the reflectivity of ice and snow is low, therefore scanners operating at 1550 nm will not be the optimum choice when surveying snow fields. Attention must also be paid to eye safety which means that the human eye must not be damaged when looking into a laser beam. [4]

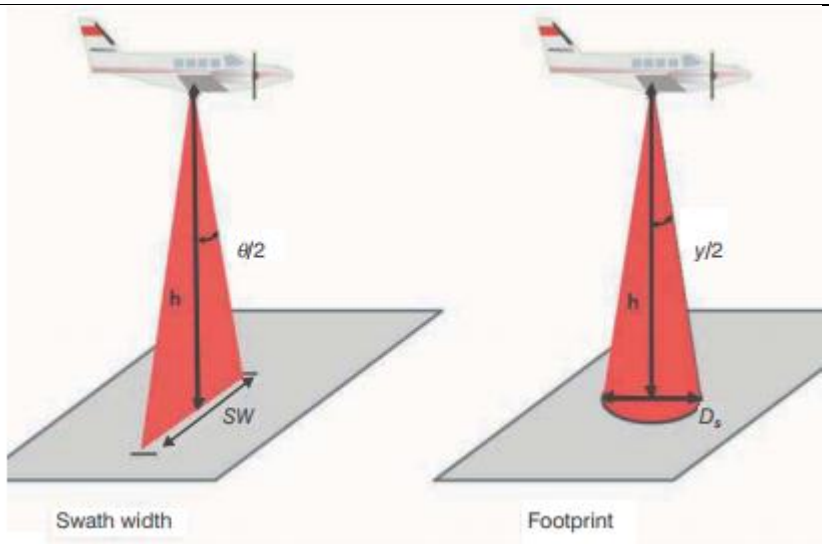


Figure (2-7): Swath and laser footprint (footprint is much smaller in reality) [4]

The swath width sw of a scanner is given by equation (2.4):

$$sw = 2h \tan\frac{\theta}{2} \qquad (2.4)$$

where θ is the full scan angle and h is the height above ground (Figure 1.14). Airborne scanners allow scan angles between about 5° and 75°. For example, the swath width will be 574 m at a height of 1000 m and a scan angle of 32°. [16]

The laser beam widens with the distance from the laser scanner. A relationship, similar to the previous one, describes the diameter Ds of the illuminated footprint on the ground.[16]

$$D_{s=2h \tan\frac{\gamma}{2}} \qquad (2.5)$$

where γ (full angle) is the beam divergence and h the height above the ground assuming the spot shape to be a circle (Figure 2.7). Typically, beam divergences are between 0.1 rad and 1 rad. So, the footprint will be 0.2 m from a survey height of 1000 m and a divergence of 0.2 rad. The laser beam profile does not have sharp edges, but the irradiance falls off gradually away from the centre of the beam. Usually, the diameter Ds is determined at a position at which the irradiance drops to $1/e^2$ times the total irradiance. [2]

2-3-3 Advantages and limitations of airborne laser scanning

As an active system operating with light, a laser scanner needs a clear view to the ground. It cannot penetrate clouds, fog and dense vegetation. The laser beam will easily go through the canopy of deciduous trees, especially in winter time when the leaves are off. However, it will not see the ground below dense conifers and in multistorey rainforest. Total reflection might be the reason if a water surface is seen only close to the nadir direction, and a wet street or a wet roof may also lead to drop-outs. [2]

Despite its limitations, airborne laser scanning has turned out to be a very effective means for producing high quality digital elevation models. Laser scanning technology has some advantages compared to other methods of generating elevation data. Some of its pros are: [6]

High measurement density and high data accuracy: The highest measurement densities (about 30 measurements/m^2) are reached from a helicopter. The standard accuracy of elevation data in the local coordinate system is 0.05– 0.20 m for height and 0.2–1.0 m for position. [2]

Fast data acquisition: For point densities of 1 point/m^2 and higher airborne laser scanning is accepted to be a very fast means of generating accurate elevation models.Fast acquisition is supported by the fact that scanning can be done at any time, day and night. [4]

Canopy penetration: If the canopy is not too dense, part of the laser light beam may penetrate to the ground enabling production of an elevation model of the forest floor. High penetration rates are reached in the case of deciduous trees in winter time, when the leaves are off.

Minimum amount of ground truth data: Terrestrial work is minimized because even for large flight blocks only few ground references are needed. [2]

2-4 Laser Ranging

All laser ranging, profiling, and scanning operations are based on the use of some type of laser-based ranging instrument—usually described as a laser ranger or laser ran gefi nder—that can measure distance to a high degree of accuracy. As will be discussed in more detail later in this chapter, this measurement of distance or range, which is always based on the precise measurement of time, can be carried out using one of the two main methods. [2]

1. The first of these involves the accurate measurement of the TOF of a very short but intense pulse of laser radiation to travel from the laser ranger to the object being measured and to return to the instrument after having been reflected from the object—hence the use of the term "pulse echo" mentioned above. Thus, the laser ranging instrument measures the precise time interval that has elapsed between the pulse being emitted by the laser ranger located at point A and its return after reflection from a ground object located at point B. [2]

$$R = \frac{v.t}{c} \qquad (2.5)$$

where R is the slant distance or range v is the speed of electromagnetic radiation, which is a known value t is the measured time interval From this, the following simple relationship can be derived: [2]

$$\Delta R = \Delta v.\frac{t}{2} + v.\frac{\Delta t}{2} \qquad (2.6)$$

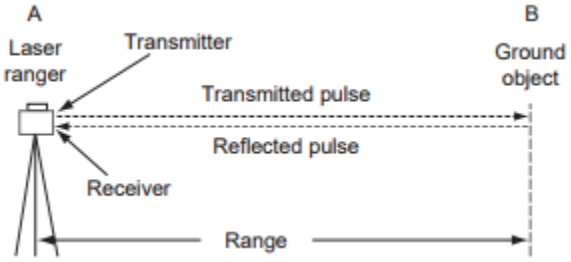


Figure (2-8): Basic operation of a laser rangefinder that is using the timed pulse or TOF method [2]

Where:

ΔR is the range precision

Δv is the velocity precision

Δt is the corresponding precision value of the time measurement

Since the speed of light is very accurately known, in practice, the range precision or resolution is determined by the precision of the time measurement. [2]

2. In the second (alternative) method, the laser transmits a continuous beam of laser radiation instead of a pulse. In this case, the range value is derived by comparing the transmitted and received versions of the sinusoidal wave pattern of this emitted beam and measuring the phase difference between them. Since the wavelength (λ) of the carrier signal of the emitted beam of laser radiation is quite short—typically around 1 μm, and there is no need for such a measuring accuracy in topographic mapping applications, a modulation signal in the form of a measuring wave pattern is superimposed on the carrier signal and its phase difference can be measured more precisely. Thus, the amplitude (or intensity) of the laser radiation will be modulated by a sinusoidal signal, which has a period Tm and wavelength λ m. The measurement of the slant distance R is then carried out through the accurate measurement of the phase difference (or the phase angle, j) between the emitted signal at point A and the signal received at the instrument after its refl ection either from the ground itself or from an object that is present on the ground at point B. This phase measurement is usually carried out using a digital pulse counting technique. This gives the fractional part of the total distance $\Delta\lambda$ (Figure 2.9). By changing the modulation pattern, the integer number. [2]

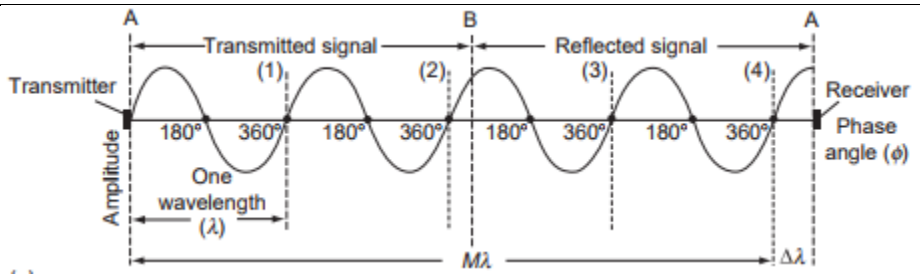


Figure (2-9): Phase comparison is carried out between the transmitted and reflected signals from a CW laser [4]

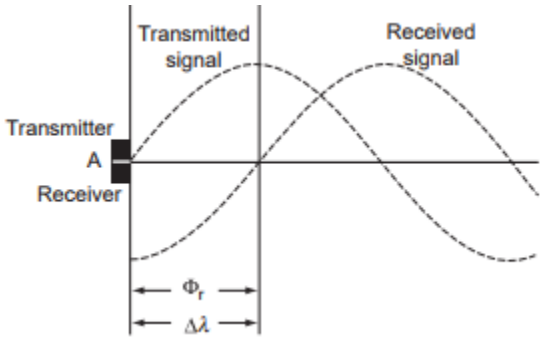


Figure (2-10): phase comparison between the two signals takes place at the laser rangefinder located at A. [2]

of wavelengths (M) can be determined and added to the fractional values to give the final slant range (R).

$$R = (M\lambda + \Delta\lambda)/2 \quad (2.5)$$

of wavelengths (M) can be determined and added to the fractional values to give the final slant range (R).

where

M is the integer number of wavelengths

λ is the known value of the wavelength

$\Delta\lambda$ is the fractional part of the wavelength = $(\phi/2\pi) \cdot \lambda$, where ϕ is the phase angle [2]

2-5 Laser scanner TX5

The Trimble TX5 3D laser scanner is a high-speed three-dimensional laser scanner for detailed measurement and documentation. The laser scanner uses laser technology to produce exceedingly detailed three-dimensional images of complex environments and geometries in only a few minutes. The resulting images are an assembly of millions of 3D measurement points. [11]

The main features are:

- High accuracy
- High resolution
- High speed.
- Intuitive control via the built in touchscreen display.
- High mobility due to its small size, light weight, and the integrated quick charge battery
- Photorealistic 3D color scans due to the integrated color camera.
- Integrated dual axis compensator to automatically level the captured scan data
- Integrated compass and altimeter to give the scans an orientation and a height information
- WLAN to remotely control the scanner. [11]

Figure (2-11): Terrestrial Laser Scanner TX5 [11]

2-5-1 How laser scanner works

In principle, the TX5 3D laser scanner works by sending an infrared laser beam into the center of its rotating mirror. The mirror deflects the laser beam on a vertical rotation around the environment being scanned; scattered light from surrounding objects is then reflected back into the scanner. [11]

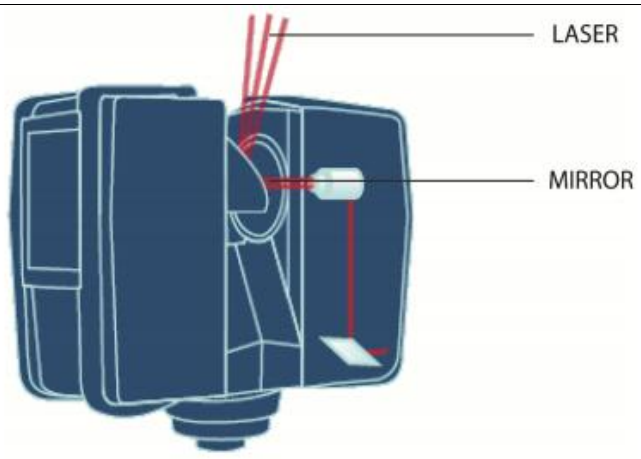


Figure (2-12): Principle of Terrestrial Laser scanner TX5 [11]

To measure the distance, the laser scanner uses phase shift technology, where constant waves of infrared light of varying length are projected outward from the scanner. Upon contact with an object, they are reflected back to the scanner. The distance from the scanner to the object is accurately determined by measuring the phase shifts in the waves of the infrared light. HYPERMODULATION greatly enhances the signal-to-noise ratio of the modulated signal with the help of a special modulation technology. The x, y, z coordinates of each point are then calculated by using angle encoders to measure the mirror rotation and the horizontal rotation of the laser scanner. These angles are encoded simultaneously with the distance measurement. Distance, vertical angle and horizontal angle make up a polar coordinate (δ, α, β), which is then transformed to a Cartesian coordinate (x, y, z). The scanner covers a 360° x 300° field of view**.** [11]

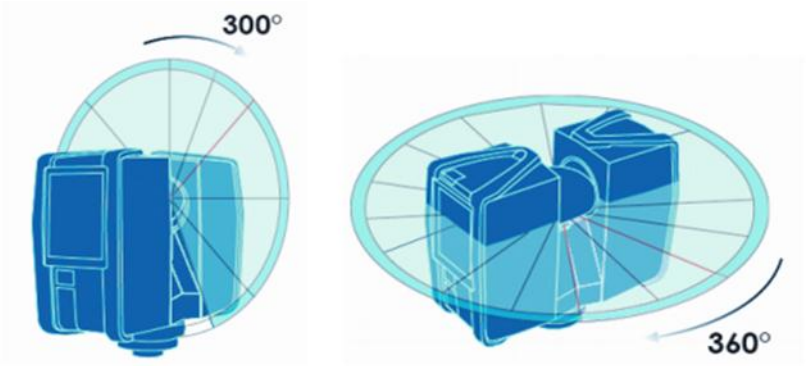


Figure (2-13): The scanner covers a 360X 300 field view [11]

Additionally, the laser scanner determines the reflectivity of the captured surfaces by measuring the intensity of the received laser beam. In general, bright surfaces reflect a greater portion of the emitted light than do dark surfaces. This reflectivity value is used to assign a corresponding grey value to each single point. The single point measurements are repeated up to 976,000 times per second. The result is a point cloud, a three-dimensional dataset of the scanner's environment (hereinafter referred to as the "laser scan" or "scan"). Depending on the selected resolution (points acquired per rotation) each point cloud consists of millions ofscan points. The laser scans are recorded to the removable SD memory card, enabling easy and secure transfer to Trimble TX5 SCENE software.[11]

2-6 3D Modeling

Building footprint, height, volume, and three-dimensional (3D) shape information can be used to estimate energy demand, quality of life, urban populations, property tax, and surface roughness .Three-dimensional building models are essential for 3D city or urban landscape models, urban flooding prediction, and assessment of urban heat island effects. Building models can be divided into two categories: simple and sophisticated. Geometric attributes for a simple building model consist of a footprint polygon and a height value. [5]

The geometric attributes for a sophisticated building model include not only a footprint polygon but also planes or other types of surfaces for various parts of the roof as well as their projections (polygons) on the ground plane. Only one fixed building height exists for a simple model, while building heights of the sophisticated model are variable. The advantage of the simple model is that it requires fewer geometric attributes to delineate a building. The buildings are represented by various types of 3D boxes; therefore, the 3D rendering is fast. Many commercial GIS software packages such as ArcGIS (www.esri.com) can display simple building models effectively. The simple building model is sufficient for applications such as numerical modeling of urban flooding and heat island effect, estimating urban population and energy demand, and large-scale 3D visualization, all of which do not require the details of buildings. The key to extracting a simple building model from LiDAR measurements is to derive footprints. The building height value can be represented using statistical height values such as mean and median of LiDAR measurements within a footprint. [16]

The disadvantage of the simple building model is the lack of detail and accuracy of building shapes. The sophisticated models overcome this limitation by offering more geometric information for 3D buildings. The sophisticated models are required by applications such as hurricane wind damage models for individual properties, property tax estimation, and detailed urban landscape modeling. The disadvantage of sophisticated models is that the 3D rendering is slow. Most existing commercial GIS software cannot display sophisticated models efficiently due to compound geometric structures. [3]

High-resolution data needed for extracting 3D building models can be derived through airborne light and detection (LiDAR) mapping systems. However, airborne LiDAR systems generate voluminous and irregularly spaced 3D point measurements of objects, including ground, building, trees, and cars scanned by the laser beneath the aircraft. The sheer volumes of point data require dedicated algorithms for automated building reconstruction. This paper focuses on presenting a framework for the construction of simple and sophisticated building models from LiDAR measurements. [15]

2-6-1 Separation of Ground and Non ground Points

The critical step in building model construction is to identify building measurements from LiDAR data. Two ways are often utilized to identify building measurements from LiDAR data. One is to separate ground, buildings, trees, and other measurements from LiDAR data simultaneously using segmentation. Examples of this method can be found in Maas and Filin and Pfeifer.[5]

The other way is to separate the ground from non-ground LiDAR measurements first and then identify the building points from nonground measurements. Numerous algorithms have been developed to identify ground measurements from LiDAR data. For example, Vosselman proposed a slope-based filter to remove non ground measurements by comparing slopes between a LiDAR point and its neighbors.[15]

Shan and Sampath applied a slope-based 1D bidirectional filter to LiDAR measurements along the cross track direction to label non ground points [5]. Zhang et al. used mathematical morphology to identify ground measurements. The non-ground measurements can be simply derived by removing identified ground data from a raw data set. To facilitate the data processing, digital surface models (DSMs) interpolated from raw LiDAR measurements and digital terrain model (DTM) interpolated from ground measurements are often produced. The image for non-ground objects is derived by subtracting DTM from DSM. Examples using this method to derive non ground objects can be found in Refs.[1]

2-6-2 Building Measurement Identification

The next step is to extract building point measurements or pixels in the data set for non-ground objects, which are dominated by trees and buildings. The distinct difference between buildings and trees is that the roof surfaces are approximately planar, while canopy surfaces are irregular. Several parameters based on this difference have been proposed for segmenting buildings and trees. For example, the first derivatives of heights are either a zero (fl at roof) or constant (sloped roof) for a planar surface, and the second derivatives of a sloped planar surface are zero. The first and second derivatives of an irregular surface should be variable. Morgan and Tempfl i applied Laplacian and Sobel operators to height surfaces to separate building and tree measurements [8]. The problem with this method is that the derivatives from LiDAR measurements for roofs are not constant because of measurement errors. Small features such as chimneys, water tanks, and pipes on a roof surface can also produce abnormal derivative values. In addition, the derivatives at the edge of a building have a large variation, making itdiffi cult to separate buildings from adjacent trees.

Another technique to separate building and tree measurements involves using a least squares method to estimate the parameters for a plane that fits a LiDAR point and its neighbors within a local window. It is expected that the deviations of roof points from their fitted planes

will be small and the plane parameters will be similar and consistent, while deviations and plane parameters for tree points will be large and variable. Compared to derivatives, the plane parameters are less sensitive to individual outliers caused by chimneys and water tanks. The drawback of this method is that plane parameters are not robust at the boundary of the buildings because fewer points are available for parameter estimation.[15]

Many airborne LiDAR systems are capable of deriving the first and last return measurements produced by multiple reflections of a laser pulse by the objects on the earth's surface. The height difference between the first and last return measurements can be used to separate building and tree measurements [14]. The height difference is usually large for tree measurements and close to zero for building measurements. A measurement is identified as a building measurement by comparing its height difference to a predefined threshold. However, this method does not work for areas covered by dense trees where laser pulses cannot penetrate. In addition, the elevation difference between the first and last returns less than 3–4 m is not reliable because of the influence of the laser pulse width, which is typically 10 ns. [1]

Hough transform has also been used to separate building points from tree measurements or to identify them directly from a raw LiDAR data set. Data in the physical space are transformed to and analyzed in the parameter space. The advantage of the Hough transform is its tolerance of gaps in the feature boundary. However, it is difficult to define the optimum cell size for voting in the parameter space, which is influenced by the error range of LiDAR measurements, the sampling density, and local height changes of a roof surface. Unfortunately, local height changes of individual surfaces are different; thus it is difficult to quantify these changes using a single value. Usually, the cell size is set empirically, and if the cell size is too large, several real-world planes could be merged into one during building identification. In contrast, if the cell size is too small, one real-world plane could be split into several smaller planes. In addition, the adjacency of point measurements is not considered by the Hough transform. Therefore, LiDAR measurements for separated but closely adjacent buildings with the same height could be mixed into the same category.[1]

The LiDAR measurements for buildings and trees can be segmented based on one or more of the above-mentioned parameters. There are two ways to perform the segmentation task. One is the point- or pixel-level classifi cation. Maas employed raw elevation , Laplace filtered height, and maximum slope from LiDAR measurements to perform a supervised maximum likelihood classification. Filin separated LiDAR measurements for buildings, vegetation, ground, and other features using an unsupervised classification based on the position of a point, the parameters of the tangent plane to the point, and the relative height difference between the point and its neighbors. Elberink and Maas used unsupervised k-means classification in their texture-based segmentation. The problem with a point-level classification is that the measurements for a building cannot be guaranteed to classify into the same category. Also, the selection of training data sets for a supervised classification can be very time consuming. An alternative way is to fi

nd a building area using a region-growing algorithm based on a seed point. This method considers the adjacency of LiDAR measurements and the robustness of region-growing processes.[14]

2-6-3 Building Model Creation

After measurements for a building are identifi ed, a raw footprint can be derived by connecting boundary points of LiDAR measurements for a building. The raw footprint polygon has to be generalized for building models because the raw footprint includes considerable noise due to irregularly spaced LiDAR points. Alharthy and Bethel employed the histogram of boundary points to generalize the footprint edges by assuming that the buildings have only two dominant directions that are perpendicular with each other. Based on the same assumption, Sampath and Shan used the least squares model to regularize the footprint edges. Recently, Zhang et al. have published a method to refine a footprint iteratively based on estimated dominant directions. Footprints with oblique edges, which are not perpendicular to dominant directions, are allowed in this method as long as the total length of the oblique edges is less than the total length of the edges parallel or perpendicular to the dominant directions.[12]

Simple building models can be derived by adding a uniform height value once a refined building footprint is derived. However, the process to derive sophisticated building models is more complicated. Schwalbe et al. categorized the methods to derive 3D building models into two categories: model-driven and data-driven. In the model-driven method, building models are identified by fitting predefined models into the LiDAR measurements For example, Maas and Vosselman estimated parameters for primitive building models based on the invariant moment analysis. Brenner extended this method to complex buildings by splitting a building into simple primitives fi rst and then fi tting individual primitives using point clouds [15]. However, building models for a study area are not always available in advance, which limits the application of the model-driven method.

In the data-driven method, building measurements are grouped first for different roof planes. Then, the 2D topology of each building, which is represented by a set of connected planar roof surfaces projected onto a horizontal plane, is derived. Rottensteiner and Jansa approximated pixels on edges with line segments first and then intersected these line segments to derive the vertices of the 2D topology. Alharthy and Bethel derived the polygon for each roof plane by connecting the boundary points and simplifying the polygon edges using the Douglas–Peucker algorithm. Polygons for each roof plane are then snapped into the 2D topology. Since neighboring roof polygons may overlap or be separated from each other, it is not easy to snap the neighboring polygons together. [9]

Raw 3D building models can be directly created from the derived 2D topologies and identified roof planes. However, the quality of such building models is poor because a 2D topology is often noisy due to irregularly spaced LiDAR measurements. Uncertainty in estimated

roof plane parameters due to errors in LiDAR measurements and segmentation also distort raw building models. A refinement of 2D topology and roof plane parameters is often needed to derive high quality building models. Many geometric constraints have been proposed to regularize and refine the 2D topology. Gruen and Wang proposed seven constraints formulated into weighted observation equations, and they enforced these constraints by using least squares adjustment (LSA). Before LSA is applied, points, lines, or planes related to each constraint should be grouped together manually, a step that limits the application of this method. Other methods realized the difficulty in enforcing many constraints simultaneously and only enforced important constraints, especially parallelism. The parallelism constraint assumes that a building has two dominant directions perpendicular to each other, and the building edges on the 2D topology should be parallel to either of the dominant directions. However, this assumption is too strict and cannot be applied to buildings with edges oblique to both dominant directions.[5]